Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 24 (2022), Issue 12 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Viewpoint

Tutorial

Corrigenda and Addendas

Research Letter

Review

Digital Platform Uses for Help and Support Seeking of Parents With Children Affected by Disabilities: Scoping Review

Oliver Gruebner[1,2], PhD; Afua van Haasteren[3], PhD; Anna Hug[1], BSc; Suzanne Elayan[4], PhD; Martin Sykora[4], PhD; Emiliano Albanese[3], PhD; John Naslund[5], PhD; Markus Wolf[6*], PhD; Marta Fadda[3*], PhD; Michael von Rhein[7*], PhD

[1]Department of Geography, University of Zurich, Zurich, Switzerland

[2]Epidemiology, Biostatistics, and Prevention Institute, University of Zurich, Zurich, Switzerland

[3]Institute of Public Health, Università della Svizzera italiana, Lugano, Switzerland

[4]Centre for Information Management, Loughborough University, Loughborough, United Kingdom

[5]Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, United States

[6]Department of Psychology, University of Zurich, Zurich, Switzerland

[7]University Children's Hospital Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Oliver Gruebner, PhD
Department of Geography
University of Zurich
Winterthurer Strasse 190
Zurich, 8057
Switzerland
Phone: 41 63 55152
Email: oliver.gruebner@uzh.ch

Abstract

Background: Receiving a diagnosis that leads to severe disability in childhood can cause a traumatic experience with long-lasting emotional stress for patients and family members. In recent decades, emerging digital technologies have transformed how patients or caregivers of persons with disabilities manage their health conditions. As a result, information (eg, on treatment and resources) has become widely available to patients and their families. Parents and other caregivers can use digital platforms such as websites or social media to derive social support, usually from other patients and caregivers who share their lived experiences, challenges, and successes on these platforms. However, gaps remain in our understanding of platforms that are most frequently used or preferred among parents and caregivers of children with disabilities. In particular, it is not clear what factors primarily drive or discourage engagement with these digital tools and what the main ethical considerations are in relation to these tools.

Objective: We aimed to (1) identify prominent digital platforms used by parents or caregivers of children with disabilities; (2) explore the theoretical contexts and reasons for digital platform use, as well as the experiences made with using these platforms reported in the included studies; and (3) identify any privacy and ethical concerns emerging in the available literature in relation to the use of these platforms.

Methods: We conducted a scoping review of 5 academic databases of English-language articles published within the last 10 years for diseases with childhood onset disability and self-help or parent/caregiver-led digital platforms.

Results: We identified 17 papers in which digital platforms used by parents of affected children predominantly included social media elements but also search engines, health-related apps, and medical websites. Information retrieval and social support were the main reasons for their utilization. Nearly all studies were exploratory and applied either quantitative, qualitative, or mixed methods. The main ethical concerns for digital platform users included hampered access due to language barriers, privacy issues, and perceived suboptimal advice (eg, due to missing empathy of medical professionals). Older and non–college-educated individuals and ethnic minorities appeared less likely to access information online.

Conclusions: This review showed that limited scientifically sound knowledge exists on digital platform use and needs in the context of disabling conditions in children, as the evidence consists mostly of exploratory studies. We could highlight that affected families seek information and support from digital platforms, as health care systems seem to be insufficient for satisfying knowledge and support needs through traditional channels.

(J Med Internet Res 2022;24(12):e37972) doi:10.2196/37972

KEYWORDS

digital place; pediatric diagnoses; conditions; disability; neuromuscular; information and support seeking; online; social media; peer support; lived experience; parents; children; youth; review; scoping review; trauma; caregivers

Introduction

Receiving the diagnosis of a disease leading to disability in childhood can cause long-lasting emotional stress for patients and family members. There is considerable evidence illustrating the adverse mental health consequences and higher levels of psychological distress experienced among children with disabilities [1,2]. Similarly, parents and caregivers (hereafter referred to as "parents" for simplification) of children with disabilities also experience elevated stress and can face challenges adapting to the care needs of their child [3]. Importantly, efforts are needed to support parents in adapting and meeting the needs of their child, as this can directly impact the child's development and well-being over the life course [4,5]. Therefore, it is critically important to determine effective approaches for supporting parents of children with disabilities, so that they can adopt necessary and desired coping strategies and feel confident in meeting the day-to-day needs of their children.

In recent decades, emerging digital technologies have transformed how patients or parents of persons with disabilities manage their health conditions [6], and information (eg, on treatment and resources) has become more widely available to them. Furthermore, social and emotional support (eg, through online self-help and peer support groups) is now more readily accessible through various online platforms. For instance, Oldenburg et al [7] present a helpful rundown of the role new media have played (eg, PatientsLikeMe) in supporting patients of children with chronic diseases, while in the work of Sykora [8], some early health-related social platforms are mentioned (eg, PatientOpinion, CarePages, CureTogether, and PatientsLikeMe), and a walkthrough of the social platform CureTogether (now defunct after being bought by 23andMe) is provided. Most recently, patients suffering from "long COVID" (referring to the recent COVID-19 pandemic) who were being dismissed by their health care professionals were able to mobilize by sharing their symptoms and locating other sufferers through social media. This ultimately resulted in a new chronic condition known as "long COVID" and the creation of what are now known as "long COVID clinics" to support patients [9]. Parents of children with debilitating diseases can potentially use digital platforms such as search websites or social media (eg, Reddit or WhatsApp groups) to derive social support, usually from other patients and parents who share their lived experiences, challenges, and successes on these platforms.

However, gaps remain in our understanding of platforms that are most frequently used or preferred among parents of children with disabling conditions. For example, it is not clear what factors drive or discourage engagement with these digital tools. In addition, there is little evidence available about the ethical concerns over services provided by digital platforms that are used by parents for information and support seeking in the context of a disabling or lethal disease of their child.

Accordingly, in this scoping literature review, we aimed to (1) identify prominent digital platforms used by parents or caregivers of children with disabilities; (2) explore the theoretical contexts and reasons for digital platform use, as well as experiences with using these platforms reported in the included studies; and (3) identify any privacy and ethical concerns emerging in the available literature in relation to the use of these platforms.

Methods

Study Design

We conducted a scoping review following the framework of Arksey and O'Malley [10]. Scoping reviews are useful in mapping and identifying available evidence [11]; therefore, we opted for this approach rather than other types of reviews, which often answer a single clinical question, because we were more concerned with broadly exploring a concept [12]. The search was performed using 5 scientific databases: PubMed, CINAHL, PsycINFO, Communication & Mass Media Complete, and Psychology & Behavioral Sciences Collection. EBSCO Host was used to concurrently search through all the databases except for PubMed. In line with Arksey and O'Malley [10], the reference lists of the articles included were screened for additional studies. Gray literature searches were also conducted on the websites of various major organizations tackling neuromuscular diseases (NMDs; Multimedia Appendix 1), in addition to using Google search engine to retrieve further studies. All the searches were conducted between July and September 2021.

Search Strategy: Identifying Relevant Studies

Based on the severity of the disease and the high psychoemotional distress it can cause to the parents of affected children, initial searches began with a primary focus on retrieving studies relating to NMDs with a pediatric onset such as Duchenne muscular dystrophy (DMD). However, these searches resulted in few studies relevant to the subject of interest. Therefore, a search strategy was adopted to include "disabilities" as a broader keyword. Table 1 details the keywords and search terms used to identify relevant studies. The inclusion criteria to identify relevant papers were (1) scientific English articles published in the last 10 years (2011- 2021) on diseases with childhood-onset disability, (2) all study types (eg, reviews, original studies), (3) use of self-help or parent/caregiver-led digital platforms (eg, internet, websites, social media or online support groups), and (4) those describing either reasons, expectations, concerns, suggestions, or experience on digital platforms. The exclusion criteria were (1) non-English articles, (2) articles published before 2011, (3) adult-onset diseases, (4) papers reporting on digital platforms maintained by medical

institutions or those designed for research, and (5) papers with a main focus on health professionals' experiences with digital platforms.

The exclusion of non-English articles was due to our inability to analyze articles in non-English languages at the time of our research; however, we must emphasize that we will endeavor to include other languages in future studies. Our focus on the past decade in our inclusion criteria is due to the relatively recent emergence of social media, which only appeared in the first decade of the 21st century and gained increasing popularity [13] in its current form from around 2009 onward. The landscape and nature of social media's interactive affordances have also evolved substantially [14], which is why we deemed that extending the study period beyond 1 decade would become problematic.

Two authors (AH and AvH) independently screened the titles, abstracts, and full texts, while a third author (MF) was consulted to establish a consensus.

Table 1. Keyword searches conducted on titles and abstracts.

Key concepts	Search terms
Parents/caregivers	parent* OR caregiver* OR carer* OR mother* OR father* AND
Children affected by disability	'child* disab*' OR 'child* disorder*' OR 'pediatric disab*' OR 'disabled persons' OR 'physical disab*'AND
Communication, exchange	communicat* OR experienc* OR challenge* OR connect* OR support* OR exchang* AND
Internet/social media support	Internet* OR online* OR 'social media*' OR webs* OR virtual* OR 'online support' OR 'Self-help Groups' OR Facebook OR Twitter OR WhatsApp OR Reddit OR Instagram OR 'mobile App*'

Charting the Data

The descriptive attributes of each article including the authors, year of publication, country, and objective of the study were extracted from each article. To facilitate the process of identifying the most prolific digital platforms, the scope of each study and its objectives, along with the respective outcomes measured, were also extracted from each full-text article included in the review.

Collating, Summarizing, and Reporting the Results

To derive an overview of the informational needs of parents of children with disabilities, the preferred online platforms, specific experiences, expectations, concerns, and suggestions for improvement highlighted by each study were identified. These were later labeled under broader concepts and organized around more general, coherent themes. In the last stage of the analysis process, common and divergent themes and topics in findings among and across all the included articles were identified.

Results

Initial Findings

Our search yielded a total of 184 scientific articles. Additionally, 18 articles were identified by reference list screenings, and 2 articles [15,16] were obtained from gray literature, bringing the total number of retrieved articles to 204. Of these, 16 records (7.8%) were identified as duplicates and excluded. Of the resulting 188 articles (100%), 153 (81.4%) were excluded according to the inclusion and exclusion criteria based on their titles and abstracts. Subsequently, we screened the full texts of the remaining 35 articles (100%) and further excluded 18 articles (51.4%) based on our inclusion and exclusion criteria, with 17 final articles (100%) included in the review. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flow diagram in Figure 1 displays the entire process involved in selecting the included papers.

Detailed information about the year of publication, study design and sample size, location, study objective, study population, main digital platform, and outcomes measured can be found in Table 2. The earliest study included in our review was conducted in 2011, while the most recent study was conducted in 2020 [15,17]. Most studies (n=14, 82.4%) were original and observational, applying either qualitative (n=5, 29.4%) [15,18-21], quantitative (n=5, 29.4%) [17,22-25], or mixed methods approaches (n=4, 23.5%) [26-29]. The remaining articles consisted of 2 (11.8%) reviews [30,31] and 1 (5.9%) case study [16]. Mothers made the bulk of the study participants in all the included studies aside from Ammari and Schoenebeck [20], where efforts were made to overrecruit fathers. In 1 (5.9%) study (Rocha and colleagues [24]), the gender of parents was not identified, likely due to the study's recruitment of participants through 2 online registries (Simons Variation in Individuals Project and GenomeConnect). The target population in the studies was most commonly defined as parents of children across a range of disorders and special needs (n=13, 76.5%), while 1 (5.9%) study's population focused on families in general [26], 2 (11.8%) on patients themselves [24,25], and 1 (5.9%) solely on married mothers with up to 5 children [19]. Most studies were conducted in the United States (n=5, 29.4%) [15-17,19,29], while participants for 6 studies (35.3%) were derived from multiple countries through online recruitment [20,23,24,27,30,31]. The remaining studies were from Australia (11.8%) [26,28], Italy (5.9%) [25], Kuwait (5.9%) [22], Norway (5.9%) [21], and the Netherlands (5.9%) [18].

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

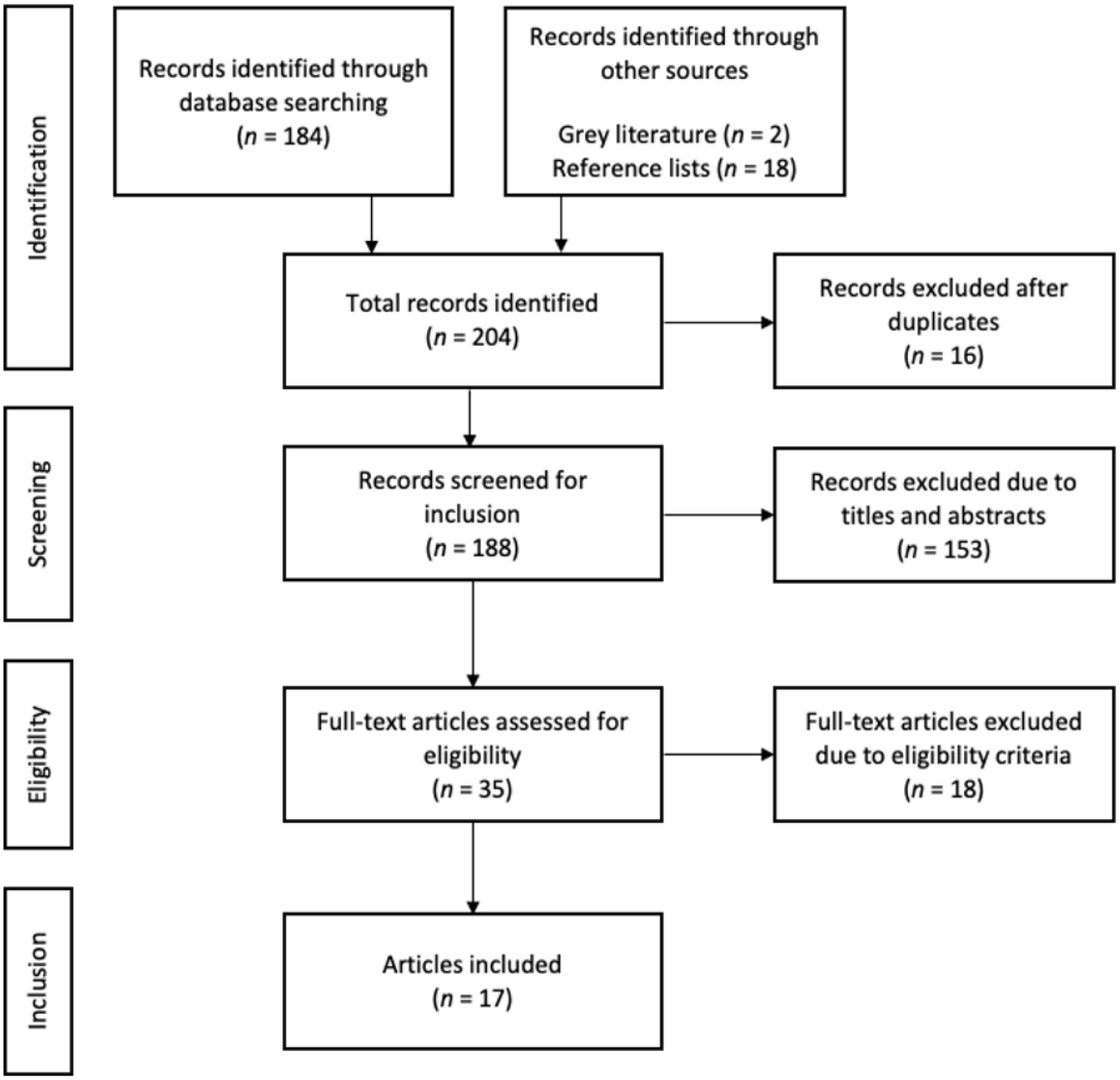


Table 2. Data extracted from the included articles.

Author, year	Study design (N)	Location	Study objective	Study population	Main digital platform[a] (outcome measured[b])
Gundersen, 2011 [21]	Interviews (N=10)	Norway	Internet use for coping with chronic illness resulting from rare genetic disorders	Parents whose children have rare genetic disorders	1, 2 (a, b, c, e)
Knapp et al, 2011 [17]	Survey (N=2371)	United States	Low-income parents of children with special needs access and use; factors related to internet use; parents' eHealth literacy, and factors associated with higher eHealth literacy	Parents of children with special health care needs	1, 2 (a, d, e)
Tozzi et al, 2013 [25]	Survey (N=516)	Italy	Details internet user profiles and how internet use affects decision-making	Patients of rare diseases	1, 2 (a, c, e)
Johnston et al, 2013 [26]	Mixed methods: survey, (N=522), focus group (N=21)	Australia	How the internet can assist families with young disabled children to make effective intervention and support decisions	Families of young children with disabilities	1, 2 (a, d, e)
Ahmed, 2014 [31]	Literature review (N=15)	Online	Summarize existing recommendations on internet use by parents of children with rare and difficult illnesses	Parents whose children have rare, difficult illnesses and special needs	1, 2 (c, d)
Ammari et al, 2014 [29]	Mixed methods: interview (N=18), survey (N=205)	United States	Use of social media sites by parents of children with special needs for information and social support; perception and management of online and offline judgment; posts perceived to be socially appropriate to post on their own online profiles versus in shared online groups; how social media sites can better support special needs families	Parents of children with special needs	1 (a, c, d)
Al-Daihani and Al-Ateeqi, 2015 [22]	Survey (N=240)	Kuwait	Information seeking behavior of parents of children with disabilities	Parents of children in a school for special needs	1, 2 (a, c, e)
Ammari and Schoenebeck, 2015 [20]	Semistructured interviews (N=43)	Online	The use of social media needs by parents with special needs children	Parents of children with special needs	1 (a, c, e)
Russell et al, 2016 [23]	Quantitative assessment of Facebook likes and posts; survey (N=49)	Canada, United Kingdom, Australia	Development and evaluation of web-based research advisory community that links parents to researchers to improve research and affected families/children's lives	Parents of children with special needs who used a Facebook group	1 (a, b, c, d)
DeHoff et al, 2016 [30]	Scoping review (N=N/A[c]), expert interviews (N=N/A)	Online	Status of research on the usefulness of digital communication like social media, in providing informational and emotional support	Parents of young children with special health care needs	1, 2 (a, b, c, e)
Fostervold, 2016 [16]	Case study (N=1)	United States	How social media posts support parents in raising their children with a disability	Parents of a child with a disability	1 (a)
Alsem et al, 2017 [18]	Semi-structured interviews (15)	Netherlands	Information needs, process of seeking and evaluating information, and the different sources of information for parents	Parents of children with disabilities	1, 2 (a, b)

Author, year	Study design (N)	Location	Study objective	Study population	Main digital platform[a] (outcome measured[b])
Nicholl et al, 2017 [27]	Mixed methods: survey (N=128), focus group, (N=8)	Ireland, Northern Ireland, United States, United Kingdom	General internet usage patterns, types of information frequently searched for, and effect of internet-sourced information on parents of children with rare conditions	Parents of children with rare conditions	1, 2 (a, b, d)
Sharaievska and Burk, 2018 [19]	Semistructured interviews (N=8)	United States	Role of online and offline support groups in the lives of families with children who have developmental disabilities	Married mothers who had 1-5 children with developmental disabilities	1 (a, e)
Rocha et al, 2018 [24]	Survey (N=103)	Online	Understand the online behavior, perspectives, and norms of rare disease communities to provide preliminary guidance to genetic counselors who wish to have discussions about social media support resources	Patients with newly described or rare genetic findings from online patient registries	1 (a, c, d)
Tracey et al, 2018 [28]	Mixed methods: survey (N=291), focus group (N=56)	Australia	Information-seeking behavior of parents and their perceptions and evaluations of the various information sources available	Parents of children with disabilities	1, 2 (a, b, c, d, e)
Terra, 2020 [15]	Semistructured and open-ended interviews (N=5)	United States	Role of social media to empower and provide community for parents raising children with profound multiple disabilities	Parents of children with profound multiple disabilities	1 (a, c)

[a]Digital platforms: (1) social media (eg, Facebook, Twitter, email), (2) internet search engines, health apps, medical websites, or not specifically mentioned otherwise.

[b]Outcome measured: (a) reasons for use, (b) expectations from use, (c) concerns/shortcomings, (d) suggestions for improvement, (e) satisfaction and experience.

[c]N/A: not available.

Digital Platforms Utilized

We classified the types of digital platforms identified in the reviewed articles into 2 categories: (1) digital platforms with social interaction options, such as social media; and (2) other platforms, such as search engines, medical websites, and health-related apps. Due to the overall aim of this review and the search strategy applied, health-related apps were not prominently found. As listed in Textbox 1, social media were the most prolific digital platforms used by caregivers and parents and were mentioned in 3 (17.7%) of 17 papers [17,21,22]. Furthermore, 1 (5.9%) study [19] focused entirely on online support groups by comparing the differences between online and offline interactions, whereas all other studies (94.1%) examined online support within the context of other digital platforms [21-26]. Some studies (n=4, 23.5%) reported on the use of internet search engines [18,25,27,28] or other online information sources (n=2, 11.8%) [17,26]. Medical websites that were frequented by caregivers were also identified in some studies (n=4, 23.5%) [21,22,25,29]. Differences in digital platform preference were evident among different age groups, as noted by Tozzi and colleagues [25]. They found that compared to younger age groups, respondents 55 years or older appeared to be less familiar with Twitter or smartphones, preferring to use email and Facebook instead [25].

Textbox 1. Digital platforms mentioned in the reviewed literature.

1. Platforms with social interaction options, such as Facebook, Twitter, and Instagram, YouTube, Pinterest, LinkedIn, Skype, Viber, MSN messenger, Yahoo! Answers (operating between June 2005 and May 2021), Yahoo Groups, Quora, Google groups, CaringBridge, CarePages (shut down in December 2017), and other online forums, blogs, discussion boards, and emails
2. Other platforms, such as search engines, medical websites (BabyCenter website, Better Start website, autism support websites), and health-related apps

Theoretical Contexts and Reasons for Digital Platform Use and Experiences Made

Overall, 5 (29.4%) studies [15,16,19-21] adopted various theoretical frameworks guiding the understanding of how social interactions and support work. First, the Ecological Model of Human Development, as used in Fostervold [16], is a theory that helps us understand the interconnectedness of family and the larger society and the resulting socialization of a child. The Symbolic Interaction Framework [32] employed in the study by Sharaievska and Burk [19] suggests that individuals' perception of reality is constructed through their interaction with the people and objects around them. Terra's thesis [15] applied 2 theories, namely, the Theory of Sense of Community and the Empowerment Theory. Based on the Theory of Sense of Community developed in 1976 and published in 1986 by McMillan and Chavis [33], this thesis "sought to explain the dynamics of the sense-of-community force" [15]. The identified components of sense of community were membership, influence, fulfillment of needs, and shared emotional connection [15]. The Empowerment Theory describes a process in which people gain understanding and control over personal, social, economic, or political forces in order to take action to better their lives, and it was utilized in the study by Terra [15] to focus on the impact of community membership on education, awareness, and action on behalf of their child and other children with disability. Another study [20] also focused on the Empowerment Theory and extended it into a new theory of "networked empowerment" that describes how parents whose children have received a special needs diagnosis find other parents, mobilize resources, and become advocates. The fifth study [21] used the theoretical framework of medical sociologist Aaron Antonovsky [34-36], who was dedicated to understanding how people manage to demonstrate resilience despite going through extremely difficult life experiences. Antonovsky contends that the explanation is to be found in people's capacity to manage stressors, that is, "demands to which there are no readily available or automatic adaptive responses" [36].

From the reviewed literature, we noted that digital platforms were predominantly used for information retrieval and social support. As noted by Gunderson [21], no 2 digital platforms were considered equivalent for deriving various types of information by their study participants. Therefore, parents chose to use either platform based on their respective needs. The criteria considered necessary to facilitate the utility of platforms were highlighted in 2 studies. According to Nicholl et al [27], the most important attributes of platforms were relevance, accurate and up-to-date information, trustworthiness, recommendation by health professionals, easy-to-understand information, helpful references, and an appealing layout. Participants in Johnston and colleagues' [26] study echoed several of these factors, adding that presentation (different languages, videos or audio recordings, pictures, easy to navigate, and information written in easy language) and connection functionality (blog, forum, access to professionals and other parents, and access to owners of the website) increased the overall utility of a platform.

The general expectation that digital platforms would have objective, up-to-date, and vital information on conditions of interest was emphasized by participants in several other studies [18,20,22-24]. The types of information sought by parents included details about services and systems available [29,30], specialists for specific conditions [20,25], social workers [20], and appropriate schools and childcare [23]. Parents used these types of information to assist them in caring for their children as well as interacting with professionals involved in their care. They often felt empowered by the readily available information on digital platforms. In several studies, parents particularly felt the need to consult digital platforms soon after a diagnosis to learn more about the condition or before an upcoming doctor's visit [18,20-22,27].

Digital platforms also provided a means of not only communicating with parents familiar with the condition of interest but also scheduling appointments with professionals, seeking second opinions or alternative therapies [25], or communicating with family and friends [27]. For example, parents used websites such as CaringBridge and CarePages to provide updates on the status of their children's health [29]. Digital platforms such as CaringBridge and CarePages offer the opportunity to post about the status of one's condition with the primary aim of assisting others frequenting these platforms. Some parents chose to share relevant scientific research on the condition faced by their children for the benefit of others, especially after gaining more experience with services and diagnoses [16].

Participants in several studies stated that digital platforms would foster a feeling of support among the participants [18,20,22-24]. By consulting the posts by parents of children with similar symptoms and care pathways, most parents became more attenuated to what to expect and how best to care for their children [25]. Moreover, some participants in a study by Ammari and Schoenebeck [20] noted that posts from other parents (eg, on health care services and medication, special education services, or specially designed clothes) provided hope and decreased their anxiety and depression after a diagnosis. Several studies reported that parent-to-parent peer support either via social media groups or online support groups was vital in reducing feelings of isolation among parents of children with special needs [15,16,19,24]. The same was true for respondents in Gunderson's [21] study, who reported that sole help from health professionals proved insufficient, especially after initial diagnosis or during the deterioration phase of a condition. Where professionals or researchers participate in forums on digital platforms, respondents stressed the importance of their posts reflecting empathy [23]. In addition, humor was considered a viable tool to minimize the emotional toll of social media posts, according to participants in the study by Ammari and colleagues [29].

Although digital platforms were preferred in most instances because virtual interactions were easier to establish and manage, some parents hoped for the development of hybrid social connections whereby virtual relationships would translate into occasional physical interactions [15]. In other studies [20,29], online interactions through social media sites were reported to facilitate social support, especially for geographically restricted families with scarce resources in their immediate vicinities. However, social media sites were also reported to not be

facilitative in linking newly diagnosed individuals and their families with experienced ones or connecting affected individuals to others with analogous experiences [29].

Privacy and Ethical Concerns in the Use of Digital Platforms

When using social media in the context of child disability, privacy issues were imminent among several parents, as personal posts relating to photos and medical questions, for instance, were often restricted to closed groups [23,24,28]. Some studies showed that the majority of participants preferred closed over open online fora, such as closed Facebook groups to discuss personal information only with members of the group [18,20,24]. Furthermore, closed Facebook pages were preferred by participants in the study by Ammari and Schoenebeck [20] for organizing and strategizing activities, whereas public groups were used to advocate for perceived necessary policy changes. While 1 study found that the number of respondents feeling rather or very comfortable with sharing medical and personal information in a closed group decreased when having professionals present [24], there was a consensus in opinion about the presence of professionals on digital platforms, as they were considered necessary by some parents to facilitate robust information sharing [19,24,26,29].

According to Fostervold [16], issues of conflict of interest and privacy also arise when participants request to be "friends" with their health professionals on social media websites. Furthermore, possible abuse of photographs of children and medical information was noted by 1 participant in the study by Rocha et al [24]. Although parents reported feeling overall less judged online than offline, they dealt with judgment online by blocking or unfriending culprits, minimizing posts, reducing their engagement, and even deleting the respective digital platform account [29].

We also found that there were differences in digital platform use according to the sociodemographics involved. For example, the study by Tozzi and colleagues [25] found that individuals who were younger, active on social media, and already prone to communicating via electronic means were the most likely to discuss information found online with physicians. Conversely, the study by Knapp and colleagues [17] found that older individuals, non–college-educated people, non-English speaking people, and ethnic minorities were less likely to access information online [17]. The same study also found that these population groups, when compared to their reference group, were less likely to show eHealth literacy based on the eHealth Literacy Scale (eHEALS), a measure to evaluate the "ability to locate, evaluate, integrate, and apply information gained from electronic platforms" [17,37]. The language barrier of digital platforms also prevented many parents from interacting with and deriving optimum utility from digital platforms [26,28].

Digital platforms on which information was obscured and difficult to find also posed a great concern for participants [18]. Additionally, the prevalence and traction of misinformation and disinformation on digital platforms were considered particularly problematic among participants of 2 studies [15,24]. Furthermore, the expectation for unrealistic lifestyles [15,29], along with depressing posts [15,20,21,25], posed a mental health worry. For some parents, the difficulty of weighing advice found on social media information against that of professionals [15,25] was also an issue of concern. Whereas posts linked to government sources were deemed important to increase the trustworthiness of information in some studies [18], other studies found this to be insufficient and advocated for posts to include information on the original cultural context [28].

Suggestions made in another study to increase the usefulness of social media platforms included targeted pages to connect children with similar ages and conditions together, consolidating pages on similar conditions, and facilitating the online interaction between more disease-experienced parents with less experienced ones [29]. Finally, health apps focused on delivering interventions were encouraged to include and prioritize social support elements to improve their overall utility [30].

Discussion

Principal Results

The available literature shows that digital platforms used by parents of children with disabilities predominantly included social media but also search engines, health-related apps, and medical websites. Information retrieval and social support seeking were the main reasons for their utilization, with the general expectation of finding and sharing objective, up-to-date, and reliable information and guidance. In addition, the main concerns for digital platform users included privacy issues and the digital divide across sociodemographic groups, including language barriers.

Social Support From Digital Places

In our review, most of the literature reported that parents used commonly available social media platforms (eg, Facebook, Twitter, Instagram) and other online forums, blogs, and discussion boards. Social media can be defined as digital platforms that provide users with the ability to share and discuss information publicly and within individual peer networks [38]. As such, they offer a social component that includes bidirectional communication among social media users that allows for social interaction and exchange. In previous studies, researchers recognized social media platforms as so-called "digital places" that can be defined as socially constructed spaces (ie, environments) with individual meaning and utility to their users, similar to geographic places [38,39]. Following the nomenclature of Glanz et al [40], respondents in our reviewed studies used these digital places for informational and emotional support. It is noteworthy that some parents felt less charged online than offline, possibly due to the virtual character of digital places and more options to defriend or retract from social contacts more easily than in the physical world. Participants in the reviewed studies expressed the general expectation to find and share objective, up-to-date, and reliable information and guidance, which seems to be closely related to a feeling of empowerment. Importantly, this need for information seems to be closely related to the need for emotional and other forms of social support. Future research may extend the focus on multiple dimensions of digital places—how individual meaning and utility of these places may influence their use in the context of disabilities in children. Furthermore,

future research should also investigate the question of how digital place use in this context might affect mental health and resilience in patients and family members, especially during the time of diagnosis and at critical events during disease progression.

Ethical and Privacy Concerns Reported in Digital Platform Use

Individual-level characteristics and social determinants of health played a role in digital platform use. For example, older and less educated individuals, as well as ethnic minorities, were less likely to access information online compared to their younger, college-educated, White counterparts. This finding lends itself to the explanation that online digital platforms and resources are not easily accessible to everyone and once more indicates a digital divide in the context of child disability, with less access for already vulnerable families. Our review also highlighted that women were the respondents in most of the reviewed studies; this may point to a gender bias, but it may also suggest that women take over the larger care burden in families with children living with a disability. However, it is worth noting that the results on sociodemographics and digital platform use were rather old, in that they were published in 2013 [25] and 2011 [17], when smartphones were not as widely used. According to the Pew Research Center, 53% of adults in the United States owned a smartphone in 2013 [41], and 85% owned a smartphone in 2021 [42]. Nevertheless, our findings from these 2 references highlight a prevalent issue where older adults are often less likely to be familiar with the most recent social media platforms.

From a geographic perspective, the reviewed studies derive from many different countries, and it is unclear whether there are patterns of digital platform use that are distinct in some regions or others. Some of the platforms might be specifically useful or even targeted to regional, national, or cultural audiences, which should be investigated in future studies.

Privacy issues were raised in various studies, highlighting that affected individuals and parents felt more confident in closed fora and that they appreciated if professionals were verifying the information being discussed. At the same time, there was a desire to try to maintain a healthy distance from professionals to discuss private issues in a safe space. This finding points to the ambivalent relationship that parents of children with disabilities may develop with the child's health care providers.

No study in our review applied an experimental design involving the evaluation of digital platforms to test for the effects of distinct platform designs on distinct dimensions of support (ie, emotional, informational, and instrumental support and appraisal). This represents a significant limitation in the available literature because it means limited evidence in this area, as well as difficult-to-draw conclusions about the effectiveness of these platforms beyond anecdotal accounts from the explorative research summarized in this review. This is particularly true for patients with NMDs on whom research in this field seems to be widely neglected thus far despite the severity of the diseases.

Our findings have major practical implications. Physicians and other health care providers, health care facilities, and health agencies should take advantage of digital platforms that provide social interaction options to meet and empower families of patients living with disabilities. This should be done by not only identifying and addressing patients' and parents' needs before, during, and after access but also by recognizing and correcting any structural conditions that may affect individuals' opportunities to use such platforms.

Limitations

Our review is biased toward high-income countries; therefore, the relevance of the findings for use across different settings globally is difficult to ascertain. Future studies should address underrepresented cultural groups, languages, races, ethnicities, and countries to broaden our understanding of social media use in the context of pediatric diagnoses leading to disabilities and the inequities associated with it.

Conclusions

To date, scarce scientifically sound knowledge is available on digital platform use and needs in the context of disabling diagnoses in children. Our study aims to help fill this gap by highlighting which digital platforms families of children with disabilities visit, what they seek in them, and why. Most importantly, our findings on the privacy and ethical concerns in the use of these platforms remind us of the role of social determinants in shaping the magnitude of individuals' access to and benefit from these platforms. As families of children with disabilities constitute an already vulnerable population, future research should seek to identify and critically examine the avoidable, unfair, and unjust conditions that may amplify forms of inequities in their access to support. This can be done by continually committing to engage a broad range of narratives, voices, and lived experiences when conducting empirical research on digital platform uses among parents of children affected by disabilities.

Acknowledgments

We acknowledge funding from the Swiss School of Public Health (SSPH+) for the project RISE (Raising Awareness for Mental Health in People Affected by Pediatric Neuromuscular Diagnoses) to authors OG and EA. The sponsor had no role in any step of the preparation of this manuscript.

Authors' Contributions

AvH and AH conducted the study, reviewed the literature, and reported the findings. MF supervised the review process. OG, SE, MS, MW, MF, and MvR designed the study. OG took the lead in writing and prepared the first draft of the manuscript. All others

contributed to the writing of the manuscript. MW, MF, and MvR supervised the overall study. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient association websites.
[DOCX File , 13 KB - jmir_v24i12e37972_app1.docx]

References

1. Maxey M, Beckert TE. Adolescents with disabilities. Adolescent Res Rev 2016 Oct 24;2(2):59-75. [doi: 10.1007/s40894-016-0043-y]
2. Mueller CE, Prout HT. Psychosocial adjustment of adolescents and young adults with intellectual disabilities. J Ment Health Res Intellect Disabil 2009 Oct 02;2(4):294-311. [doi: 10.1080/19315860903308572]
3. Guyard A, Michelsen SI, Arnaud C, Lyons A, Cans C, Fauconnier J. Measuring the concept of impact of childhood disability on parents: validation of a multidimensional measurement in a cerebral palsy population. Res Dev Disabil 2012 Sep;33(5):1594-1604. [doi: 10.1016/j.ridd.2012.03.029] [Medline: 22534397]
4. Thompson R, Gustafson K. Adaptation to chronic childhood illness. In: American Psychological Association. Washington, DC: APA PsycBooks; 1996.
5. Varni JW, Katz ER, Colegrove R, Dolgin M. Family functioning predictors of adjustment in children with newly diagnosed cancer: a prospective analysis. J Child Psychol Psychiatry 1996 Mar;37(3):321-328. [doi: 10.1111/j.1469-7610.1996.tb01409.x] [Medline: 8707915]
6. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
7. Oldenburg B, Taylor CB, O'Neil A, Cocker F, Cameron LD. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]
8. Sykora M. A treatise on Web 2.0 with a case study from the financial markets. Loughborough University. 2012. URL: https://dspace.lboro.ac.uk/2134/11473 [accessed 2022-11-21]
9. Rushforth A, Ladds E, Wieringa S, Taylor S, Husain L, Greenhalgh T. Long Covid - the illness narratives. Soc Sci Med 2021 Oct;286:114326 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114326] [Medline: 34425522]
10. 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]
11. Anderson S, Allen P, Peckham S, Goodwin N. Asking the right questions: scoping studies in the commissioning of research on the organisation and delivery of health services. Health Res Policy Syst 2008 Jul 09;6:7 [FREE Full text] [doi: 10.1186/1478-4505-6-7] [Medline: 18613961]
12. Munn Z, Peters MDJ, 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 Dec 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
13. Thelwall M. Social network sites: users and uses. In: Zelkowitz MV, editor. Advances in Computers. Amsterdam, the Netherlands: Elsevier; 2009:19-73.
14. Sykora M. Web 1.0 to Web 2.0: an observational study and empirical evidence for the historical r(evolution) of the social web. IJWET 2017;12(1):70. [doi: 10.1504/ijwet.2017.084024]
15. Terra AA. Connections social media and parents raising children with profound multiple disabilities [dissertation]. University of the Pacific. 2020. URL: https://scholarlycommons.pacific.edu/uop_etds/3665/ [accessed 2022-11-20]
16. Fostervold Wells LM. Using social media for parental support in raising a child with a disability: a critical analysis of Facebook postings [thesis]. St Cloud State University. 2016. URL: https://repository.stcloudstate.edu/cfs_etds/12 [accessed 2022-11-21]
17. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]
18. Alsem MW, Ausems F, Verhoef M, Jongmans MJ, Meily-Visser JMA, Ketelaar M. Information seeking by parents of children with physical disabilities: An exploratory qualitative study. Res Dev Disabil 2017 Jan;60:125-134. [doi: 10.1016/j.ridd.2016.11.015] [Medline: 27914304]
19. Sharaievska I, Burk B. Recreation in families with children with developmental disabilities: caregivers' use of online and offline support groups. TRJ 2018;52(1):42-57. [doi: 10.18666/trj-2018-v52-i1-8446]

20. Ammari T, Schoenebeck S. Networked empowerment on Facebook groups for parents of children with special needs. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18-23; Seoul, Republic of Korea p. 2805-2814. [doi: 10.1145/2702123.2702324]
21. Gundersen T. 'One wants to know what a chromosome is': the internet as a coping resource when adjusting to life parenting a child with a rare genetic disorder. Sociol Health Illn 2011 Jan;33(1):81-95. [doi: 10.1111/j.1467-9566.2010.01277.x] [Medline: 20937053]
22. Al-Daihani SM, Al-Ateeqi HI. Parents of children with disabilities in Kuwait: a study of their information seeking behaviour. Health Info Libr J 2015 Jun 21;32(2):131-40; quiz 141 [FREE Full text] [doi: 10.1111/hir.12102] [Medline: 25899617]
23. Russell DJ, Sprung J, McCauley D, Kraus de Camargo O, Buchanan F, Gulko R, et al. Knowledge exchange and discovery in the age of social media: the journey from inception to establishment of a parent-led web-based research advisory community for childhood disability. J Med Internet Res 2016 Nov 11;18(11):e293 [FREE Full text] [doi: 10.2196/jmir.5994] [Medline: 27836818]
24. Rocha HM, Savatt JM, Riggs ER, Wagner JK, Faucett WA, Martin CL. Incorporating social media into your support tool box: points to consider from genetics-based communities. J Genet Couns 2018 Apr 13;27(2):470-480 [FREE Full text] [doi: 10.1007/s10897-017-0170-z] [Medline: 29130143]
25. Tozzi AE, Mingarelli R, Agricola E, Gonfiantini M, Pandolfi E, Carloni E, et al. The internet user profile of Italian families of patients with rare diseases: a web survey. Orphanet J Rare Dis 2013 May 16;8:76 [FREE Full text] [doi: 10.1186/1750-1172-8-76] [Medline: 23680013]
26. Johnston C, Tracey D, Mahmic S, Papps F. Getting the best from DisabilityCare Australia: families, information and decision making. Western Sidney University. 2013. URL: https://researchdirect.westernsydney.edu.au/islandora/object/uws:26945 [accessed 2022-11-21]
27. Nicholl H, Tracey C, Begley T, King C, Lynch AM. Internet use by parents of children with rare conditions: findings from a study on parents' web information needs. J Med Internet Res 2017 Feb 28;19(2):e51 [FREE Full text] [doi: 10.2196/jmir.5834] [Medline: 28246072]
28. Tracey D, Johnston C, Papps FA, Mahmic S. How do parents acquire information to support their child with a disability and navigate individualised funding schemes? J Res Spec Educ Needs 2017 Jul 24;18(1):25-35. [doi: 10.1111/1471-3802.12390]
29. Ammari T, Schoenebeck S, Morris M. Accessing social support and overcoming judgment on social media among parents of children with special needs. 2014 Presented at: 8th International Conference on Weblogs and Social Media; May 16; Ann Arbor, MI p. 22-31. [doi: 10.1609/icwsm.v8i1.14503]
30. DeHoff BA, Staten LK, Rodgers RC, Denne SC. The role of online social support in supporting and educating parents of young children with special health care needs in the United States: a scoping review. J Med Internet Res 2016 Dec 22;18(12):e333 [FREE Full text] [doi: 10.2196/jmir.6722] [Medline: 28007689]
31. Ahmed S. Internet and parents with children with special needs: a meta-analysis of PubMed articles. CSCW Workshop. 2014. URL: https://dsg.tuwien.ac.at/team/breskovic/downloads/ACM_CSCW_CCCIC_2014_papers/05_ahmed.pdf [accessed 2022-11-21]
32. White JM, Klein DM. Family Theories. Los Angeles, CA: SAGE; 2008.
33. McMillan DW, Chavis DM. Sense of community: A definition and theory. J Community Psychol 1986 Jan;14(1):6-23. [doi: 10.1002/1520-6629(198601)14:1<6::aid-jcop2290140103>3.0.co;2-i]
34. Antonovsky A. Unraveling the Mystery of Health: How People Manage Stress and Stay Well. San Francisco, CA: Jossey-Bass; 1987.
35. Antonovsky A. The structure and properties of the sense of coherence scale. Soc Sci Med 1993 Mar;36(6):725-733. [doi: 10.1016/0277-9536(93)90033-z]
36. Antonovsky A. Health, Stress, and Coping. San Francisco, CA: Jossey-Bass; 1979.
37. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
38. Shankardass K, Robertson C, Shaughnessy K, Sykora M, Feick R. A unified ecological framework for studying effects of digital places on well-being. Soc Sci Med 2019 Apr;227:119-127. [doi: 10.1016/j.socscimed.2018.09.022] [Medline: 30287115]
39. Moon G. Conceptions of space and community in British health policy. Soc Sci Med 1990 Jan;30(1):165-171. [doi: 10.1016/0277-9536(90)90339-t]
40. Glanz K, Rimer B, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2008.
41. Smith A. Smartphone ownership 2013. Pew Research Center. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_Smartphone_adoption_2013_PDF.pdf [accessed 2022-11-21]
42. Perrin A. Mobile technology and home broadband 2021. Pew Research Center. URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2022-11-20]

Abbreviations

DMD: Duchenne muscular dystrophy
eHEALS: eHealth Literacy Scale
NMD: neuromuscular disease
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by R Kukafka; submitted 14.03.22; peer-reviewed by M Dinleyici, D Gustafson; comments to author 28.08.22; revised version received 23.09.22; accepted 25.10.22; published 06.12.22.

Please cite as:
Gruebner O, van Haasteren A, Hug A, Elayan S, Sykora M, Albanese E, Naslund J, Wolf M, Fadda M, von Rhein M
Digital Platform Uses for Help and Support Seeking of Parents With Children Affected by Disabilities: Scoping Review
J Med Internet Res 2022;24(12):e37972
URL: https://www.jmir.org/2022/12/e37972
doi:10.2196/37972
PMID:36472896

Review

Digital Engagement of Older Adults: Scoping Review

Abraham Sahilemichael Kebede[1], MPH; Lise-Lotte Ozolins[2], PhD; Hanna Holst[2], PhD; Kathleen Galvin[1], PhD

[1]School of Sport and Health Sciences, University of Brighton, Brighton, United Kingdom

[2]Department of Health and Caring Sciences, Linnaeus University, Växjö, Sweden

Corresponding Author:
Abraham Sahilemichael Kebede, MPH
School of Sport and Health Sciences
University of Brighton
Village Way
Brighton, BN1 9PH
United Kingdom
Phone: 44 7502375585
Email: a.s.kebede@brighton.ac.uk

Abstract

Background: Digital technologies facilitate everyday life, social connectedness, aging at home, well-being, and dignified care. However, older adults are disproportionately excluded from these benefits. Equal digital opportunities, access, and meaningful engagement require an understanding of older adults' experience across different stages of the technological engagement life cycle from nonuse and initial adoption to sustained use, factors influencing their decisions, and how the experience changes over time.

Objective: Our objectives were to identify the extent and breadth of existing literature on older adults' perspective on digital engagement and summarize the barriers to and facilitators for technological nonuse, initial adoption, and sustained digital technology engagement.

Methods: We used the Arksey and O'Malley framework for the scoping review process. We searched MEDLINE, PsycINFO, CINAHL, Web of Science, and ACM digital library for primary studies published between 2005 and 2021. The inclusion and exclusion criteria were developed based on the Joanna Briggs Institute (participants, content, and context) framework. Studies that investigated the digital engagement experience as well as barriers to and facilitators of older adults' digital technology engagement were included. The characteristics of the study, types of digital technology, and digital engagement levels were analyzed descriptively. Content analysis was used to generate tentative elements using a congruent theme, and barriers and facilitators were mapped over the capability, opportunity, and motivation behavior change model (COM-B) and the theoretical domain framework. The findings were reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews).

Results: In total, 96 publications were eligible for the final charting and synthesis. Most of the studies were published over the past 5 years, investigated the initial adoption stage of digital engagement, and focused on everyday technologies. The most cited barriers and facilitators across the engagement stages from each COM-B component were capability (eg, physical and psychological changes and lack of skill), opportunity (eg, technological features, environmental context, and resources), and motivation (eg, optimism from perceived usefulness and beliefs about capability).

Conclusions: The COM-B model and theoretical domain framework provide a guide for identifying multiple and intertwined barriers and facilitators at each stage of digital engagement. There are limited studies looking into the whole spectrum of older adults' digital technology experience; in particular, studies on technological nonuse and sustained use stages are rare. Future research and practice should focus on tailored interventions accounting for the barriers to older adults' digital engagement and addressing capabilities, motivation, and opportunities; affordable, usable, and useful digital technologies, which address the changes and capability requirements of older adults and are cocreated with a value framework; and lifelong learning and empowerment to develop older adults' knowledge and skills to cope with digital technology development.

International Registered Report Identifier (IRRID): RR2-10.2196/25616

(J Med Internet Res 2022;24(12):e40192) doi:10.2196/40192

KEYWORDS

digital divide; digital engagement; older adults; older people; sustained engagement; technology acceptance; technological nonuse

Introduction

Background

Globally, remarkable progress has been made in medical interventions, health care and technological advancement, contributing to unprecedented decline in mortality rate and increase in life expectancy [1]. There are currently 703 million older adults (≥65 year), and this number is projected to double by 2050 [2]. Harnessing the numerous potentials of rapidly developing digital technology plays an important role in ensuring a better and more inclusive society, better health and social care, and economic support for the older population. However, recent surveys have indicated that a significant proportion of this age group has limited or no access to a range of digital technologies [3-6]. In addition, the diversity and quality of technology use are limited to fewer and familiar functionalities such as communication. For example, using a smartphone as a classic phone or for simply obtaining information [7].

Nowadays, an increasing number of older adults are digitally engaging and becoming competent technology users through improved accessibility features, user-centered and experience-based designs, and further education that equips older adults with essential digital skills. However, there is a long way to closing the digital divide between the ages, and the primary technological design ethos continues to be the supply side (digital developers') presupposition that *one size does fit all* which fails to account for older adult's physical and mental capability, accessibility needs, age-related changes, and lack of skill and support [8,9].

Recently, the SARS-CoV-2 pandemic has further increased the reliance on digital technology for everyday living, working from home, shopping, financial transactions, e-learning, communication, entertainment, and health service delivery (eg, remote consultation through e-consult and e-pharmacy) [10,11]. However, lack of access, awareness, and skills exacerbated existing digital inequality [12]. Beyond mere accessibility and use issues, older adults' digital experience constitutes pragmatic versus hedonic aspects, motivation based on functional, usability and aesthetic dimensions, and emotional ambivalence [13]. The perceived benefits of technologies in restoring autonomy, a sense of independence, improving the quality of life [13], decision-making [14], mobility, and social connectedness [15] constitute a positive experience. Intrusiveness, privacy and safety concerns, nonease of use, vulnerability, and social stigma can be sources of mixed feelings [13].

A scoping review that captures the nature and breadth of literature and older adults' experiences and factors influencing their digital engagement is pertinent and timely, given the fast-paced nature of this discipline. A recent review to develop a system-level framework for health technology adoption and scale-up highlighted the importance of investigating nonadoption and sustainability, the shortage of studies in this area, and the role of barrier and facilitator research as an input for organizational-level adoption [16]. Similar indications have been made in a recent scoping review that summarized the definition and models of technological adoption, which underscored the importance of research that captures the entire spectrum of the digital technology acceptance cycle, including the continued use of technology with all its temporal aspects of engagement and the quality of technology users' experiences over a long period [17,18]. This review will summarize studies that investigated older adults' digital engagement, including nonuse, initial adoption, and sustained digital engagement and the driving factors (see Textbox 1 for key concept definitions).

Textbox 1. Definitions of key review terms.

Digital technologies

- are electronic tools, systems, devices, and resources that generate, store or process data (eg, computers, smartphones, internet, information communication technology, video streaming, social media, internet games, multimedia, etc). Two main overarching categories of digital technology were investigated in this review based on the scope of functionalities [13]:
  - Everyday technologies include devices and services such as the internet, smartphones, computers, smart watches, messaging apps, social media, tablets, e-banking systems, gaming, and other technologies used to support daily living [13].
  - Remote or assistive care technologies use information communication technology devices and telecommunications networks to deliver health and social care remotely, often at home or in health and social care settings. Examples include telecare, telemedicine, ehealth, mobile health, telephone health consultations, remote monitoring technologies, and tracking technologies (alarms, sensors, fall detection devices, and wearables).

Digital engagement level

- Older adults' digital technology engagement or disengagement is conceptualized as a 3-staged continuum from technological nonuse and initial adoption to sustained engagement. See the review protocol by Kebede et al [19] for details of this typology.
  - Initial adoption: user decisions to accept or reject digital technology and the drivers that influence user's adoption
  - Sustained engagement: successful and maintained use of digital technologies after adoption was characterized by prolonged use of digital technology. For example, according to Ofcom, 3 months of regular use of internet qualify the minimum sustained engagement [20]. Additionally, willingness of the user to actively engage in co-designing and cocreating processes.
  - Nonuse: this will include studies that investigated technology abandonment, older adults' perspective on nonadoption, and associated justifications.

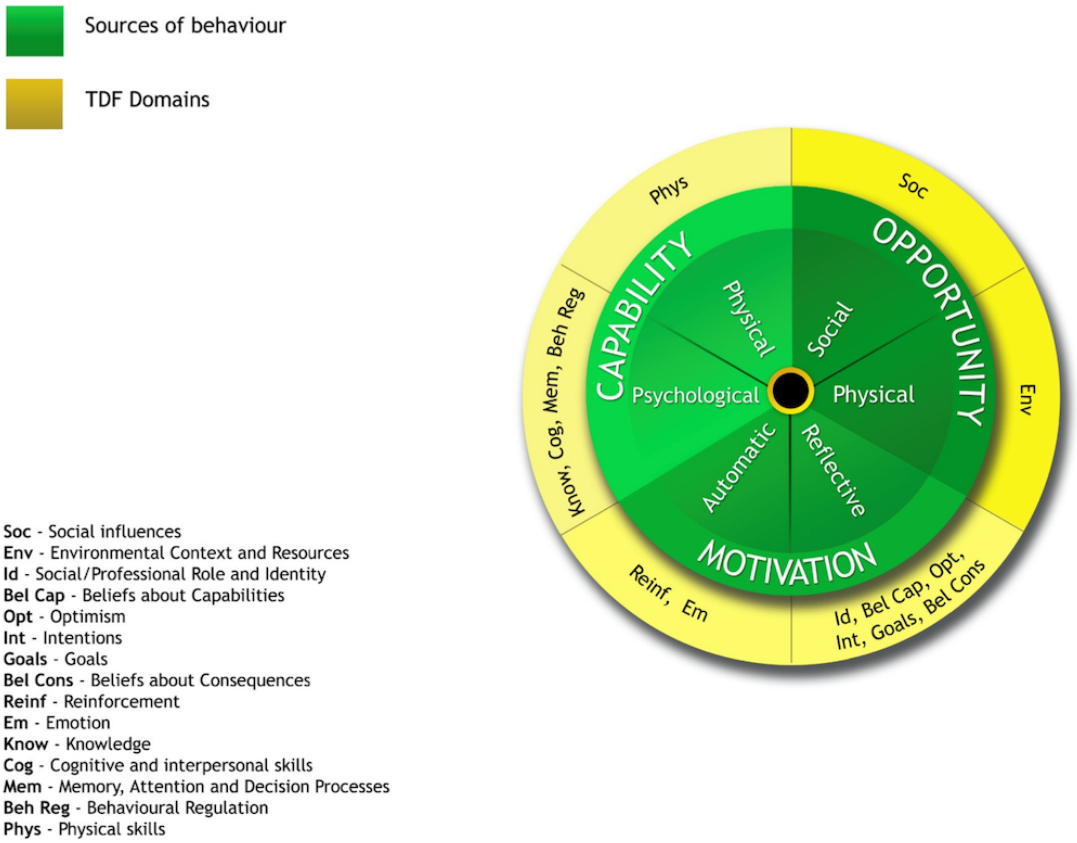
Theoretical Framework

We used the capability, opportunity, and motivation behavior change model (COM-B) and theoretical domain framework (TDF) models at the hub of the behavioral change wheel to facilitate the synthesis of the barriers to and facilitators of digital engagement among older adults. These frameworks are widely used to identify salient determinants of behavior and develop specific intervention recommendations, particularly in health system research, health care providers and service users' behavior [21]. Furthermore, the application in synthesizing evidence generated using quantitative and qualitative methodologies has increased owing to robust, structured, and replicable nature of the models [21,22].

The COM-B and TDF models are organized into 14 constructs and 3 main components. The *physical and psychological capability domain* (skills, knowledge, memory, attention, and decision process), the *automatic and reflective motivation* domain referring to the intrinsic processes for behavior and decision-making based on whether conscious and unconscious cognitive processes that influence older adults' behavior and decisions to engage digitally were included (beliefs about capabilities, optimism, consequences, intention, goals, reinforcement, and emotion), and *opportunity domain* (environmental context and social influence) [23]; see Figure 1 depicting COM-B and TDF behavioral change wheel.

The framework was adopted and customized to fit the purpose of this review and map the factors influencing older adults' digital engagement. For example, in the capability domain, physical and psychological changes attributed to age and aging-related processes had been included as identities that influence digital engagement. The environmental context, which reflects factors that are physically external to the individual, for example, the technology-related features, was grounded in this domain.

Figure 1. The behavioral change wheel combining theoretical domain framework (TDF) domains and capability, opportunity, and motivation behavior change model (COM-B) components [21,23].



Review Aim

Although our preliminary assessment indicates that there are reviews on older adults' digital engagement, previous reviews have focused on the effect of technologies on specific health or social outcomes, and there is little evidence showing the whole spectrum of users' experience journey throughout the technological engagement life cycle, especially on nonuse and sustained digital engagement [18]. Details of this engagement typology, nonuse, initial adoption, and sustained use have been published elsewhere [19]. In line with the mainstream technological models, most studies and reviews have focused on the individual motivation aspect of behavior. A comprehensive, systematic, and robust theoretical framework that helps understand individual motivations, abilities, and external social, environmental, and technological factors is required.

Therefore, in this scoping review, our aim was to map the existing literature on older adults' digital engagement, including technological nonuse, initial adoption, and sustained use using COM-B and TDF models to answer the following questions:

1. What is the extent and nature of existing evidence on older adults' digital technology engagement?
2. What are the barriers to and facilitators of older adults' digital engagement?
3. What are the gaps in research that can inform future research priorities regarding older adults' digital technology engagement?

Methods

Overview

We conducted a systematic scoping review of the literature guided by the Arksey and O'Malley framework and recent methodological developments to conceptually map the nature and extent of the literature and factors influencing older adults' digital engagement [24,25]. An extension of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) was used to present the result of the final review [26].

Eligibility

The inclusion and exclusion criteria of our scoping review were developed based on the participants, concept, and context guidelines of the Joanna Briggs Institute (see Textbox 2 for summary inclusion and exclusion criteria). Primary studies including participants with a mean age of ≥65 years that investigated everyday technologies and remote care technologies on nonuse, initial adoption, and sustained use of technology were included. Peer-reviewed studies published in English and from a global context were included in this review, whereas anecdotal evidence, reviews, and unpublished works were excluded.

Textbox 2. Articles inclusion and exclusion criteria.

Eligibility criteria for the systematic scoping review

- Inclusion criteria
  - Study types: any type of original published peer reviewed research paper using qualitative, quantitative, or mixed methodology
  - Period: any paper published between 2005 and 2021
  - Language: English
  - Population: older adults with mean age of ≥65 years as study participants
  - Concept: studies on digital engagement, both every day and remote care technologies, investigating experiences, use, barriers, and facilitators.
- Exclusion criteria
  - Study types: systematic reviews, conference papers, protocols, case studies, opinion and editorial letters, and unpublished works
  - Period: studies before 2005 and studies after 2021
  - Language: any other language
  - Population: studies primarily involving care givers, family members, or digital developers

Search Strategy

A comprehensive search strategy of major electronic databases such as MEDLINE, PsycINFO, CINAHL, Web of Science, Association of Computing Machinery Digital Library, Google Scholar, and Library and Information Science and Technology Abstracts was conducted to locate relevant studies (see Multimedia Appendix 1 for a detailed search strategy). We developed a comprehensive search strategy combining major subject headings and free texts, and their thesaurus, plural forms, and spellings in collaboration with an experienced university research librarian. Other relevant studies were also identified and included through reference checking and citation tracking.

Screening

All relevant articles identified in our search strategy underwent 2-stage screening process: title and abstract screening and full-text screening. The Evidence for Policy and Practice Information reviewer software (version 4; Evidence for Policy and Practice Information and Co-ordinating Centre) was used to facilitate the screening process. The articles were screened against the inclusion and exclusion criteria developed by the authors (ASK, LO, HH, and KG).

Data Charting and Analysis

We reported the study characteristics, types of digital technologies investigated, and level of digital engagement under investigation with numbers and percentages using frequency tables and charts. The factors influencing older adults' digital engagement reported in the primary studies were extracted and charted. The summary findings, barriers, and facilitators identified from each included study were uploaded to the NVivo (version 12; QSR International). Conventional content analysis was used to determine the presence of certain sentence fragments, words, themes, or concepts in the text and coded into conceptually congruent categories [27]. This was done by reading the charts from individual primary studies and coding them line-by-line into tentative themes. We followed an iterative process (reading, coding, and revisiting the codes) to establish interconnections among the resultant elements and categorized them into COM-B and TDF constructs [28].

Results

Description of Included Studies

Of the total 11,412 articles identified from our search results, 1856 (16.26%) duplicates were removed. In total, 1141 (11.94%) full-text articles were obtained by screening the title and abstracts of 9556 records. Finally, 8.41% (96/1141) of articles were included in the review by assessing 1141 full-text articles against eligibility criteria. The main reasons for exclusion were non-English studies, published before 2005, mean age of the study participants <65 years, and studies with insufficient information on older adults' digital engagement perspective (see Figure 2 that shows the PRISMA-ScR flow diagram for details of screening and eligible articles).

Most (61/96, 64%) of the studies were published in the last 5 years (between 2016 and 2021), and 28% (27/96) were published between 2010 and 2015. Geographically, most of the literature was from North America (43/96, 45%), followed by Europe (32/96, 33%), Asia (13/96, 14%), and Australia (8/96, 8%). Methodologically, 47% (45/96) of studies used a qualitative method, 36% (35/96) quantitative methods, and 17% (16/96) mixed methods. Table 1 summarizes the characteristics of the articles included in this review (see Multimedia Appendix 2 [29-124] for details on the characteristics of the extracted studies).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

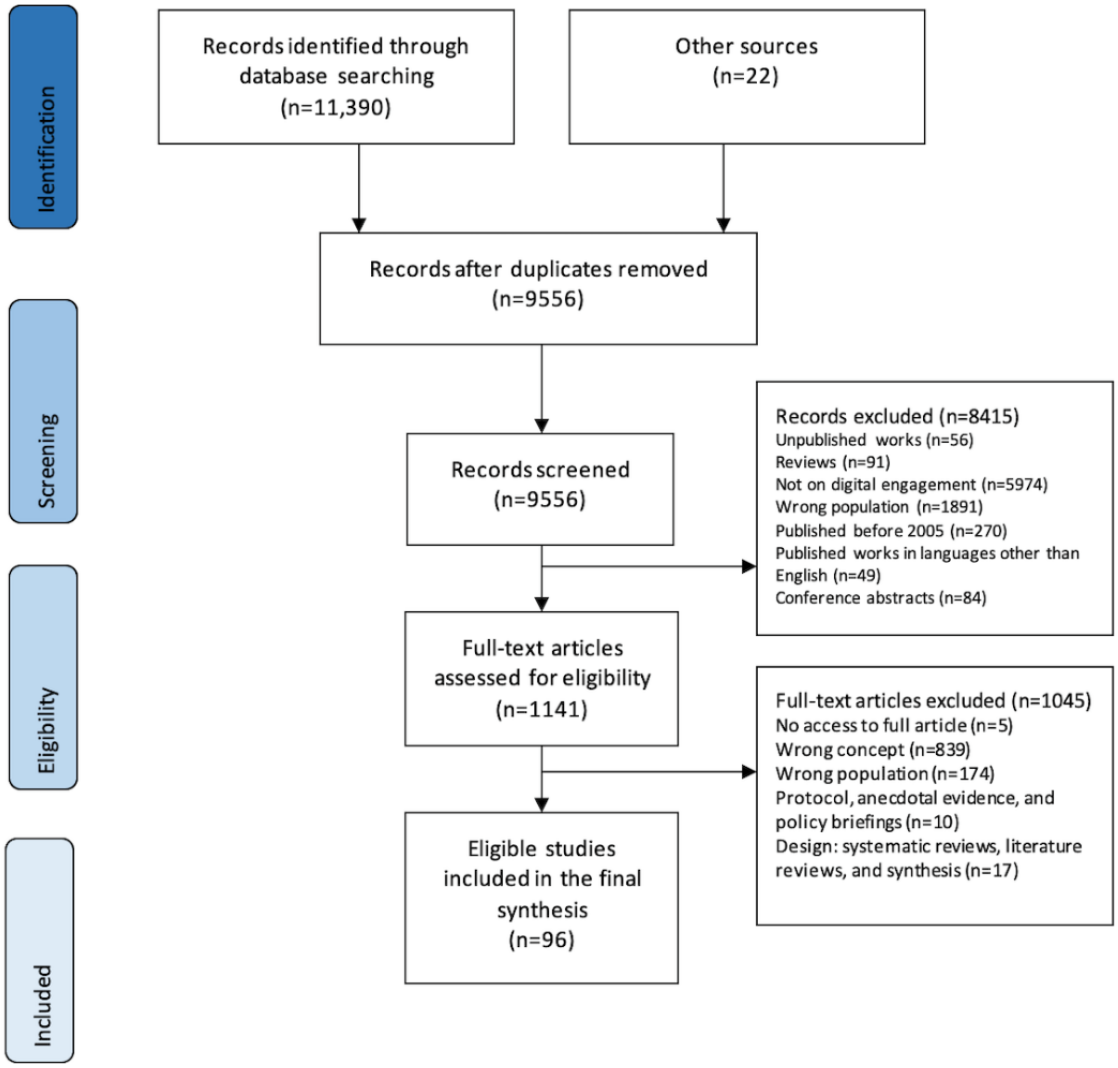


Table 1. Characteristics of included papers (n=96).

Key characteristics	Studies, n (%)
Year of publication	
2005-2010	8 (8.4)
2011-2015	27 (28)
2016-2021	61 (64)
Study settings[a]	
North America	43 (45)
Europe	32 (33)
Australia	8 (8)
Asia	13 (14)
Others	2 (2)
Study design	
Qualitative	45 (47)
Quantitative	35 (36)
Mixed method	16 (17)

[a]Two studies were cross-continental.

Digital Technology Engagement

Table 2 presents the specific types of digital technologies that were investigated. Most (54/96, 56%) of the studies investigated everyday digital technologies. Among these, 26% (14/54) of studies investigated multiple technologies, followed by the internet (10/54, 19%), information communication technologies (10/54, 19%), social networking sites (7/54, 13%), and computers (6/54, 11%). The rest 44% (42/96) of the studies, were on remote care or assistive technologies. In this category, telehealth or telecare and robots were investigated in 24% (10/42) studies. Furthermore, remote monitoring technologies, tracking technologies, mobile health, and eHealth were investigated in 7% (3/42) of studies.

Most (57/96, 59%) of the articles investigated the initial adoption stage of the digital engagement, followed by sustained digital engagement (13/96, 14%). Only 2% (2/96) of the articles studied digital technology nonuse. A significant number of studies (24/96, 25%) investigated >1 or all engagement levels (Table 3).

Table 2. Types of digital technology studied (n=96).

Types of digital technology	Studies, n (%)
Everyday technologies (n=54)	
Mobile phones	2 (2)
Gaming technologies	5 (5)
Computers	6 (6)
Social networking sites	7 (7)
Internet or ICT[a]	20 (21)
Multiple technologies	14 (15)
Remote or Assistive technologies (n=42)	
Gerontechnology	1 (1)
Assistive devices	3 (3)
mHealth[b]	3 (3)
Tracking technology	5 (5)
Remote monitoring	7 (7)
eHealth	3 (3)
Robots	10 (10)
Telehealth or telecare	10 (10)

[a]ICT: information and communication technology.

[b]mHealth: mobile health.

Table 3. Digital engagement level studied (n=96).

Levels of digital engagement	Studies, n (%)
Initial adoption	57 (59)
Sustained engagement	13 (14)
Nonuse	2 (2)
Multiple engagement levels	24 (25)

Narrative Summary on the Barriers and Facilitators

Overview

A significant overlap between the barriers to and facilitators of older adults' digital technology nonuse, adoption, and sustained digital engagement was identified. Of the 96 included studies, 39% (37/96) of the articles reported environmental context and resources as barriers and facilitators, followed by beliefs about capabilities (29/96, 30%) and physical and cognitive capabilities (26/96, 27%); social influences, beliefs about consequences, and knowledge each were cited in >20 studies. We will present the narrative synthesis below using the 3 stages of the engagement continuum and finally summarize the barriers and facilitators identified using the COM-B and TDF framework models (see Table 4 for the summary of barriers and facilitators).

Table 4. Summary of barriers and facilitators of older adults' digital engagement.

COM-B[a]	TDF[b] domains	Barriers	Facilitators
Physical capability and psychological capability	Skills (n=13)	• Difficulty in navigating and maintaining digital technologies [29,30] • Difficult to discover, locate, and use accessibility features [31] • Difficulty in finding information on website [32] • Lack of training and lack of digital competency and technical skills [33-35] • Mismatch between materiality and capability [33]	• Familiarity and experience [36-39] • Interpersonal dynamics and skills [40] • Skill to manipulate accessibility features [31,41]
Physical capability and psychological capability	Knowledge (n=23)	• Digital illiteracy [32,42,43] • Limited exposure to modern digital technologies [29,44] • Unaware of existing digital technology [31,45-47] • Lack of operational or technical knowledge [36,44,48,49] • Lack of instruction and assistance [50,51] • Understanding of what information the system collects and how it is communicated [52] • Language barriers [53]	• Awareness of the digital technology existence [33] • Prior knowledge [37,54] • Previous history or have heard stories of fall [55] • Adequate trainings [44,52,56-58] • Availability of written guide [48] • Knowledge of accessibility futures, for example, how to adjust font size [31]
Physical capability and psychological capability	Physical and cognitive identity (n=26)	• Old age-related perceptions of ability changes [31,44,59] • Health-related barriers [39,50,60-62] • Reduced sensory perception or physical (impaired vision, hearing, and dexterity) and cognitive limitations (memory loss and forgetfulness) [29,33,36,37,39,43,44,48,49,51,53,55,63-66] • Inactive lifestyle [51]	• Higher subjective well-being [67] • Good physical functions [51,68,69] • Higher cognitive functions [70]
Reflective motivation	Beliefs about capabilities (n=29)	• Perceived difficulty [71] • Inability to upgrade software [53] • Inability to attach wearable chips [29] • Perceived lack of digital technology competence [34] • Performance or effort expectancy [72] • Lack of confidence and self-efficacy [37,43,73-75]	• Positive attitude to oneself [44] • Willingness to learn or adopt technology [36,50,76] • Use of digital technologies at work [77] • Self-efficacy, self-confidence, and self-esteem [39,44,65,72,78-81] • Higher educational status [68,69,82,83] • Perceived ease of use [36,38,39,78,84-86]
Reflective motivation	Optimism (n=21)	• Comparison oneself with younger generation and feeling of inadequacy [47,50] • Failing to meet perceived need or lack of relevance [40,45,87,88] • Aversion and limited or lack of interest [37,43-45,51] • Pre-established negative attitudes [34,56,89] • Technophobia [32]	• Technological optimism [90,91] • Perceived digital technology benefits [43,84,90,92-95] • Positive technological experience [37] • Availability of need-based trainings [93] • Curiosity [37] • Enthusiastic attitude [91]
Reflective motivation	Beliefs about consequence (n=24)	• Intrusiveness: privacy [34,44,61,63,74,96-100], safety [32,45], and security concerns [37,43,73] • Mistrust [54,64] • Perceived lack of benefits [101] • Lack of reliability and uncertainty about the reliability [32,66,85,87] • Lack of accountability related to remote care technologies [32] • Fear of addiction or habit forming nature especially with internet-based digital technologies [64,102]	• Ability to regulate internet identity [96] • Interactive features that give timely and tailored feedback [101] • Reduced isolation or connectedness [61,76] • Ability to monitor health [87,88] • Positive health-seeking behavior [37]
Reflective motivation	Intention (n=1)	—[c]	• Higher intentions to use digital technologies [84]

COM-B[a]	TDF[b] domains	Barriers	Facilitators
Reflective motivation	Goals (n=9)	• Preference to spend time on family and other valuable activities [103]	• Independence and sense of autonomy [55,56,102] • Perceived playfulness and the fun associated with digital technology [38,92] • Goal-monitoring ability [85] • Sense of connection or connectedness and interaction [104] • Way of keeping in touch with family and friends [74]
Automatic motivation	Reinforcement (n=13)	• Poor instructions [51,105] • Preference for inactive lifestyle at old age (satisfied with current activity performance) [51]	• Convenience: technologies which makes activities easier and faster [32,40] • Received a tailored and personalized support and trainings [39,43,44,63,68,76] • Safe learning environment (accessible, appropriately placed, inclusive, one-to-one and personalized support) [76,81] • Technologies that can be customized to older adults needs, abilities and preferences [33,76] • User satisfactions [106]
Automatic motivation	Emotion (n=15)	• Fear and frustration from digital technologies complexity [43,44,47,62,64,71,73,87] • Fear of withdrawal from face-to-face input from their physician [80,107] • Fear owing to lack of knowledge [36] • Lack of emotional reciprocity [108] • Digital shopping assistant with digital assistant style or task oriented or formal [109]	• Digital shopping assistant with social assistant style or reciprocity, conversational [109] • Mismatched appearance vs robot attributes such as voice and facial expressions [110] • Robots with certain enjoyment and attractiveness [110] • Enjoyable games [78]
Physical opportunity and social opportunity	Environmental context and resources (n=37)	• Technological factors • Perceived or actual complexity of technology [30,41,44,45,85] • Lack of user friendliness [75] • Technologies without adaptive design features [44] • Poorly designed user interfaces [36] • Having to charge devices many times (battery life) [55] • Poor output quality [77], poor video and audio quality [111], small size of icons and texts [36], and color [53] • Device malfunction and slow and repeated freezing [29,45,48,112] • Require captcha [41] • Relentless pace of digital technology development [66] • Suboptimal performance [75] • Inaccurate measurement and technologies with non-standard scales [75,113] • Lack of technological aesthetic values, for example, wearables [45] • Environmental factors • Physical infrastructure access [54] • Economic barriers and financial limitation [30] • Cost: direct [36,37,42,44-46,51,53,63,66,67,69,73,84,101,114] and opportunistic cost associated with technologies, electrical consumption [115], and cost related to maintenance [100]	• Technological factors • Ease of use and simplicity [32,40,63] • Simple log procedure [85] • Quality of outputs (quality videos, audios, and text) [77] • Waterproof [51] • Sleep-tracking ability [51] • Touch screen [38] • Connectivity [40] • Audible feedback [36,66] • Automated call [55] • Large icon and display [36] • Instant feedback [36] • Alarms and reminder future [49] • Accessibility features such as font adjustment [76] • Remote technologies integrated within mainstream technologies, for example, fall detection devices integrated with cell phones [55] • Environmental factors • Older adults' digital technology ownership (owning computer, smartphone, broadband etc) [116] • Free of charge, financial incentives [77]; affordable [55,117]; provided through existing financial schemes (eg, insurance) [55]

COM-B[a]	TDF[b] domains	Barriers	Facilitators
Physical opportunity and social opportunity	• Social influences (n=25)	• Perceived isolation or helplessness [55]; loss of social contact [47]; living alone [55,68]; lack of social assistance [44,47,74,82] • Digital alienation and social disapproval [98] • Negative learning experience (isolating and insulting learning environment; facilitators' judgemental attitudes [76] • Stigma from wearing wearables (alarm going in public) [29,45,55,98] • Perceptions of prejudice and discrimination or stigma from sense of powerlessness and dependency [44,78,90,98] • Care through intergenerational support [57] • Cultural expectations (mothers do not call; instead, children have to call) [40] • Cold and shallow forms for digital communications for gossip and self-obsessiveness [34]	• Digital kinship and maintaining social connection [88,107] • Formal or informal social engagements [79] • Peer or family support availability [41,51,66,80,93,96] • Having someone around to help in the house [55] • Encouragement and recommendation by physicians or nurses to use digital technology [73,77,107]

[a]COM-B: capability, opportunity, and motivation behavior change model.

[b]TDF: theoretical domain framework.

[c]Not available.

Technological Nonuse

There is a research gap regarding technological nonuse. Only 2 studies have investigated the determinants of technological nonuse among older adults as a primary outcome [34,103]. The remaining studies investigated nonuse as a secondary outcome or as a comparator to use. Older adults' motivation and attitude play a significant role in their decision to reject digital technology. A study reported nonuse among older adults as justified rejection [103]. Some of these justifications were based on value judgments and the inability to foresee the relevance of the technology [33,45,103].

The perception of old age as an identity, that is, not identifying oneself as an old person was indicated as a reason to disengage, particularly from technologies designed for this specific demographic group [103]. For example, wearables such as fall detection devices and remote trackers, can comport a sense of dependency. Furthermore, having an unfavorable attitude toward digital technology formed by past personal experiences of privacy and safety concerns were found to be important factors for technological nonuse [33,89,102]. The lack of meaningful involvement in decision-making regarding use, data, privacy, and security contributed to the nonuse of digital technology. For example, studies on remote monitoring technology have indicated that users are not well informed about how and by whom their data will be handled [52,111].

Initial Adoption

Physical and cognitive capability changes have been reported to influence older adults' initial adoption of digital technology. These changes include reduced or loss of sensory perception (visual and hearing), impaired dexterity, and impaired cognitive function [29,33,36,37,39,43,44,48,49,51,53,55,63-66]. These changes cause a mismatch between the capabilities and materiality of technology. Meanwhile, greater subjective well-being, "good" physical function, and higher cognitive function facilitate better initial technological engagement [51,67-70].

Knowledge and skills in operating digital technologies were another widely reported capability theme [29,30,34]. Familiarity with digital technologies through a work context and subsequent skill acquisition facilitate initial adoption [36-39]. Attaining digital competence among older adults was highly dependent on awareness of existing technology and availability of support and instruction [32,33,36,37,42-44,48,49,53]. Personalized training, availability of written guidelines, and opportunities for need-based learning in a safe environment have been reported to facilitate skill acquisition and initial digital engagement [44,48,71,76,85]. A safe environment for learning characterized by accessible, appropriately placed, inclusive, one-to-one, personalized support geared toward one's ability and preference empowered and facilitated digital technology adoption by older adults [50,51,76,81]. Discouraging learning environment characterized by features such as judgmental delivery, isolating, and insulting impersonalized, fast-paced, and incomprehensible jargons were reported as barriers [76].

Studies have reported technological features that are unmatched with older adults' physical capabilities as barriers to digital engagement. Some of the mismatches include poor sound quality and impaired hearing, small text font or icons size and impaired vision, and difficulty maneuvering buttons, and deteriorating dexterity [55]. These factors were found to be particularly significant in speech- and alarm-based technologies, such as fall detection devices and remote monitoring technologies [55,93]. Poorly designed user interfaces that are difficult to interact with due to the requirement of several factor authentications and inputs, slow and freezing, [30,36,48] poor connectivity [112], and lack of notification system [30] were identified as barriers to digital engagement among older adults. By contrast, simple login procedures, accessible, customizable and easy access technologies, including large displays, touch screens, high-definition sound and pictures, high-quality outputs and the ability to give printouts to facilitate engagement

[31,38,41,84,85,113,116]. Automated technologies with instant feedback and interactive features; and the ability to track performance were received more favorably [55,84,90,101].

Few studies have discussed the peculiar features of the technologies used in health and social care settings. Lack of communication support among the users, technology, and health care providers; biomedical parameters including vital signs presented on nonstandard scales; and lack of professional interpretation of those parameters were reported as barriers to digital engagement [113]. False alarms from fall detection devices and remote monitoring technologies and the associated stigma and discrimination from wearing wearables were also mentioned as barriers [29,45,55,118].

Social influence, recommendation and support from relations, plays a pivotal role in digital technology adoption among older adults. A recommendation received from someone trusted, for example, doctors, nurses, and family members, influenced older adults' intention to adopt or reject digital technology [36,84]. Furthermore, the constant support with technical difficulty by having someone around was found to facilitate internet and social network technology adoption [68,106,116]. Studies on assistive technology have reported that perceived isolation or lack of companionship or living alone increases acceptance [33,55]. Interpersonal skills facilitate greater engagement in web-based communication [40].

Older adults' attitudinal factors toward digital technologies, such as perceived difficulty, self-efficacy, and benefits were important motivation-related determinants [34,56,89]. The perception that digital technology is not appropriate for older adults was reported to be a barrier to engagement [47,50,71]. Lack of confidence and interest, aversion and skepticism toward digital technology, and lack of relevance or necessity to adopt digital technology were salient barriers that hinder older adults' motivation to engage digitally [37,43,73-75]. Awareness of the perceived benefits such as expedited health care [63], information that allows for goal setting and goal monitoring [85], the opportunity for self-development (skills, esteem, and identity) [78,119], previous history of fall [55], improved task performance [44], and social connectedness [103] were among the main motivational reasons for older adults to digitally engage.

The fear of digital technology intrusiveness was cited several times as a barrier to adopting digital technology [34,44,61,63,74,96-99]. Safety concerns, security, and mistrust are common reasons for digitally disengaging, particularly, in web-based digital technologies [44,50,63,96,111]. In addition, fear of web-based scammers or impersonators was identified as a salient barrier to digital engagement [64]. Furthermore, fear and frustration from the amount of distraction from repetitive and redundant adverts was mentioned [96]. Studies have shown older adults' preference for social interaction with value (eg, intentional and meaningful activities, such as family or exercise) instead of web-based interactions with extended social network [34,45].

Sustained Digital Technology Engagement

There were many commonalities between the barrier and facilitator themes on digital technology adoption and sustained digital engagement [45,52,102,111]. Technological features that are simple and customizable to older adults' needs facilitate sustainable, better and prolonged engagement [38]. Features that require multiple inputs and multi-factor authentication that could be inaccessible to older adults discourage sustained engagement [45,48,76,96]. High output quality of digital technology, such as voice, picture, sound, and other outputs, was found to be equally necessary for sustained engagement [53]. For web-based technologies, slow and freezing interfaces led to dissatisfaction and frustration [48].

Sustained use, according to many studies, was highly dependent on the perceived self-efficacy of individuals [81]. Confidence was affected by knowledge of the technology, experience and familiarity, and willingness and ability to learn [37,68,79,120]. Studies also reported technologies addictive features and repetitive distractions were among the barriers to long term technology use [45,96,102]. Safety concerns, security, and mistrust are common privacy issues associated with web-based digital technologies [50,96].

Discussion

Principal Findings

This scoping review provides a synthesis of the literature on older adults' experiences and facilitators of and barriers to digital engagement. We conceptualized digital technology engagement as a three-stage continuum (nonuse, initial adoption, and sustained use) to capture the entire range of individuals' experiences from technology abandonment and acceptance to actual and continued use. A process predicated on ongoing negotiations or renegotiations and nonlinear progression between stages. Our review included 96 primary studies exploring a range of everyday and remote care technologies and demonstrated the complexity and multiple intertwined factors at personal, sociocultural, and environmental levels influencing digital engagement among older adults. We mapped these factors over the COM-B and TDF behavioral change models to facilitate articulation and provide a basis for future interventions that improve digital engagement among older adults. Environmental context and resources, beliefs about capabilities, and physical and cognitive capabilities were the most cited factors across the engagement stages. There is little research on the nonuse and sustained-use stages, as most studies in our review investigated the initial adoption stage of digital engagement.

Comparison With Prior Works

One central theme across engagement stages was older adults' digital knowledge and skill capabilities [29,30]. Over the past years, older adults' digital competence, access to digital technology, and interest in further education have significantly improved [5,125]. However, a significant proportion of older adults have insufficient or lack the required digital skills. For example, only 1 in 4 European older adults have basic digital skills [126]. According to the European Union (EU) digital competence framework, digital literacy comprises 5 indicators:

information and data literacy, communication and collaboration, digital content creation, safety, and problem-solving [127]. Such guidelines with broader definitions and detailed outlines of digital skills could help guide the development of curricula to equip older adults with essential basic digital skills. Innovative and interactive practical learning delivery modalities, for example, web-based learning and digital games, could be used [121]. Although older adults' digital skills reflect familiarity and varying levels of exposure through education or work contexts in the past, the changing requirements related to capability, as well as rapid technological development, necessitate continued training and support.

Not surprisingly, the costs of procuring and maintaining technology and indirect costs (eg, electricity consumption) were cited several times as barriers to older adults' digital engagement [36,44,67,73,84,101]. This aligns with previous reviews that low income predicts low technology ownership and low access to quality support and digital engagement in general [128,129]. For example, Choi et al [130] reported a strong correlation between discontinuing internet use and low income among homebound older adults. "Digital poverty," that is, inability to fully use available digital platforms owing to lack of finance, access (eg, geographic exclusion) and lack of skill, is a growing practical and policy concerns even among economically developed countries. According to the recent report from the United Kingdom House of Commons, a significantly lower proportion of households with income between £6000 and £10,000 "GBP £1 (US $1.42) have home internet access compared with those households who earn £40,000 and above (51% vs 99%); this divide has even worsened during the COVID-19 pandemic with the increasing hybrid ways of coping [131].

Our review demonstrated that the usability of technology is highly dependent on its material features (physical property, functionality, and interoperability). Previous studies have reported that older adults find it cumbersome when technologies have multiple buttons, multi-factor authentications, poor quality user interfaces, and outputs [30,41,44,45,85]. These difficulties could emanate from the inherent complexity of technologies, design failures, or lack of necessary training and skill sets to operate technology. However, it is noteworthy to understand the extreme heterogeneity in older users' experience, background, and diversity of applications and to take a precautionary approach when making a technological design recommendation based on barriers and facilitator studies. Continued efforts to strike a balance between usable, enjoyable, and secure technologies through value-based design ethos that considers older adults' physical, psychological, and contextual needs must be promoted. This includes accessibility features that allow older adults to customize technology according to their needs.

We found that fear of safety and invasion of privacy were barriers to digital engagement and a growing concern among older adults, regulatory bodies, and researchers [52]. This was in line with previous findings on the growing digital distrust and apprehension among users owing to technology intrusiveness; increased web-based activities; use of personal data for health and financial reasons; increasing number of data breaches; data monetization; and lack of transparency on why, how, and by whom data will be handled [52,111]. Privacy regulations such as the EU General Data Protection Regulation have improved the privacy accountability of suppliers and raised users' awareness of their privacy rights [132]. However, older adults' awareness and proactive prevention of personal data are significantly lower than those of their younger counterparts [133]. Addressing these serious concerns requires cross-cutting interventions that ensures older adults' empowerment, simultaneously strengthening legal frameworks and institutions and cross-sectoral partnerships. For example, regulatory bodies need to capitalize on and constantly update existing privacy regulations to enforce and protect individuals. In addition, concerned stakeholders need to provide continuous education on safety, privacy rights, and regulations that will improve the older adults' privacy efficacy, privacy concerns, and trust of older adults. Businesses and service providers also need to play their part in implementing privacy regulations, establishing a clear communication protocol and transparency. Although this will primarily benefit users, recent reports have indicated that firms with effective privacy protection systems have a significantly higher return on investment; "beyond meeting compliance requirement-good privacy is indeed good for business and individuals" [134].

Digital technology takes on multiple explicit and implicit meanings for its users. In line with previous studies, our review demonstrated the role of technology in promoting active and independent living and enhanced personal autonomy, power, and control [96,113,119]. However, technology could also imply a sense of dependency and decline contrary to the primary purpose of promoting independence [28]. For example, studies have reported assistive technologies symbolizing an image of "being old" opposite to the desired or ideal self-image perceived by older adults and could be associated with agist stigma and discrimination. This apparent latent tension between the individual's identity and perception of society aligns with the mainstream identity theory that describes the role of self-image and the perception of others in individuals' decisions [135]. These symbolic properties and their influence on the adoption or rejection of technology need further research.

The COM-B and TDF mapping in our review ensured that a wide range of emergent determinants were explored. These comprehensive frameworks cover intrinsic factors pertaining to individuals' abilities and motivations and extrinsic factors related to social, technological, and environmental factors. These factors can be used by researchers, technology developers, caregivers, and program implementors to inform the development of implementation models for optimal digital engagement among older adults. Previous studies have given a tremendous emphasis on the individual motivational aspect of behavior, including beliefs about consequences and beliefs about capabilities, such as perceived usefulness and ease of use [42,136]. These themes have been widely explored in previous technological acceptance models and theories and have attracted considerable interest for research [44,80,84,92]. However, looking beyond motivation and addressing all other moderating factors are required to close the digital divide between age groups.

Future Directions

This review highlights several areas that require further research. First, research needs to move beyond the prevailing focus of the classic technological acceptance models and theories on initial adoption and individual motivational factors. Accordingly, conceptualizing digital engagement as a continuum instead of a one-time decision could help understand individuals' journeys holistically, the impacts of disengagement on well-being, and how it marginalizes older adults. Second, there is a need for a standard definition and validated measuring tools for the nonuse and sustained-use stages of digital engagement. Third, theorizing older adults' digital technology nonadoption, uptake, and continued engagement using in-depth and contextually situated methodologies is needed. Such theorizing from older adults' experiential accounts could help illuminate the meaning of technology; the interaction between the material and symbolic properties of technology; digital engagement meanings on identity, interpersonal relations, capabilities, motivations, affect, and emotions; and how all these influence adoption, decisions, dignity, and well-being.

Strengths and Limitations

This review followed a systematic approach to review evidence on digital technology engagement of older adults, which included identifying review questions, comprehensive search across all major databases for technology and health, application of explicit inclusion and exclusion criteria, use of a systematic and structured theoretical framework to map out the evidence, and synthesis of the findings. However, this review has several limitations. First, only studies published in English were included, and most of the studies were from North America (43/96, 45%) and Europe (32/96, 33%), which might affect the transferability and coverage of the identified literature. Second, in this review, we sought to understand the barriers to and facilitators of overall digital engagement instead of technology-specific engagements and might have missed the nuanced variations. Third, the sustained digital engagement stage is not a well-defined research outcome and a consensus on its definition has yet to be reached. To ascertain what studies could fall under this category, we followed either the description in the primary studies if they explicitly declared that they were investigating the sustainability of digital use, or we ascertained through careful reading of the description of the study to see whether sustained use over a significant time was one of the objectives or implicated in the study. Finally, we did not hold a formal stakeholder consultation because of time and resource constraints. Instead, the findings of this review have been widely discussed in a consortium and many conferences and other informal gatherings involving older adults.

Conclusions

The digital engagement of older adults can be conceptualized as a three-stage continuum (nonuse, initial adoption, and sustained use) and a negotiated process possessing acceptance, rejection, and temporal characteristics. Little research has been conducted on nonuse and sustained engagement stages. Most studies in this review investigated the initial adoption stage. Using the COM-B and TDF models enabled us to identify a wide range of salient intrinsic and extrinsic determinants across engagement stages. Considering the barriers identified, including but not limited to the changing capability requirements, cost, access to technology, safety and privacy concerns, and design stereotypes and assumptions, could improve older adults' digital experiences, facilitate better digital engagement, and optimize future digital interventions and scale-up. Furthermore, empowering older adults with digital skills through a learner-centered approach on a need-to-know basis should be promoted. Future research aimed at understanding older adults' everyday world of experience, the meaning of digital technologies, and how they cope with this fast-paced digital development is critical for promoting meaningful digital engagement. The range of contexts and values, which older adults avoid, adopt, or continue to use, in digital technology and standardized tools that measure these outcomes require further research.

Acknowledgments

This project received funding from the European Union's Horizon 2020 research and innovation program under grant 813928.

Authors' Contributions

ASK developed the research questions, study methods, and inclusion and exclusion criteria. KG refined the research questions and helped develop the research and study methods. ASK, KG, L-LO, and HH reviewed the papers and meaningfully contributed to the drafting and editing of the manuscript. The final version of the manuscript has been read and approved by all the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOC File , 39 KB - jmir_v24i12e40192_app1.doc]

Multimedia Appendix 2
Characteristics of included studies.

[DOCX File , 66 KB - jmir_v24i12e40192_app2.docx]

References

1. Organisation for Economic Co-operation and Development. Health at a Glance 2017: OECD Indicators. Paris, France: OECD Publishing; Nov 10, 2017.
2. World Population Ageing: Highlights 2019. Department of Economic and Social Affairs, United Nations. New York, NY, USA: United Nations; 2020. URL: https://tinyurl.com/3rh89ma3
3. Internet access – households and individuals, Great Britain: 2018. Office for National Statistics. 20198 Aug 7. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2018 [accessed 2022-02-19]
4. Share of adults in the United States who use the internet in 2019, by age group. Statista. 2019. URL: https://www.statista.com/statistics/266587/percentage-of-internet-users-by-age-groups-in-the-us/ [accessed 2021-07-26]
5. Anderson M. Mobile Technology and Home Broadband 2019. Pew Research Center. 2019 Jun 13. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2021-03-10]
6. Secretary-general's roadmap for digital cooperation. United Nations. 2021. URL: https://www.un.org/en/content/digital-cooperation-roadmap/ [accessed 2022-01-01]
7. Andone I, Błaszkiewicz K, Eibes M, Trendafilov B, Montag C, Markowetz A. How age and gender affect smartphone usage. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 2016 Presented at: UbiComp '16; September 12-16, 2016; Heidelberg, Germany p. 9-12. [doi: 10.1145/2968219.2971451]
8. 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]
9. Pirhonen J, Lolich L, Tuominen K, Jolanki O, Timonen V. “These devices have not been made for older people's needs” – older adults' perceptions of digital technologies in Finland and Ireland. Technol Soc 2020 Aug;62:101287. [doi: 10.1016/j.techsoc.2020.101287]
10. Whitelaw S, Mamas MA, Topol E, Van Spall HG. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
11. Nguyen MH, Gruber J, Fuchs J, Marler W, Hunsaker A, Hargittai E. Changes in digital communication during the COVID-19 global pandemic: implications for digital inequality and future research. Soc Media Soc 2020 Jul;6(3):2056305120948255 [FREE Full text] [doi: 10.1177/2056305120948255] [Medline: 34192039]
12. Litchfield I, Shukla D, Greenfield S. Impact of COVID-19 on the digital divide: a rapid review. BMJ Open 2021 Oct 12;11(10):e053440 [FREE Full text] [doi: 10.1136/bmjopen-2021-053440] [Medline: 34642200]
13. Damant J, Knapp M, Freddolino P, Lombard D. Effects of digital engagement on the quality of life of older people. Health Soc Care Community 2017 Nov;25(6):1679-1703. [doi: 10.1111/hsc.12335] [Medline: 26919220]
14. Kuerbis A, Mulliken A, Muench F, Moore AA, Gardner D. Older adults and mobile technology: factors that enhance and inhibit utilization in the context of behavioral health. Ment Health Addict Res 2017;2(2):1-11. [doi: 10.15761/mhar.1000136]
15. Davidson S. Digital Inclusion Evidence Review 2018. ageUK. 2018 Nov. URL: https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/age_uk_digital_inclusion_evidence_review_2018.pdf [accessed 2020-11-05]
16. 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]
17. Karapanos E, Zimmerman J, Forlizzi J, Martens JB. Measuring the dynamics of remembered experience over time. Interact Comput 2010 Sep;22(5):328-335. [doi: 10.1016/j.intcom.2010.04.003]
18. Nadal C, Sas C, Doherty G. Technology acceptance in mobile health: scoping review of definitions, models, and measurement. J Med Internet Res 2020 Jul 06;22(7):e17256 [FREE Full text] [doi: 10.2196/17256] [Medline: 32628122]
19. Kebede AS, Ozolins LL, Holst H, Galvin K. The digital engagement of older people: systematic scoping review protocol. JMIR Res Protoc 2021 Jul 05;10(7):e25616 [FREE Full text] [doi: 10.2196/25616] [Medline: 36260392]
20. Statement: Delivering the Broadband Universal Service. Ofcom. 2019. URL: https://www.ofcom.org.uk/consultations-and-statements/category-1/delivering-broadband-universal-service [accessed 2022-05-20]
21. 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]
22. Richardson M, Khouja CL, Sutcliffe K, Thomas J. Using the theoretical domains framework and the behavioural change wheel in an overarching synthesis of systematic reviews. BMJ Open 2019 Jun 22;9(6):e024950 [FREE Full text] [doi: 10.1136/bmjopen-2018-024950] [Medline: 31229999]
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 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

24. 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]
25. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
27. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
28. 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]
29. Demiris G, Chaudhuri S, Thompson HJ. Older adults' experience with a novel fall detection device. Telemed J E Health 2016 Sep;22(9):726-732 [FREE Full text] [doi: 10.1089/tmj.2015.0218] [Medline: 26959299]
30. Baker S, Warburton J, Hodgkin S, Pascal J. The supportive network: rural disadvantaged older people and ICT. In: Ageing and Society. 2016 Apr 07 Presented at: ITAP '16; October 20-22, 2016; Kochi, Japan p. 1291-1309. [doi: 10.1017/s0144686x16000350]
31. Franz RL, Wobbrock JO, Cheng Y, Findlater L. Perception and adoption of mobile accessibility features by older adults experiencing ability changes. In: Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility. 2019 Oct Presented at: ASSETS '19; October 28-30, 2019; Pittsburgh, PA, USA p. 267-278. [doi: 10.1145/3308561.3353780]
32. Choudrie J, Ghinea G, Songonuga VN. Silver surfers, e-government and the digital divide: an exploratory study of UK local authority websites and older citizens. Interact Comput 2013 Feb 06;25(6):417-442. [doi: 10.1093/iwc/iws020]
33. Greenhalgh T, Wherton J, Sugarhood P, Hinder S, Procter R, Stones R. What matters to older people with assisted living needs? A phenomenological analysis of the use and non-use of telehealth and telecare. Soc Sci Med 2013 Sep;93:86-94 [FREE Full text] [doi: 10.1016/j.socscimed.2013.05.036] [Medline: 23906125]
34. Lüders M, Brandtzæg PB. 'My children tell me it's so simple': a mixed-methods approach to understand older non-users' perceptions of Social Networking Sites. New Media Soc 2016 Jul 09;19(2):181-198. [doi: 10.1177/1461444814554064]
35. Nguyen MH, Hunsaker A, Hargittai E. Older adults' online social engagement and social capital: the moderating role of Internet skills. Inf Commun Soc 2022;25(7):942-958. [doi: 10.1080/1369118x.2020.1804980]
36. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382 [FREE Full text] [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
37. Nymberg VM, Bolmsjö BB, Wolff M, Calling S, Gerward S, Sandberg M. 'Having to learn this so late in our lives…' Swedish elderly patients' beliefs, experiences, attitudes and expectations of e-health in primary health care. Scand J Prim Health Care 2019 Mar;37(1):41-52 [FREE Full text] [doi: 10.1080/02813432.2019.1570612] [Medline: 30732519]
38. Astell AJ, Joddrell P, Groenewoud H, de Lange J, Goumans M, Cordia A, et al. Does familiarity affect the enjoyment of touchscreen games for people with dementia? Int J Med Inform 2016 Jul;91:e1-e8. [doi: 10.1016/j.ijmedinf.2016.02.001] [Medline: 26897552]
39. Leone C, Lim JS, Stern A, Charles J, Black S, Baecker R. Communication technology adoption among older adult veterans: the interplay of social and cognitive factors. Aging Ment Health 2018 Dec;22(12):1666-1677. [doi: 10.1080/13607863.2017.1381946] [Medline: 29019416]
40. Barbosa Neves B, Franz R, Judges R, Beermann C, Baecker R. Can digital technology enhance social connectedness among older adults? A feasibility study. J Appl Gerontol 2019 Jan;38(1):49-72. [doi: 10.1177/0733464817741369] [Medline: 29166818]
41. Ferreira SM, Sayago S, Blat J. Going beyond telecenters to foster the digital inclusion of older people in Brazil: lessons learned from a rapid ethnographical study. Inf Technol Dev 2016 Jun 20;22(sup1):26-46. [doi: 10.1080/02681102.2015.1091974]
42. Chung JE, Park N, Wang H, Fulk J, McLaughlin M. Age differences in perceptions of online community participation among non-users: an extension of the Technology Acceptance Model. Comput Human Behav 2010 Nov;26(6):1674-1684. [doi: 10.1016/j.chb.2010.06.016]
43. Hill R, Betts LR, Gardner SE. Older adults' experiences and perceptions of digital technology: (dis)empowerment, wellbeing, and inclusion. Comput Human Behav 2015 Jul;48:415-423. [doi: 10.1016/j.chb.2015.01.062]
44. Chen K, Chan AH. Use or non-use of gerontechnology--a qualitative study. Int J Environ Res Public Health 2013 Sep 30;10(10):4645-4666 [FREE Full text] [doi: 10.3390/ijerph10104645] [Medline: 24084674]
45. Orellano-Colón EM, Mann WC, Rivero M, Torres M, Jutai J, Santiago A, et al. Hispanic older adult's perceptions of personal, contextual and technology-related barriers for using assistive technology devices. J Racial Ethn Health Disparities 2016 Dec;3(4):676-686 [FREE Full text] [doi: 10.1007/s40615-015-0186-8] [Medline: 27294762]

46. 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]
47. Heinz M, Martin P, Margrett JA, Yearns M, Franke W, Yang H, et al. Perceptions of technology among older adults. J Gerontol Nurs 2013 Jan;39(1):42-51. [doi: 10.3928/00989134-20121204-04] [Medline: 23244061]
48. Damodaran L, Olphert CW, Sandhu J. Falling off the bandwagon? Exploring the challenges to sustained digital engagement by older people. Gerontology 2014;60(2):163-173 [FREE Full text] [doi: 10.1159/000357431] [Medline: 24356488]
49. Gitlow L. Technology use by older adults and barriers to using technology. Phys Occup Ther Geriatr 2014 Aug 12;32(3):271-280. [doi: 10.3109/02703181.2014.946640]
50. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
51. Kononova A, Li L, Kamp K, Bowen M, Rikard RV, Cotten S, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]
52. Berridge C. Breathing room in monitored space: the impact of passive monitoring technology on privacy in independent living. Gerontologist 2016 Oct;56(5):807-816. [doi: 10.1093/geront/gnv034] [Medline: 26035900]
53. Navabi N, Ghaffari F, Jannat-Alipoor Z. Older adults' attitudes and barriers toward the use of mobile phones. Clin Interv Aging 2016 Sep 30;11:1371-1378 [FREE Full text] [doi: 10.2147/CIA.S112893] [Medline: 27757025]
54. Haase KR, Cosco T, Kervin L, Riadi I, O'Connell ME. Older adults' experiences with using technology for socialization during the COVID-19 pandemic: cross-sectional survey study. JMIR Aging 2021 Apr 23;4(2):e28010 [FREE Full text] [doi: 10.2196/28010] [Medline: 33739929]
55. Chaudhuri S, Kneale L, Le T, Phelan E, Rosenberg D, Thompson H, et al. Older adults' perceptions of fall detection devices. J Appl Gerontol 2017 Aug;36(8):915-930 [FREE Full text] [doi: 10.1177/0733464815591211] [Medline: 26112030]
56. Hernández-Encuentra E, Pousada M, Gómez-Zúñiga B. ICT and older people: beyond usability. Educ Gerontol 2009 Feb 04;35(3):226-245. [doi: 10.1080/03601270802466934]
57. Berridge C, Chan KT, Choi Y. ensor-based passive remote monitoring and discordant values: qualitative study of the experiences of low-income immigrant elders in the United States. JMIR Mhealth Uhealth 2019 Mar 25;7(3):e11516 [FREE Full text] [doi: 10.2196/11516] [Medline: 30907741]
58. Rolandi E, Vaccaro R, Abbondanza S, Casanova G, Pettinato L, Colombo M, et al. Loneliness and social engagement in older adults based in Lombardy during the COVID-19 lockdown: the long-term effects of a course on social networking sites use. Int J Environ Res Public Health 2020 Oct 28;17(21):7912 [FREE Full text] [doi: 10.3390/ijerph17217912] [Medline: 33126634]
59. Kamin ST, Lang FR, Beyer A. Subjective technology adaptivity predicts technology use in old age. Gerontology 2017;63(4):385-392. [doi: 10.1159/000471802] [Medline: 28423371]
60. Axelsson SW, Wikman AM, Näslund A, Nyberg L. Older people's health-related ICT-use in. Gerontechnology 2013;12(1):36-43. [doi: 10.4017/gt.2013.12.1.010.00]
61. Beer JM, Takayama L. Mobile remote presence systems for older adults: acceptance, benefits, and concerns. In: Proceedings of the 6th International Conference on Human-Robot Interaction. 2011 Mar Presented at: HRI '11; March 6-9, 2011; Lausanne, Switzerland p. 19-26. [doi: 10.1145/1957656.1957665]
62. Choi NG, Dinitto DM. Internet use among older adults: association with health needs, psychological capital, and social capital. J Med Internet Res 2013 May 16;15(5):e97 [FREE Full text] [doi: 10.2196/jmir.2333] [Medline: 23681083]
63. Parker SJ, Jessel S, Richardson JE, Reid MC. Older adults are mobile too!Identifying the barriers and facilitators to older adults' use of mHealth for pain management. BMC Geriatr 2013 May 06;13:43 [FREE Full text] [doi: 10.1186/1471-2318-13-43] [Medline: 23647949]
64. Gatto SL, Tak SH. Computer, Internet, and e-mail use among older adults: benefits and barriers. Educ Gerontol 2008 Aug 22;34(9):800-811. [doi: 10.1080/03601270802243697]
65. Bell C, Fausset C, Farmer S, Nguyen J, Harley L, Fain WB. Examining social media use among older adults. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media. 2013 May Presented at: HT '13; May 1-3, 2013; Paris, France p. 158-163. [doi: 10.1145/2481492.2481509]
66. Fisher K, Easton K. The meaning and value of digital technology adoption for older adults with sight loss: a mixed methods study. Technol Disabil 2019 Apr 02;30(4):177-184. [doi: 10.3233/tad-180205]
67. Chopik WJ. The benefits of social technology use among older adults are mediated by reduced loneliness. Cyberpsychol Behav Soc Netw 2016 Sep;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
68. Berner J, Aartsen M, Deeg D. Predictors in starting and stopping Internet use between 2002 and 2012 by Dutch adults 65 years and older. Health Informatics J 2019 Sep;25(3):715-730 [FREE Full text] [doi: 10.1177/1460458217720398] [Medline: 28747085]

69. Brody BL, Field LC, Roch-Levecq A, Depp C, Edland SD, Minasyan L, et al. Computer use among patients with age-related macular degeneration. Ophthalmic Epidemiol 2012 Aug;19(4):190-195 [FREE Full text] [doi: 10.3109/09286586.2012.672618] [Medline: 22775273]
70. Berner JS, Rennemark M, Jogréus C, Berglund J. Factors associated with change in Internet usage of Swedish older adults (2004-2010). Health Informatics J 2013 Jun;19(2):152-162 [FREE Full text] [doi: 10.1177/1460458212462151] [Medline: 23715214]
71. Berkowsky RW, Cotten SR, Yost EA, Winstead VP. Attitudes towards and limitations to ICT use in assisted and independent living communities: findings from a specially-designed technological intervention. Educ Gerontol 2013 Nov 01;39(11):797-811 [FREE Full text] [doi: 10.1080/03601277.2012.734162] [Medline: 24244065]
72. de Veer AJ, Peeters JM, Brabers AE, Schellevis FG, Rademakers JJ, 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]
73. Cimperman M, Brenčič MM, Trkman P, Stanonik MD. Older adults' perceptions of home telehealth services. Telemed J E Health 2013 Oct;19(10):786-790 [FREE Full text] [doi: 10.1089/tmj.2012.0272] [Medline: 23931702]
74. Genoe R, Kulczycki C, Marston H, Freeman S, Musselwhite C, Rutherford H. E-leisure and older adults: findings from an international exploratory study. Ther Recreat J 2018;52(1):1-18. [doi: 10.18666/trj-2018-v52-i1-8417]
75. Hamblin K. Telecare, obtrusiveness, acceptance and use: an empirical exploration. Br J Occup Ther 2016 Sep 30;80(2):132-138. [doi: 10.1177/0308022616667751]
76. Betts LR, Hill R, Gardner SE. "There's not enough knowledge out there": examining older adults' perceptions of digital technology use and digital inclusion classes. J Appl Gerontol 2019 Aug;38(8):1147-1166. [doi: 10.1177/0733464817737621] [Medline: 29165038]
77. Chang CP, Lee TT, Mills ME. Experience of home telehealth technology in older patients with diabetes. Comput Inform Nurs 2017 Oct;35(10):530-537. [doi: 10.1097/CIN.0000000000000341] [Medline: 28291156]
78. Lin X, Chen T. A qualitative approach for the elderly's needs in service robots design. In: Proceedings of the 2018 International Conference on Service Robotics Technologies. 2018 Presented at: ICSRT '18; March 16-19, 2018; Chengdu, China p. 67-72. [doi: 10.1145/3208833.3208846]
79. Kim J, Lee HY, Christensen MC, Merighi JR. Technology access and use, and their associations with social engagement among older adults: do women and men differ? J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):836-845. [doi: 10.1093/geronb/gbw123] [Medline: 28073816]
80. Middlemass JB, Vos J, Siriwardena AN. Perceptions on use of home telemonitoring in patients with long term conditions - concordance with the Health Information Technology Acceptance Model: a qualitative collective case study. BMC Med Inform Decis Mak 2017 Jun 26;17(1):89 [FREE Full text] [doi: 10.1186/s12911-017-0486-5] [Medline: 28651588]
81. Tyler M, Simic V, De George-Walker L. Older adult Internet super-users: counsel from experience. Act Adapt Aging 2018 Jan 26;42(4):328-339. [doi: 10.1080/01924788.2018.1428472]
82. Krug RD, Xavier AJ, d'Orsi E. Factors associated with maintenance of the use of internet, EpiFloripa Idoso longitudinal study. Rev Saude Publica 2018 Apr 09;52:37 [FREE Full text] [doi: 10.11606/S1518-8787.2018052000216] [Medline: 29641661]
83. Bujnowska-Fedak MM, Pirogowicz I. Support for e-health services among elderly primary care patients. Telemed J E Health 2014 Aug;20(8):696-704 [FREE Full text] [doi: 10.1089/tmj.2013.0318] [Medline: 24359252]
84. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017;32(6):E1-E7 [FREE Full text] [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
85. Preusse KC, Mitzner TL, Fausset CB, Rogers WA. Older adults' acceptance of activity trackers. J Appl Gerontol 2017 Feb;36(2):127-155 [FREE Full text] [doi: 10.1177/0733464815624151] [Medline: 26753803]
86. Beer JM, Prakash A, Smarr CA, Chen TL, Hawkins K, Nguyen H, et al. Older users' acceptance of an assistive robot: attitudinal changes following brief exposure. Gerontechnology 2017 Mar 29;16(1):21-36 [FREE Full text] [doi: 10.4017/gt.2017.16.1.003.00] [Medline: 31178671]
87. Davenport RD, Mann W, Lutz B. How older adults make decisions regarding smart technology: an ethnographic approach. Assist Technol 2012;24(3):168-183. [doi: 10.1080/10400435.2012.659792] [Medline: 23033734]
88. Aure CF, Kluge A, Moen A. Older adults' engagement in technology-mediated self-monitoring of diet: a mixed-method study. J Nurs Scholarsh 2021 Jan;53(1):25-34 [FREE Full text] [doi: 10.1111/jnu.12619] [Medline: 33316147]
89. van Deursen AJ, Helsper EJ. A nuanced understanding of Internet use and non-use among the elderly. Eur J Commun 2015 Apr 09;30(2):171-187. [doi: 10.1177/0267323115578059]
90. Wu YH, Wrobel J, Cornuet M, Kerhervé H, Damnée S, Rigaud AS. Acceptance of an assistive robot in older adults: a mixed-method study of human-robot interaction over a 1-month period in the Living Lab setting. Clin Interv Aging 2014 May 8;9:801-811 [FREE Full text] [doi: 10.2147/CIA.S56435] [Medline: 24855349]
91. Gilly MC, Celsi MW, Schau HJ. It don't come easy: overcoming obstacles to technology use within a resistant consumer group. J Consum Aff 2012;46(1):62-89. [doi: 10.1111/j.1745-6606.2011.01218.x]
92. Chen CK, Tsai TH, Lin YC, Lin CC, Hsu SC, Chung CY, et al. Acceptance of different design exergames in elders. PLoS One 2018 Jul 5;13(7):e0200185 [FREE Full text] [doi: 10.1371/journal.pone.0200185] [Medline: 29975755]

93. Wang L, Rau PP, Salvendy G. Older adults' acceptance of information technology. Educ Gerontol 2011 Dec;37(12):1081-1099. [doi: 10.1080/03601277.2010.500588]
94. Chen TL, Bhattacharjee T, Beer JM, Ting LH, Hackney ME, Rogers WA, et al. Older adults' acceptance of a robot for partner dance-based exercise. PLoS One 2017 Oct 18;12(10):e0182736 [FREE Full text] [doi: 10.1371/journal.pone.0182736] [Medline: 29045408]
95. Brickwood KJ, Williams AD, Watson G, O'Brien J. Older adults' experiences of using a wearable activity tracker with health professional feedback over a 12-month randomised controlled trial. Digit Health 2020 Apr 26;6:2055207620921678 [FREE Full text] [doi: 10.1177/2055207620921678] [Medline: 32426152]
96. Fischl C, Asaba E, Nilsson I. Exploring potential in participation mediated by digital technology among older adults. J Occup Sci 2017 Jun 22;24(3):314-326. [doi: 10.1080/14427591.2017.1340905]
97. Zambianchi M, Rönnlund M, Carelli MG. Attitudes towards and use of information and communication technologies (ICTs) among older adults in Italy and Sweden: the influence of cultural context, socio-demographic factors, and time perspective. J Cross Cult Gerontol 2019 Sep;34(3):291-306. [doi: 10.1007/s10823-019-09370-y] [Medline: 31079248]
98. Kong L, Woods O. Smart eldercare in Singapore: negotiating agency and apathy at the margins. J Aging Stud 2018 Dec;47:1-9. [doi: 10.1016/j.jaging.2018.08.001] [Medline: 30447861]
99. Courtney KL. Privacy and senior willingness to adopt smart home information technology in residential care facilities. Methods Inf Med 2008;47(1):76-81. [doi: 10.3414/me9104] [Medline: 18213432]
100. Claes V, Devriendt E, Tournoy J, Milisen K. Attitudes and perceptions of adults of 60 years and older towards in-home monitoring of the activities of daily living with contactless sensors: an explorative study. Int J Nurs Stud 2015 Jan;52(1):134-148 [FREE Full text] [doi: 10.1016/j.ijnurstu.2014.05.010] [Medline: 24951084]
101. Callari TC, Ciairano S, Re A. Elderly-technology interaction: accessibility and acceptability of technological devices promoting motor and cognitive training. Work 2012;41 Suppl 1:362-369. [doi: 10.3233/WOR-2012-0183-362] [Medline: 22316750]
102. Seifert A, Schelling HR. Seniors online: attitudes toward the Internet and coping with everyday life. J Appl Gerontol 2018 Jan;37(1):99-109. [doi: 10.1177/0733464816669805] [Medline: 27655237]
103. Weaver CK, Zorn T, Richardson M. Goods not wanted. Older people's narratives of computer use rejection. Inf Commun Soc 2010 Aug;13(5):696-721. [doi: 10.1080/13691180903410535]
104. Ballantyne A, Trenwith L, Zubrinich S, Corlis M. 'I feel less lonely': what older people say about participating in a social networking website. Qual Ageing Older Adults 2010 Sep 13;11(3):25-35. [doi: 10.5042/qiaoa.2010.0526]
105. Barg-Walkow LH, Harrington CN, Mitzner TL, Hartley JQ, Rogers WA. Understanding older adults' perceptions of and attitudes towards exergames. Gerontechnology 2017;16(2):81-90 [FREE Full text] [doi: 10.4017/gt.2017.16.2.003.00] [Medline: 32025226]
106. Uei S, Tsai C, Yang M. Telecare service use among Taiwanese aged 60 and over: satisfaction, trust, and continued use intention. Soc Behav Pers 2013 Sep 01;41(8):1309-1318. [doi: 10.2224/sbp.2013.41.8.1309]
107. Lindberg J, Bhatt R, Ferm A. Older people and rural eHealth: perceptions of caring relations and their effects on engagement in digital primary health care. Scand J Caring Sci 2021 Dec;35(4):1322-1331 [FREE Full text] [doi: 10.1111/scs.12953] [Medline: 33448031]
108. Bajones M, Fischinger D, Weiss A, De La Puente P, Wolf D, Vincze M, et al. Results of field trials with a mobile service robot for older adults in 16 private households. J Hum Robot Interact 2020 Feb 07;9(2):1-27. [doi: 10.1145/3368554]
109. Chattaraman V, Kwon W, Gilbert JE, Ross K. Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. Compute Human Behav 2019 Jan;90:315-330. [doi: 10.1016/j.chb.2018.08.048]
110. Beuscher LM, Fan J, Sarkar N, Dietrich MS, Newhouse PA, Miller KF, et al. Socially assistive robots: measuring older adults' perceptions. J Gerontol Nurs 2017 Dec 01;43(12):35-43 [FREE Full text] [doi: 10.3928/00989134-20170707-04] [Medline: 28700074]
111. Arnaert A, Klooster J, Chow V. Technology innovations. Attitudes toward videotelephones: an exploratory study of older adults with depression. J Gerontol Nurs 2007;33(9):5-13.
112. Bakas T, Sampsel D, Israel J, Chamnikar A, Ellard A, Clark JG, et al. Satisfaction and technology evaluation of a telehealth robotic program to optimize healthy independent living for older adults. J Nurs Scholarsh 2018 Nov;50(6):666-675. [doi: 10.1111/jnu.12436] [Medline: 30320967]
113. Demiris G, Thompson H, Boquet J, Le T, Chaudhuri S, Chung J. Older adults' acceptance of a community-based telehealth wellness system. Inform Health Soc Care 2013 Jan;38(1):27-36 [FREE Full text] [doi: 10.3109/17538157.2011.647938] [Medline: 22571733]
114. Chou CC, Chang CP, Lee TT, Chou HF, Mills ME. Technology acceptance and quality of life of the elderly in a telecare program. Comput Inform Nurs 2013 Jul;31(7):335-342. [doi: 10.1097/NXN.0b013e318295e5ce] [Medline: 23728446]
115. Chu L, Chen HW, Cheng PY, Ho P, Weng IT, Yang PL, et al. Identifying features that enhance older adults' acceptance of robots: a mixed methods study. Gerontology 2019;65(4):441-440. [doi: 10.1159/000494881] [Medline: 30844813]

116. Gordon NP, Hornbrook MC. Older adults' readiness to engage with eHealth patient education and self-care resources: a cross-sectional survey. BMC Health Serv Res 2018 Mar 27;18(1):220 [FREE Full text] [doi: 10.1186/s12913-018-2986-0] [Medline: 29587721]
117. Baldassar L, Wilding R. Migration, aging, and digital kinning: the role of distant care support networks in experiences of aging well. Gerontologist 2020 Feb 24;60(2):313-321. [doi: 10.1093/geront/gnz156] [Medline: 31812983]
118. Wu YH, Cristancho-Lacroix V, Fassert C, Faucounau V, de Rotrou J, Rigaud AS. 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]
119. Aguilar A, Boerema C, Harrison J. Meanings attributed by older adults to computer use. J Occup Sci 2010 Apr;17(1):27-33. [doi: 10.1080/14427591.2010.9686669]
120. Andonian LC. Meanings and experiences associated with computer use of older immigrant adults of lower socioeconomic status: les sens et les expériences associés à l'utilisation des ordinateurs chez les immigrants âgés ayant un faible statut socioéconomique. Can J Occup Ther 2018 Apr;85(2):146-157. [doi: 10.1177/0008417418754395] [Medline: 29366333]
121. Lin CT, Chuang SS. A study of digital learning for older adults. J Adult Dev 2018 Jul 18;26(2):149-160. [doi: 10.1007/s10804-018-9314-0]
122. Arief M, Rissanen S, Saranto K. Influence of previous work experience and education on Internet use of people in their 60s and 70s. J Innov Health Inform 2018 Oct 18;25(3):132-141 [FREE Full text] [doi: 10.14236/jhi.v25i3.868] [Medline: 30398457]
123. Bluethmann SM, Coa KI, Alfano CM, Hesse BW. Electronic health information exchange opportunities for self-management of care: responses from older adults with and without cancer history in the United States. Curr Oncol Rep 2018 Mar 23;20(4):30 [FREE Full text] [doi: 10.1007/s11912-018-0674-1] [Medline: 29572671]
124. Holden RJ, Campbell NL, Abebe E, Clark DO, Ferguson D, Bodke K, Brain Health Patient Safety Laboratory. Usability and feasibility of consumer-facing technology to reduce unsafe medication use by older adults. Res Social Adm Pharm 2020 Jan;16(1):54-61 [FREE Full text] [doi: 10.1016/j.sapharm.2019.02.011] [Medline: 30853507]
125. Digital skills to connect: exploring the digital experiences of people aged 50-70 and the organisations supporting them during COVID-19. Center for Ageing Better. 2021 Jul. URL: https://ageing-better.org.uk/sites/default/files/2021-07/Digital-Skills-to-Connect.pdf [accessed 2022-02-11]
126. Ageing in the digital era – UNECE highlights key actions to ensure digital inclusion of older persons. United Nations Economic Commission for Europe. 2021. URL: https://unece.org/media/press/358156 [accessed 2022-02-11]
127. Vuorikari R, Punie Y, Carretero Gomez S, van den Brande G. DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model. Office of the European Union. Luxembourg, Luxembourg: Publications Office of the European Union; 2016. URL: https://publications.jrc.ec.europa.eu/repository/handle/JRC101254 [accessed 2022-09-29]
128. Honeyman M, Maguire D, Evans H, Davies A. Digital technology and health inequalities: a scoping review. Public Health Wales Research and Evaluation. Cardiff, UK: Public Health Wales NHS Trust; 2020. URL: https://phw.nhs.wales/publications/publications1/digital-technology-and-health-inequalities-a-scoping-review/ [accessed 2022-01-18]
129. Reddick CG, Enriquez R, Harris RJ, Sharma B. Determinants of broadband access and affordability: an analysis of a community survey on the digital divide. Cities 2020 Nov;106:102904 [FREE Full text] [doi: 10.1016/j.cities.2020.102904] [Medline: 32921864]
130. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
131. Tackling the digital divide - House of Commons, 4 November 2021. Local Government Association, United Kingdom. 2021 Nov 3. URL: https://www.local.gov.uk/parliament/briefings-and-responses/tackling-digital-divide-house-commons-4-november-2021 [accessed 2022-09-07]
132. 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). Official Journal of the European Union. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2022-10-03]
133. Protecting data privacy to maintain digital trust: The importance of protecting data privacy during the pandemic and beyond. Cisco 2020 Consumer Privacy Survey. San Jose, CA, USA: Cisco Systems; 2020. URL: https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cybersecurity-series-2020-cps.pdf [accessed 2022-09-08]
134. From privacy to profit: Achieving positive return on privacy investment. Cisco data privacy benchmark study 2020. Cisco Cybersecurity Series 2020. San Jose, CA, USA: Cisco Systems; 2020 Jan. URL: https://www.cisco.com/c/dam/global/en_uk/products/collateral/security/2020-data-privacy-cybersecurity-series-jan-2020.pdf [accessed 2022-09-28]
135. Burke PJ, Stets JE. Identity Theory. Oxford, UK: Oxford University Press; 2009.
136. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]

Abbreviations

COM-B: capability, opportunity, and motivation behavior change model
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for Scoping Reviews
TDF: theoretical domain framework

Edited by G Eysenbach, T Leung; submitted 09.06.22; peer-reviewed by R Frost, S Kim, K Meyer; comments to author 06.07.22; revised version received 21.08.22; accepted 23.10.22; published 07.12.22.

Please cite as:
Kebede AS, Ozolins LL, Holst H, Galvin K
Digital Engagement of Older Adults: Scoping Review
J Med Internet Res 2022;24(12):e40192
URL: https://www.jmir.org/2022/12/e40192
doi:10.2196/40192
PMID:36477006

Review

Applications of Artificial Intelligence to Obesity Research: Scoping Review of Methodologies

Ruopeng An[1*], MPP, PhD; Jing Shen[2*], PhD; Yunyu Xiao[3], PhD

[1]Brown School, Washington University in St. Louis, St. Louis, MO, United States

[2]Department of Physical Education, China University of Geosciences, Beijing, China

[3]Weill Cornell Medical College, Cornell University, Ithaca, NY, United States

[*]these authors contributed equally

Corresponding Author:
Jing Shen, PhD
Department of Physical Education, China University of Geosciences
No. 29, Xueyuan Road, Haidian District
Beijing, 100083
China
Phone: 86 010 82322397
Email: shenjing@cugb.edu.cn

Abstract

Background: Obesity is a leading cause of preventable death worldwide. Artificial intelligence (AI), characterized by machine learning (ML) and deep learning (DL), has become an indispensable tool in obesity research.

Objective: This scoping review aimed to provide researchers and practitioners with an overview of the AI applications to obesity research, familiarize them with popular ML and DL models, and facilitate the adoption of AI applications.

Methods: We conducted a scoping review in PubMed and Web of Science on the applications of AI to measure, predict, and treat obesity. We summarized and categorized the AI methodologies used in the hope of identifying synergies, patterns, and trends to inform future investigations. We also provided a high-level, beginner-friendly introduction to the core methodologies to facilitate the dissemination and adoption of various AI techniques.

Results: We identified 46 studies that used diverse ML and DL models to assess obesity-related outcomes. The studies found AI models helpful in detecting clinically meaningful patterns of obesity or relationships between specific covariates and weight outcomes. The majority (18/22, 82%) of the studies comparing AI models with conventional statistical approaches found that the AI models achieved higher prediction accuracy on test data. Some (5/46, 11%) of the studies comparing the performances of different AI models revealed mixed results, indicating the high contingency of model performance on the data set and task it was applied to. An accelerating trend of adopting state-of-the-art DL models over standard ML models was observed to address challenging computer vision and natural language processing tasks. We concisely introduced the popular ML and DL models and summarized their specific applications in the studies included in the review.

Conclusions: This study reviewed AI-related methodologies adopted in the obesity literature, particularly ML and DL models applied to tabular, image, and text data. The review also discussed emerging trends such as multimodal or multitask AI models, synthetic data generation, and human-in-the-loop that may witness increasing applications in obesity research.

(J Med Internet Res 2022;24(12):e40589) doi:10.2196/40589

KEYWORDS

artificial intelligence; deep learning; machine learning; obesity; scoping review

Introduction

Background

The double burden of malnutrition, characterized by the coexistence of overnutrition (eg, overweight and obesity) and undernutrition (eg, stunting and wasting), is present at all levels of the population: country, city, community, household, and individual [1]. Obesity is a leading cause of preventable death and consumes substantial social resources in many high-income and some low- and middle-income economies [2]. Worldwide, the obesity rate has nearly tripled since 1975 [3]. In 2016, 13% of the global population, or 650 million adults, were obese [4].

More than 340 million children and adolescents aged 5 to 19 years and 39 million children aged <5 years were overweight or obese [4]. By 2025, the global obesity prevalence is projected to reach 18% among men and 21% among women [5].

Health data are now available to researchers and practitioners in ways and quantities that have never existed before, presenting unprecedented opportunities for advancing health sciences through state-of-the-art data analytics [6]. By contrast, dealing with large-scale, complex, unconventional data (eg, text, image, video, and audio) requires innovative analytic tools and computing power only available in recent years [7,8]. Artificial intelligence (AI), characterized by machine learning (ML) and deep learning (DL), has become increasingly recognized as an indispensable tool in health sciences, with relevant applications expanding from disease outbreak prediction to medical imaging and patient communication to behavioral modification [9-14]. Over the past decade, an upsurge of the scientific literature adopting AI in health research has been witnessed [15,16]. These investigations applied a wide range of AI models: from *shallow* ML algorithms (eg, decision trees (DTs) and k-means clustering) and *deep* neural networks [17] to various data sources (eg, clinical and observational) and types (eg, tabular, text, and image) [18]. This boom in AI applications raises many questions [19-21]: How do AI-based approaches differ from conventional statistical analyses? Do AI techniques provide additional benefits or advantages over traditional methods? What are the typical AI applications and algorithms applied in obesity research? Is AI a buzzword that will eventually fall out of fashion, or will the upward trend of AI adoption to study obesity continue in the future?

Synthesizing and Disseminating AI Methodologies Adopted in Obesity Research

Three previous studies reviewed the applications of AI in weight loss interventions through diet and exercise [22-24]. They found preliminary but promising evidence regarding the effectiveness of AI-powered tools in decision support and digital health interventions [22-24]. However, to our knowledge, no study has been conducted to summarize AI algorithms, models, and methods applied to obesity research. This study remains the first methodological review on the applications of AI to measure, predict, and treat childhood and adult obesity. It serves 2 purposes: synthesizing and disseminating AI methodologies adopted in obesity research. First, we focused on summarizing and categorizing AI methodologies used in the obesity literature in the hope of identifying synergies, patterns, and trends to inform future scientific investigations. Second, we provided a high-level, beginner-friendly introduction to the core methodologies for interested readers, aiming to facilitate the dissemination and adoption of various AI techniques.

Methods

The scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [25].

Study Selection Criteria

Studies that met all of the following criteria were included in the review: (1) study design: experimental or observational studies; (2) analytic approach: use of AI, including ML and DL (ie, deep neural networks), in measuring, predicting, or intervening obesity-related outcomes; (3) study participants: humans of all ages; (4) outcomes: obesity or body weight status (eg, BMI, body fat percentage [BFP], waist circumference [WC], and waist-to-hip ratio [WHR]); (5) article type: original, empirical, and peer-reviewed journal publications; (6) time window of search: from the inception of an electronic bibliographic database to January 1, 2022; and (7) language: articles written in English.

Studies that met any of the following criteria were excluded from the review: (1) studies focusing on outcomes other than obesity (eg, diet, physical activity, energy expenditure, and diabetes); (2) studies that used a rule-based (*hard-coded*) approach rather than example-based ML or DL; (3) articles not written in English; and (4) letters, editorials, study or review protocols, case reports, and review articles.

Search Strategy

A keyword search was performed in 2 electronic bibliographic databases: PubMed and Web of Science. The search algorithm included all possible combinations of keywords from the following two groups: (1) “artificial intelligence,” “computational intelligence,” “machine intelligence,” “computer reasoning,” “machine learning,” “deep learning,” “neural network,” “neural networks,” or “reinforcement learning” and (2) “obesity,” “obese,” “overweight,” “body mass index,” “BMI,” “adiposity,” “body fat,” “waist circumference,” “waist to hip,” or “waist - to - hip.” The Medical Subject Headings terms “artificial intelligence” and “obesity” were included in the PubMed search. Multimedia Appendix 1 documents the search algorithm used in PubMed. Two coauthors of this review independently conducted title and abstract screening on the articles identified from the keyword search, retrieved potentially eligible articles, and evaluated their full texts. The interrater agreement between the 2 coauthors was assessed with Cohen kappa (κ=0.80). Discrepancies were resolved through discussion.

Data Extraction and Synthesis

A standardized data extraction form was used to collect the following methodological and outcome variables from each included study: authors; year of publication; country; data collection period; study design; sample size; training, validation, and test set size; sample characteristics; the proportion of female participants; age range; AI models used; input data source; input data format; input features; outcome data type; outcome measures; unit of analysis; main study findings; and implications for the effectiveness and usefulness of AI in measuring, predicting, or intervening obesity-related outcomes.

Methodological Review

We classified AI methodologies adopted by the included studies into 2 primary categories: ML and DL models. Among the ML models, methods were organized into 2 subcategories: unsupervised and supervised learning. Among the DL models, methods were classified into 3 subcategories: tabular data

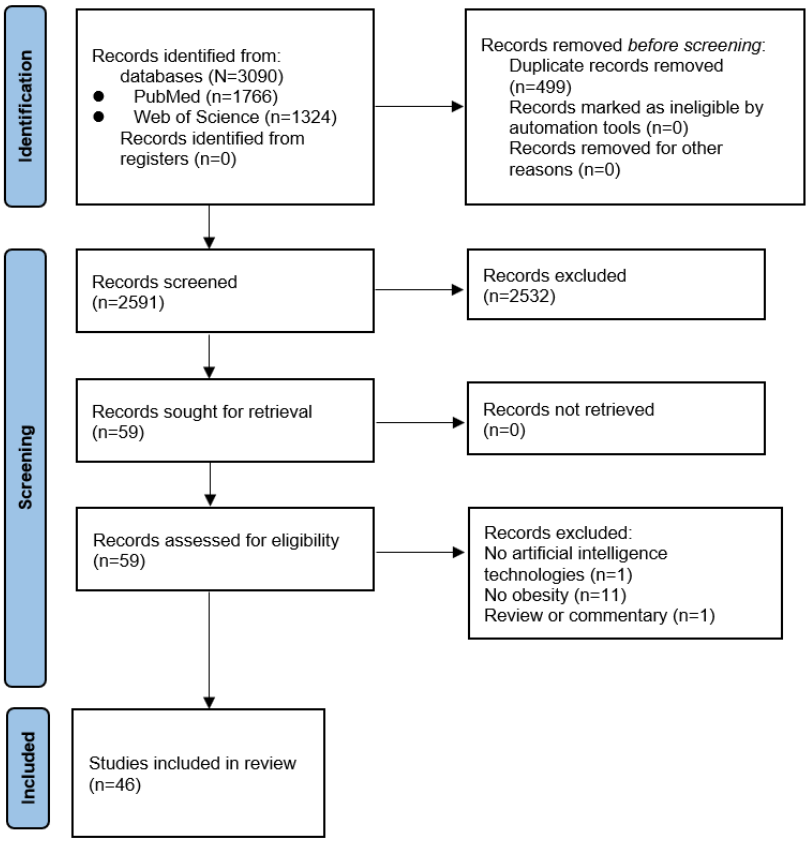
modeling, computer vision (CV), and natural language processing (NLP). Rather than enumerating every single model performed by the included studies, which is unnecessary and unilluminating, we focused on the popular models used by multiple studies.

Results

Identification of Studies

Figure 1 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. We identified a total of 3090 articles through the keyword search, and after removing 499 (16.15%) duplicates, 2591 (83.85%) unique articles underwent title and abstract screening. Of these 2591 articles, 2532 (97.72%) were excluded, and the full texts of the remaining 59 (2.28%) were reviewed against the study selection criteria. Of these 59 articles, 13 (22%) were excluded. The reasons for exclusion were as follows: no adoption of AI technologies (1/13, 8%), no obesity-related outcomes (11/13, 85%), and commentary rather than original empirical research (1/13, 8%). Therefore, of the 3090 articles identified initially through the keyword search, 46 (1.49%) were included in the review [26-71].

Figure 1. Identification of studies via databases and registers.



Study Characteristics

Table 1 summarizes the key characteristics of the 46 included studies. An increasing trend in relevant publications was observed. The earliest study included in the review was published in 1997; others were published in, or after, 2008; for example, 2% (1/46) each in 2008, 2012, and 2017; 4% (2/46) each in 2014 and 2016; 7% (3/46) each in 2009 and 2015; 9% (4/46) in 2018; 15% (7/46) in 2019; 20% (9/46) in 2020; and 26% (12/46) in 2021. Of the 46 studies, 16 (35%) were conducted in the United States [28,32,33,37,42,46,48,50-53,57,58,60,62,63]; 6 (13%) in China [39,40,45,56,64,65]; 3 (7%) each in the United Kingdom [27,68,69] and Korea [35,43,49]; 2 (4%) each in Italy [36,71], Turkey [41,70], Finland [44,59], Germany [54,55], and India [36,71]; and 1 (2%) each in Saudi Arabia [26], Iran [67], Serbia [66], Portugal [61], Spain [47], Singapore [38], Australia [34], and Indonesia [29]. Of the 46 studies, 32 (70%) adopted a cross-sectional study design [26,27,29-32,37,39-42,46-50,52,55-58,60-63,65-71], 7 (15%) a prospective study design [28,33,38,43,45,54,59], 6 (13%) a retrospective study design [34-36,51,53,64], and 1 (2%) a cotwin control design [44]. Sample sizes varied substantially across the included studies, ranging from 20 to 5,265,265. Of the 46 studies, 7 (15%) had a sample size of between 20 and 82; 11 (24%) between 130 and 600; 19 (41%) between 1061 and 9524; 6 (13%) between 16,553 and 49,805; 2 (4%) between 244,053 and 618,898; and 1 (2%) study had a sample size of 5,265,265. Of the 46 studies, 23 (50%) focused on adults, 14 (30%) on children and adolescents, 1 (2%) on people of all ages, and the remaining 8 (17%) did not report the age range of participants.

Table 1. Characteristics of the studies included in the review.

Authors, year	Country	Data collection period	Study design	Sample size	Training set size	Validation set size; test set size	Sample characteristics	Female participants (%)	Age (years)	AI[a] model
Abdel-Aal and Mangoud [26], 1997	Saudi Arabia	1995	Cross-sectional	1100	800	N/A; 300	Patients	N/A[b]	≥20	NN[c] (AIM[d] abductive)
Positano et al [71], 2008	Italy	N/A	Cross-sectional	20	N/A	N/A	Participants with varying levels of obesity	N/A	Mean 52 (SD 16)	Fuzzy c-means
Ergün [70], 2009	Turkey	N/A	Cross-sectional	82	41	N/A; 41	Participants with different ranges of obesity	N/A	N/A	LR[e], MLP[f]
Yang et al [69], 2009	United Kingdom	N/A	Cross-sectional	507	N/A	N/A	Patients	N/A	N/A	SVM[g]
Zhang et al [68], 2009	United Kingdom	1988 to 2003	Cross-sectional	16,553	11,091	N/A; 5462	Children	N/A	Birth to 3	NB[h], SVM, DT[i], NN
Heydari et al [67], 2012	Iran	2010	Cross-sectional	414	248	N/A; 104	Healthy military personnel	N/A	Mean 34.4 (SD 7.5)	NN, LR
Kupusinac et al [66], 2014	Serbia	N/A	Cross-sectional	2755	1929	413; 413	Adults	48.3	18 to 88	NN
Shao [65], 2014	China	N/A	Cross-sectional	248	174	N/A; 74	N/A	N/A	N/A	MR[j], MARS[k], SVM, NN
Chen et al [64], 2015	China	N/A	Retrospective	476	N/A	N/A	Participants with different ranges of obesity	62.4	22 to 82	NN (ELM[l])
Dugan et al [63], 2015	United States	N/A	Cross-sectional	7519	6767	N/A; 752	Children	49	2 to 10	DT, RF[m], NB, NN (BN[n])
Nau et al [62], 2015	United States	2010	Cross-sectional	22,497	15,073	N/A; 7424	Children	N/A	10 to 18	RF
Almeida et al [61], 2016	Portugal	2009 to 2013	Cross-sectional	3084	1537	N/A; 664	School-age children	49.7	9	LR, NN
Lingren et al [60], 2016	United States	N/A	Cross-sectional	428	257	N/A; 86	Children	N/A	1 to 6	SVM, NB
Seyednasrollah et al et al [59], 2017	Finland	1980 to 2012	Prospective	2262	1625	N/A; 637	Adults	N/A	≥18	GB[o]
Hinojosa et al [58], 2018	United States	2003 to 2007	Cross-sectional	5,265,265	N/A	N/A	School-age children: grades 5, 7, and 9	N/A	N/A	RF
Maharana and Nsoesie [57], 2018	United States	2017	Cross-sectional	1695	508	N/A; 339	Adults	N/A	≥18	NN (CNN[p])
Wang et al [56], 2018	China	2014 to 2015	Cross-sectional	139	111	N/A; 28	Participants with different ranges of obesity	36.7	27 to 53	SVM, KNN[q], DT, LR
Duran et al [55], 2018	Germany	1999 to 2004	Cross-sectional	1999	1333	N/A; 666	Children	42.8	8 to 19	NN

Authors, year	Country	Data collection period	Study design	Sample size	Training set size	Validation set size; test set size	Sample characteristics	Female participants (%)	Age (years)	AI[a] model
Gerl et al [54], 2019	Germany	2012; 1991 to 1994	Prospective	1061	796	206; 250	N/A	53.8	N/A	Cubist, LASSO[r], PLS[s], GB, RF, LM[t]
Hammond et al [53], 2019	United States	2008 to 2016	Retrospective	3449	482	N/A; 207	Children	49.2	4.5 to 5.5	LASSO, RF, GB
Hong et al [52], 2019	United States	2008	Cross-sectional	1237	1400	N/A; 600	Patients	N/A	≥18	LR, SVM, DT, RF
Ramyaa et al [51], 2019	United States	1993 to 1994	Retrospective	48,508	33,956	N/A; 14,552	Postmenopausal women	100	50 to 79	SVM, KNN, DT, PCA[u], RF, NN
Scheinker et al [50], 2019	United States	2018	Cross-sectional	3138	N/A	N/A	Census population	49.9	All ages	LM, GB
Shin et al [49], 2019	Korea	N/A	Cross-sectional	163	143	N/A; 20	Amateur athletes	37.4	17 to 25	NN
Stephens et al [48], 2019	United States	N/A	Cross-sectional	23	N/A	N/A	Youth with obesity symptoms	57	Range 9.78-18.54	NN
Blanes-Selva et al [47], 2020	Spain	N/A	Cross-sectional	49,805	39,844	N/A; 9961	Patients	N/A	N/A	PU[v] learning
Dunstan et al [46], 2020	United States	2008	Cross-sectional	79	N/A	N/A	Adults	N/A	≥20	SVM, RF, GB
Fu et al [45], 2020	China	1999 to 2003	Prospective	2125	1143	381; 382	Children	40.6	4 to 7	GB
Kibble et al [44], 2020	Finland	N/A	Cotwin control	43	N/A	N/A	Young adult monozygotic twin pairs	53	22 to 36	GFA[w]
Park et al [43], 2020	Korea	N/A	Prospective	76	75	N/A; 1	Adolescents	6.8; N/A	Mean 11.94 (SD 3.13); mean 13.42 (SD 3.25)	LASSO
Phan et al [42], 2020	United States	2017 to 2018	Cross-sectional	18,700 images	14,960	N/A; 3740	Adolescents and adults	N/A	N/A	LM, NN (CNN)
Taghiyev et al [41], 2020	Turkey	2019	Cross-sectional	500	325	N/A; 175	Female patients	100	≥18	DT, LR
Xiao et al [40], 2020	China	2007 to 2010	Cross-sectional	9524	N/A	N/A	Residents	54	≥18	LR, NN (CNN)
Yao et al [39], 2020	China	N/A	Cross-sectional	67; 24	N/A	N/A	Smartphone users	N/A; 41.7	Mean 25.19; range 18-46	NN
Alkutbe et al [27], 2021	United Kingdom	2014; 2015 to 2016	Cross-sectional	1223	977	N/A; 246	Children	61.8	8 to 12	GB
Bhanu et al [38], 2021	Singapore	2003 to 2006	Prospective	130	104	N/A; 26	Older adults	69.5	Mean 67.85 (SD 7.90)	NN (U-Net)

Authors, year	Country	Data collection period	Study design	Sample size	Training set size	Validation set size; test set size	Sample characteristics	Female participants (%)	Age (years)	AI[a] model
Cheng et al [37], 2021	United States	2003 to 2004; 2005 to 2006	Cross-sectional	7162	N/A	N/A	Adults	48.6	20 to 85	NB, KNN, MEFC[x], DT, NN (MLP)
Delnevo et al [36], 2021	Italy	N/A	Retrospective	221	176	N/A; 45	Participants with different ranges of obesity	N/A	N/A	GB, RF
Lee et al [35], 2021	Korea	2015 to 2020	Retrospective	3159	2370	N/A; 789	Obstetric patients and their newborns	100	20 to 44	LM, RF, NN
Lin et al [34], 2021	Australia	2010 to 2019	Retrospective	2495	882	N/A; 1613	Participants with different ranges of obesity	67.4	21 to 36	Two-step cluster analysis, k-means
Pang et al [33], 2021	United States	2009 to 2017	Prospective	27,203	21,762	N/A; 5441	Children	49.2	<2	DT, NB, LR, SVM, GB, NN
Park et al [32], 2021	United States	2014 to 2016	Cross-sectional	5000 tweets	4500	N/A; 500	Twitter users	60.7	Mean 51.91 (SD 17.20)	NB, SVM, NN (CNN, LSTM[y])
Rashmi et al [31], 2021	India	2020	Cross-sectional	600 images	420	120; 60	Children	50	8 to 11	SVM, NB, RF
Snekhalatha and Sangamithirai [30], 2021	India	N/A	Cross-sectional	2700 images	2000	500; 200	Adults	N/A	Mean 45 (SD 2.5)	NN (VGG, ResNet, DenseNet)
Thamrin et al [29], 2021	Indonesia	2018	Cross-sectional	618,898	557,008	N/A; 61,890	Adults	N/A	≥18	DT, NB, LR

Authors, year	Country	Data collection period	Study design	Sample size	Training set size	Validation set size; test set size	Sample characteristics	Female participants (%)	Age (years)	AI[a] model
Zare et al [28], 2021	United States	2003 to 2019	Prospective	244,053	162,702	N/A; 81,351	Children	49	5 to 6	DT, LR, RF, NN

[a]AI: artificial intelligence.
[b]N/A: not applicable.
[c]NN: neural network.
[d]AIM: abductory induction mechanism.
[e]LR: logistic regression.
[f]MLP: multilayer perceptron.
[g]SVM: support vector machine.
[h]NB: naïve Bayes.
[i]DT: decision tree.
[j]MR: multiple regression.
[k]MARS: multivariate adaptive regression splines.
[l]ELM: extreme learning machine.
[m]RF: random forest.
[n]BN: BayesNet.
[o]GB: gradient boosting.
[p]CNN: convolutional neural network.
[q]KNN: k-nearest neighbor.
[r]LASSO: least absolute shrinkage and selection operator.
[s]PLS: partial least squares.
[t]LM: linear model.
[u]PCA: principal component analysis.
[v]PU: positive and unlabeled.
[w]GFA: group factor analysis.
[x]MEFC: multiobjective evolutionary fuzzy classifier.
[y]LSTM: long short-term memory.

Data Sources and Outcome Measures

Table 2 summarizes the data sources and outcome measures of the studies included in the review. Input data were obtained from a variety of sources, including health surveys (eg, National Health and Nutrition Examination Survey), electronic health records, magnetic resonance imaging (MRI) scans, social media data (eg, tweets), and geographically aggregated data sets (eg, InfoUSA and Dun & Bradstreet). Of the 46 studies, 34 (74%) analyzed tabular data (eg, spreadsheet data) [26-29,33-37,39,41,44-47,49-51,53-56,58-68,70], 8 (17%) analyzed digital image data [30,31,38,40,42,43,57,71], and 4 (9%) analyzed text data [32,48,52,69]. Obesity-related measures used across the studies included anthropometrics (eg, body weight, BMI, BFP, WC, and WHR) and biomarkers.

Table 2. Data sources and measures of outcomes in the studies included in the review.

Authors, year	Input data source	Input data format	Input features (independent variables)	Outcome data type	Outcome measures	Unit of analysis
Abdel-Aal and Mangoud [26], 1997	Medical survey data	Tabular	13 health parameters	Continuous	WHR[a]	Individual
Positano et al [71], 2008	MRI[b]	Image	Subcutaneous adipose tissue and visceral adipose tissue	Binary	Abdominal adipose tissue distribution	Individual
Ergün [70], 2009	Obtained from participants	Tabular	24 obesity parameters	Binary	Classification of obesity	Individual
Yang et al [69], 2009	Clinical data	Text	Clinical discharge summaries	Binary	Obesity status	Individual
Zhang et al [68], 2009	Objective measure	Tabular	Data recorded regarding the weight of the child during the first 2 years of the child's life	Binary	Obesity	Individual
Heydari et al [67], 2012	Questionnaire and objective measure	Tabular	Age, systole, diastole, weight, height, BMI, WC[c], HC[d], and triceps skinfold and abdominal thicknesses	Binary	Obesity	Individual
Kupusinac et al [66], 2014	Objective measure	Tabular	Gender, age, and BMI	Continuous	BFP[e]	Individual
Shao [65], 2014	Objective measure	Tabular	13 body circumference measurements	Continuous	BFP	Individual
Chen et al [64], 2015	Objective measure	Tabular	18 blood indexes and 16 biochemical indexes	Continuous	Overweight	Individual
Dugan et al [63], 2015	Questionnaire and objective measure	Tabular	167 clinical data attributes	Continuous	Obesity	Individual
Nau et al [62], 2015	Two secondary data sources (InfoUSA and Dun & Bradstreet)	Tabular	44 community characteristics	Binary	Obesogenic and obesoprotective environments	Community
Almeida et al [61], 2016	Objective measure	Tabular	Age, sex, BMI *z* score, and calf circumference	Continuous	BFP	Individual
Lingren et al [60], 2016	EHR[f]	Tabular	EHR data	Binary	Obesity	Individual
Seyednasrollah et al [59], 2017	Objective measure	Tabular	Clinical factors and genetic risk factors	Binary	Obesity	Individual
Hinojosa et al [58], 2018	Objective measure	Tabular	School environment	Binary	Obesity	School
Maharana and Nsoesie [57], 2018	Objective measure	Image	Built environment	Continuous	Prevalence of obesity	Census tract
Wang et al [56], 2018	Objective measure	Tabular	Single-nucleotide polymorphisms	Binary	Obesity risk	Individual
Duran et al [55], 2018	NHANES[g]	Tabular	Age, height, weight, and WC	Binary	Excess body fat	Individual
Gerl et al [54], 2019	Objective measure	Tabular	Human plasma lipidomes	Binary and continuous	Obesity: BMI, WC, WHR, and BFP	Individual
Hammond et al [53], 2019	EHR and publicly available data	Tabular	EHR data	Binary and continuous	Obesity status	Individual
Hong et al [52], 2019	EHR	Text	Discharge summaries	Binary	Identification of obesity	Individual
Ramyaa et al [51], 2019	Questionnaire	Tabular	Energy balance components	Binary and continuous	Energy stores: body weight	Individual

Authors, year	Input data source	Input data format	Input features (independent variables)	Outcome data type	Outcome measures	Unit of analysis
Scheinker et al [50], 2019	2018 Robert Wood Johnson Foundation County Health Rankings	Tabular	Demographic factors, socioeconomic factors, health care factors, and environmental factors	Continuous	Obesity prevalence	County
Shin et al [49], 2019	Objective measure	Tabular	Upper body impedance and lower body anthropometric data	Continuous	BFP	Individual
Stephens et al [48], 2019	From recorded dialogue	Text	Dialogue	Binary	Weight management program	Individual
Blanes-Selva et al [47], 2020	EHR of HULAFE[h]	Tabular	32 variables	Binary	Identification of obesity	Individual
Dunstan et al [46], 2020	Euromonitor data set	Tabular	National sales of a small subset of food and beverage categories	Continuous	Nationwide obesity prevalence	Country
Fu et al [45], 2020	Clinical data	Tabular	Demographic characteristics, maternal anthropometrics, perinatal clinical history, laboratory tests, and postnatal feeding practices	Binary	Obesity	Individual
Kibble et al [44], 2020	Clinical data	Tabular	42 clinical variables	Binary	Mechanisms of obesity	Individual
Park et al [43], 2020	Openly accessible database	Image	Neuroimaging biomarkers	Continuous	BMI	Individual
Phan et al [42], 2020	Objective measure	Image	Neighborhood built environment characteristics	Binary, continuous	Obesity	State
Taghiyev et al [41], 2020	EHR	Tabular	Results of blood tests	Binary	Obesity	Individual
Xiao et al [40], 2020	Objective measure	Image	Vertical greenness level	Binary	Obesity	Individual
Yao et al [39], 2020	Objective measure	Tabular	Characteristics of body movement captured by smartphone's built-in motion sensors	Continuous	BMI	Individual
Alkutbe et al [27], 2021	Self-reported and objective measures	Tabular	Weight, height, age, and gender	Binary and continuous	BFP	Individual
Bhanu et al [38], 2021	MRI	Image	SAT[i] and VAT[j]	Binary	Abdominal fat	Individual
Cheng et al [37], 2021	Objective measure	Tabular	Physical activity	Binary	Obesity	Individual
Delnevo et al [36], 2021	Questionnaire	Tabular	Positive and negative psychological variables	Binary and continuous	BMI values and BMI status	Individual
Lee et al [35], 2021	Objective measure	Tabular	64 independent variables: nationwide multicenter ultrasound data and maternal and delivery information	Continuous	BMI	Individual
Lin et al [34], 2021	Objective measure	Tabular	Key clinical variables	Binary	Obesity classification criterion	Individual
Pang et al [33], 2021	EHR data from pediatric big data repository	Tabular	Demographic variables and 54 clinical variables	Binary	Obesity	Individual
Park et al [32], 2021	Corpus of geotagged tweets	Text	Tweets	Binary and continuous	BMI and obesity	Individual
Rashmi et al [31], 2021	Objective measure	Image	600 thermograms	Binary	Obesity	Individual
Snekhalatha and Sangamithirai [30], 2021	Objective measure	Image	Thermal imaging	Binary	Diagnosis of obesity	Individual

Authors, year	Input data source	Input data format	Input features (independent variables)	Outcome data type	Outcome measures	Unit of analysis
Thamrin et al [29], 2021	Publicly available health data	Tabular	Risk factors for obesity	Binary	Obesity	Individual
Zare et al [28], 2021	BMI panel data set	Tabular	Kindergarten BMI *z* score	Binary	Obesity by grade 4	Individual

[a]WHR: waist-hip ratio.

[b]MRI: magnetic resonance imaging.

[c]WC: waist circumference.

[d]HC: hip circumference.

[e]BFP: body fat percentage.

[f]EHR: electronic health record.

[g]NHANES: National Health and Nutrition Examination Survey.

[h]HULAFE: Hospital Universitari i Politècnic La Fe.

[i]SAT: subcutaneous adipose tissue.

[j]VAT: visceral adipose tissue.

Main Findings

Table 3 summarizes the estimated effects and main findings of the studies included in the review. Four key findings have emerged.

First, the studies found that ML or DL models were generally effective in detecting clinically meaningful patterns of obesity or relationships between covariates and weight outcomes; for example, ML and DL models were found useful in classifying obesity severity [30,47,52], identifying anthropometric [34] and genetic characteristics of obesity [56], and predicting obesity onset in children [28,53,63]. ML algorithms (eg, random forest [RF] and conditional RF) revealed meaningful relationships between school and neighborhood environments and overweight and obesity [45,58,62]. DL algorithms (eg, convolutional neural network [CNN]) effectively extracted built environment features from satellite images to assess their associations with the local obesity rate [57].

Second, most (18/22, 82%) of the studies comparing AI models with conventional statistical methods reported that the AI models achieved higher prediction accuracy on test data, whereas others (4/22, 18%) found similar model performances; for example, ML and DL models were found to explain a larger proportion of variations in county-level obesity prevalence than conventional statistical approaches [50]. ML models showed flexibility in handling various variable types [36,41] and large-scale data sets [32] and producing robust, generalizable inferences [41,54,64,65] with higher prediction accuracy [61,66]. By contrast, Cheng et al [37] reported that ML algorithms and conventional statistical approaches had similar performance.

Third, some (5/46, 11%) of the studies comparing the performances of different AI models yielded mixed results, reflecting the interdependence between model and data or task; for example, logistic regressions were reported to achieve higher prediction accuracy than DTs, naïve Bayes (NB) [29], and DL [35]. By contrast, Heydari et al [67] found that logistic regressions and DL models performed equally well in solving classification problems. Zhang et al [68] and Ergün [70] reported that data mining and DL techniques outperformed logistic regressions in classification accuracy.

Fourth, newer studies increasingly adopted state-of-the-art DL models to address CV and NLP tasks; for example, chatbots built on NLP models were used to support pediatric obesity treatment [48]. CNN-based CV models were used to construct indicators for the built environment using images from Google Street View [42]. DL-based tools were used to efficiently visualize and analyze abdominal visceral adipose tissue and subcutaneous adipose tissue [38].

Table 3. Estimated effects and main findings of the studies included in the review.

Authors, year	Estimated effects of AI[a] technologies on obesity prevention or treatment	Main findings
Abdel-Aal and Mangoud [26], 1997	• Models for WHR[b] as a continuous variable predict the actual values within an error rate of 7.5% at the 90% confidence limits. • Categorical models predict the correct logical value of WHR with an error in only 2 of the 300 evaluation cases. • Analytical relationships derived from simple categorical models explain global observations on the total survey population to an accuracy rate as high as 99%. • Simple continuous models represented as analytical functions highlight global relationships and trends. • There is a strong correlation between WHR and diastolic blood pressure, cholesterol level, and family history of obesity.	• Compared with other statistical and neural network approaches, AIM[c] abductive networks provide a faster and more automated model synthesis.
Positano et al [71], 2008	• CV[d] values in VAT[e], SAT[f], and VAT/SAT ratio assessment by the standard algorithm without image inhomogeneities correction were 10.7%, 11.9%, and 17.3%, respectively. Correlation coefficients were r=0.97, r=0.93, and r=0.95, respectively (all *P*<.001). • When correction for field inhomogeneities was applied, VAT, SAT, and VAT/SAT ratio CVs became 9.8%, 6.7%, and 13.1%, respectively. Correlation coefficients became r=0.97, *P*<.001 for VAT; r=0.99, *P*<.001 for SAT; and r=0.97, *P*<.001 for VAT/SAT ratio.	• The CV between manual and unsupervised analyses was significantly improved by inhomogeneities correction in SAT evaluation. Systematic underestimation of SAT was also corrected. A less critical performance improvement was found in VAT measurement. • The compensation of signal inhomogeneities improves the effectiveness of the unsupervised assessment of abdominal fat. • Correction of intensity distortions is necessary for SAT evaluation but less significant in VAT measurement.
Ergün [70], 2009	• The classification rate of neural networks in obesity is 90.2%, and the classification rate of logistic regression in obesity is 87.8%. • After these classifications, in obesity, the BMI is more affected than the divergent arteries.	• The classifying performance of a neural network is better than that of logistic regression.
Yang et al [69], 2009	• The implemented method achieved the macroaveraged F-measure of 81% for the textual task and 63% for the intuitive task. The microaveraged F-measure showed an average accuracy of 97% for textual annotations and 96% for intuitive annotations.	• Text mining may provide an accurate and efficient prediction of disease statuses from clinical discharge summaries.
Zhang et al [68], 2009	• Prediction at 8 months' accuracy is improved very slightly, in this case by using neural networks, whereas for prediction at 2 years, the obtained accuracy is enhanced by >10%, in this case by using Bayesian methods.	• SVM[g] and Bayesian algorithms seem to be the best algorithms for predicting overweight and obesity from the Wirral database. • The incorporation of nonlinear interactions could be important in childhood obesity prediction. Data mining techniques are becoming sufficiently well established to offer the medical research community a valid alternative to logistic regression.
Heydari et al [67], 2012	• Regarding logistic regression and neural networks, the respective values were 80.2% and 81.2% for correct classification 80.2% and 79.7% for sensitivity, and 81.9% and 83.7% for specificity; the values for the area under the receiver operating characteristic curve were 0.888 and 0.884, respectively, and the values for the kappa statistic were 0.600 and 0.629, respectively. • Abdominal thickness, weight, BMI, and HC[h] were significantly associated with obesity.	• Neural networks and logistic regression were good classifiers for obesity detection but were not significantly different with regard to classification.
Kupusinac et al [66], 2014	• The predictive accuracy of an ANN[i] solution is 80.43%. • ANN showed higher predictive accuracy ranging from +1.23% to +3.12%.	• An ANN is a new approach to predicting BFP[j] with the same complexity and costs but with higher predictive accuracy.

Authors, year	Estimated effects of AI[a] technologies on obesity prevention or treatment	Main findings
Shao [65], 2014	• Although the 13 body circumference measurements are involved in the real data set, the proposed models can provide better predictions with fewer body circumference measurements. It is much more convenient to predict BFP with fewer body circumference measurements for most people.	• Compared with traditional single-stage approaches, the proposed hybrid models—multiple regression, ANN, multivariate adaptive regression splines, and support vector regression techniques—can effectively predict BFP.
Chen et al [64], 2015	• The most important correlated indexes are creatinine, hemoglobin, hematocrit, uric acid, red blood cells, high-density lipoprotein, alanine transaminase, triglyceride, and γ-glutamyl transpeptidase.	• The ELM[k] performs much more efficiently than the SVM and BPNN[l] and with higher recognition rates. • The proposed ELM-based approach for overweight detection in biomedical applications holds promise as a new, accurate method for identifying participants' overweight status. It provides a viable alternative to traditional overweight modeling tools by offering excellent predictive ability.
Dugan et al [63], 2015	• The ID3[m] model trained on the CHICA[n] data set demonstrated the best overall performance with an accuracy of 85% and sensitivity of 89%. In addition, the ID3 model had a positive predictive value of 84% and a negative predictive value of 88%. • Being overweight between the ages of 12 and 24 months is a key risk factor for obesity after the second birthday. Furthermore, it is more of a risk factor if the child was not overweight before 12 months.	• Data from a production clinical decision support system can be used to build an accurate ML[o] model to predict obesity in children after the age of 2 years.
Nau et al [62], 2015	• After examining 44 community characteristics, the researchers identified 13 features of the social, food, and physical activity environment that, in combination, correctly classified 67% of communities as obesoprotective or obesogenic using the mean BMI z score as a surrogate. Social environment characteristics emerged as the most critical classifiers and might leverage intervention.	• CRF[p] allows consideration of the neighborhood as a system of risk factors.
Almeida et al [61], 2016	• All BFP-grade predictive models presented a good global accuracy (≥91.3%) for obesity discrimination. Both overfat and obese as well as obese prediction models showed, respectively, good sensitivity (78.6% and 71%), specificity (98% and 99.2%), and reliability for positive or negative test results (≥82% and ≥96%). • For boys, the order of parameters, by relative weight in the predictive model, was BMI z score, height, WHtR[q] squared variable (_Q), age, weight, CC[r]_Q, and HC[s]_Q (adjusted R^2=0.847 and RMSE[t]=2.852); for girls, it was BMI z score, WHtR_Q, height, age, HC_Q, and CC_Q (adjusted R^2=0.872 and RMSE=2.171).	• BFP can be graded and predicted with relative accuracy from anthropometric measurements (excluding skinfold thickness). Fitness and cross-validation results showed that the multivariable regression model performed better in this population than in some previously published models.
Lingren et al [60], 2016	• Overall, the rule-based algorithm performed the best: 0.895 (CCHMC[u]) and 0.770 (BCH[v]).	• The rule-based exclusion algorithm performed better than the ML algorithm. The best feature set for ML used Unified Medical Language System concept unique identifiers; International Classification of Diseases, Ninth Revision, codes; and R[x]Norm codes.
Seyednasrollah et al [59], 2017		• WGRS19 improves the prediction of adulthood obesity. The model helps screen children with a high risk of developing obesity. Predictive accuracy is highest among young children (aged 3-6 years), whereas among older children (aged 9-18 years), the risk can be identified using childhood clinical factors.

Authors, year	Estimated effects of AI[a] technologies on obesity prevention or treatment	Main findings
	• Replication in the BHSw confirmed the researchers' findings that WGRSx19 and WGRS97 are associated with BMI. WGRS19 improved the accuracy of predicting adulthood obesity in the training data (area under the curve=0.787 vs area under the curve=0.744; *P*<.001) and validation data (area under the curve=0.769 vs area under the curve=0.747; *P*=.03). WGRS97 improved the accuracy in the training data (area under the curve=0.782 vs area under the curve=0.744; *P*<.001) but not in the validation data (area under the curve=0.749 vs area under the curve=0.747; *P*=.79). Higher WGRS19 is associated with a higher BMI at 9 years and WGRS97 at 6 years.	
Hinojosa et al [58], 2018	• Violent crime, English learners, socioeconomic disadvantage, fewer physical education and fully credentialed teachers, and diversity index were positively associated with obesity. By contrast, the academic performance index, physical education participation, mean educational attainment, and per capita income were negatively associated with obesity. The most highly ranked built or physical environment variables were distance to the nearest highway and green spaces, 10th and 11th most important, respectively.	• An RFy algorithm effectively identifies the relative importance of school environment attributes.
Maharana and Nsoesie [57], 2018	• Features of the built environment explained 64.8% (RMSE=4.3) of the variation in obesity prevalence across all US census tracts. Individually, the variation explained was 55.8% (RMSE=3.2) for Seattle, Washington (213 census tracts); 56.1% (RMSE=4.2) for Los Angeles, California (993 census tracts); 73.3% (RMSE=4.5) for Memphis, Tennessee (178 census tracts); and 61.5% (RMSE=3.5) for San Antonio, Texas (311 census tracts).	• CNN[z] can be used to automate the extraction of features of the built environment from satellite images for studying health indicators. Understanding the association between specific features of the built environment and obesity prevalence can lead to structural changes that could encourage physical activity and decrease obesity prevalence.
Wang et al [56], 2018	• The SVM model significantly outperformed other classifiers based on the same training features. The SVM model exhibits 70.77% accuracy, 80.09% sensitivity, and 63.02% specificity. • The selected SNPs[aa] were effective in the detection of obesity risk.	• The ML-based method provides a feasible means for conducting preliminary analyses of genetic characteristics of obesity.
Duran et al [55], 2018	• In female participants, the sensitivity of the BMI, WC[bb], and ANN approaches to predict excess body fat was 0.751 (95% CI 0.730 - 0.771), 0.523 (95% CI 0.487 - 0.559), and 0.782 (95% CI 0.754 - 0.810), respectively. • In male participants, the sensitivity of the BMI, WC, and ANN approaches to predict excess body fat was 0.721 (95% CI 0.699 - 0.743), 0.572 (95% CI 0.549 - 0.594), and 0.795 (95% CI 0.768 - 0.821).	• The diagnostic performance in identifying excess body fat was better in male participants when an ANN approach was used than when BMI and WC z scores were applied. • The ANN and BMI z scores performed comparably and significantly better, respectively, than WC z scores in female participants.
Gerl et al [54], 2019	• The lipidome, based on a LASSO[cc] model, predicted BFP the best (R^2=0.73). In this model, the strongest positive predictor and strongest negative predictor were sphingomyelin molecules, which differ by only 1 double bond, implying the involvement of an unknown desaturase in obesity-related aberrations of lipid metabolism. • The regression was used to probe the clinically relevant information in the plasma lipidome and found that the plasma lipidome also includes information on body fat distribution because WHR (R^2=0.65) was predicted more accurately than BMI (R^2=0.47).	• ML can model and validate obesity estimates better than classical clinical parameters such as total triglycerides and cholesterol.
Hammond et al [53], 2019	• LASSO regression predicted obesity with an area under the receiver operating characteristic curve of 81.7% for girls and 76.1% for boys. • In each of the separate models for boys and girls, the researchers found that the weight-for-length z score, BMI between 19 and 24 months, and the last BMI measure recorded before the age of 2 years were the most important features for prediction.	• Comparable to cohort-based studies, EHR[dd] data with area under the receiver operating characteristic curve values could be used to predict obesity at the age of 5 years, reducing the need for investment in additional data collection.
Hong et al [52], 2019		

Authors, year	Estimated effects of AI[a] technologies on obesity prevention or treatment	Main findings
	• As the results of the 4 ML classifiers showed, the RF algorithm performed the best with micro F1-score 0.9466 and macro F1-score 0.7887 and micro F1-score 0.9536 and macro F1-score 0.6524 for intuitive classification (reflecting medical professionals' judgments) and textual classification (reflecting the decisions based on explicitly reported information of diseases), respectively. • The MIMIC[ee]-III obesity data set was successfully integrated for prediction with minimal configuration of the NLP[ff]2FHIR[gg] pipeline and ML models.	• The FHIR-based EHR phenotyping approach could effectively identify the obesity status and multiple comorbidities using semistructured discharge summaries.
Ramyaa et al [51], 2019	• SVM, neural network, and KNN[hh] algorithms performed modestly for the numerical predictions, with mean approximate errors of 6.70 kg, 6.98 kg, and 6.90 kg, respectively. • K-means cluster analysis improved prediction using numerical data and identified 10 clusters suggestive of phenotypes, with a minimum mean approximate error of approximately 1.1 kg. A classifier was used to phenotype participants into the identified clusters, with mean approximate errors of <5 kg for 15% of the test set (approximately, n=2000). SVM performed the best (54.5% accuracy), followed closely by the bagged tree ensemble and KNN algorithms.	• SVM regression was the best-suited predictive and inferential tool for this task, closely followed by neural network and KNN algorithms. Although the overall data model showed a good fit and predictive ability, clustering produced relatively superior fit statistics.
Scheinker et al [50], 2019	• Multivariate linear regression and gradient boosting machine regression (the best-performing ML model) of obesity prevalence using all county-level demographic, socioeconomic, health care, and environmental factors had R^2 values of 0.58 and 0.66, respectively (P<.001).	• ML may be used to explain more variation in county-level obesity prevalence than traditional epidemiologic models. The top-performing ML model explained two-thirds of the variation in county-level obesity prevalence, significantly more than conventional multivariate linear models.
Shin et al [49], 2019	• The performance of the proposed system was compared with those of 2 commercial systems that were designed to measure body composition using either a whole body or upper body impedance value. The results showed that the correlation coefficient (R^2) value was improved by approximately 9%, and the SE of the estimate was reduced by 28%.	• The test results validated that the inclusion of anthropometric data helped to improve accuracy, primarily when a DL[ii] approach was used to predict the regression values.
Stephens et al [48], 2019	• Adolescent patients reported experiencing positive progress toward their goals 81% of the time. The 4123 messages exchanged and patients' reported usefulness ratings (96% of the time) illustrate that adolescents engaged with the chatbot and viewed it as helpful.	• An AI chatbot is feasible as an adjunct to treatment. The feasibility and benefit of support through AI, specifically in a pediatric setting, could be scaled to serve larger groups of patients.
Blanes-Selva et al [47], 2020	• The PU[jj] learning algorithm presented a high sensitivity (98%) and predicted that approximately 18% of the patients without a diagnosis were obese.	• The implementation of the PU learning methodology in identifying obesity produced results that were satisfactory, providing high sensitivity, and consistent with the World Health Organization's obesity report.
Dunstan et al [46], 2020	• Using only 5 categories, RF could predict obesity prevalence with absolute error <10% for approximately 60% of the countries considered and absolute error <20% for 87%. • The most relevant food category with regard to predicting obesity consists of baked goods and flours, followed by cheese and carbonated drinks.	• RF shows the best performance for predicting obesity from food, followed closely by XGB[kk].
Fu et al [45], 2020		

Authors, year	Estimated effects of AI[a] technologies on obesity prevention or treatment	Main findings
	• The 2 most important features—trajectory of infant BMI z score change and maternal BMI at enrollment—were identified from the ML algorithm. • The aforementioned features showed similar predictive capacity compared with all features (area under the curve=0.68 vs 0.68; *P*=.83; DeLong test). The sensitivity analyses identified the same 2 features (ie, trajectory of infant BMI z score change and maternal BMI at enrollment), and the ranking of these features' Shapley additive explanations value was unchanged. • In the independent test cohort, the area under the curve for childhood overweight and obesity classification using the aforementioned 2 features was 0.71 (95% CI 0.66 to 0.76), which was comparable to that based on all features (0.72, 95% CI 0.67 to 0.76).	• An ML algorithm is applied to identify risk factors contributing to childhood overweight or obesity based on a large longitudinal study and addresses the relationships between all collected features and outcomes without any assumption. • A novel unified framework, Shapley additive explanations, is used to interpret predictions, and the identified predictive factors are robust.
Kibble et al [44], 2020	• New potential links between cytokines and weight gain are identified, as well as associations among dietary, inflammatory, and epigenetic factors.	• An integrative ML method called group factor analysis was used to identify the links between multimolecular-level interactions and the development of obesity.
Park et al [43], 2020	• The actual and predicted ΔBMI showed a significant intraclass correlation value with a low RMSE, and classification between people with increased BMI and those with nonincreased BMI resulted in a high area under the receiver operating characteristic curve value using only the degree centrality values obtained at the baseline visit.	• The constructed model using functional connectivity of the selected regions provides robust neuroimaging biomarkers for predicting BMI progression.
Phan et al [42], 2020	• A DNN[ll] was used for neighborhood indicator recognition and achieved high accuracies (85%-93%) for the separate recognition tasks.	• DL techniques were used to create indicators for neighborhood-built environment characteristics.
Taghiyev [41], 2020	• The proposed hybrid system demonstrated 91.4% accuracy, which is higher than that of other classifiers (ie, 4.6% higher than the performance of logistic regression and 2.3% higher than the performance of DT[mm]).	• The proposed hybrid system provides a more accurate classification of patients with obesity and a practical approach to estimating the factors affecting obesity.
Xiao et al [40], 2020	• All aspects of horizontal greenery, vertical greenery, and proximity of green levels affected body weight; however, only the VGI[nn] consistently had an adverse effect on weight and obesity.	• The VGI of the DL approach using Baidu Street View images could effectively capture the eye-level greenness in high-density–population areas. Thus, VGI can be used to effectively promote walking and other physical activities to prevent obesity.
Yao et al [39], 2020	• Jogging may be a more suitable activity of daily living for BMI prediction than walking and walking up stairs.	• The proposed DL model with the motion entropy–based filtering strategy outperforms the baseline approaches significantly.
Alkutbe et al [27], 2021	• For the gradient boosting models, the predicted fat percentage values were more aligned with the actual value than those in regression models. Gradient boosting achieved better performance than the regression equation because it combined multiple simple models into a single composite model to take advantage of this weak classifier. • The developed predictive model archived RMSE values of 3.12 for girls and 2.48 for boys.	• ML models and newly developed centile charts could be valuable tools for estimating and classifying BFP.
Bhanu et al [38], 2021	• The accuracy of segmentation was superficial SAT: 0.92, deep SAT: 0.88, and VAT: 0.9. The average Hausdorf distance was <5 mm. Automated segmentation significantly correlated R^2>0.99 (*P*<.001) with ground truth for all 3-fat compartments. Predicted volumes were within 1.96 SD from Bland-Altman analysis.	• DL-based, comprehensive superficial SAT, deep SAT, and VAT analysis tools showed high accuracy and reproducibility and provided a comprehensive fat compartment composition analysis and visualization in <10 seconds.
Cheng et al [37], 2021		

Authors, year	Estimated effects of AI[a] technologies on obesity prevention or treatment	Main findings
	• Physical activity was an important factor in predicting weight status, with gender, age, and race or ethnicity being less important factors associated with weight outcomes. • The durations of vigorous-intensity activity in 1 week and moderate-intensity activity in 1 week were essential attributes.	• With physical activity and basic demographic information of all methods analyzed, the random subspace classifier algorithm achieved the highest overall accuracy and area under the receiver operating characteristic curve value. • In general, most algorithms showed similar performance. • Logistic regression was middle ranking in terms of overall accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve value among all methods.
Delnevo et al [36], 2021	• The psychological variables in use allow one to predict both BMI values (with a mean absolute error of 5.27-5.50) and BMI status with an accuracy of >80% (metric: F1-score).	• Certain psychological variables such as depression are highly predictive of BMI. • ML has several advantages over traditional statistics and can be used to compare the impact of many variables on predicting a chosen outcome and can handle various types of variables.
Lee et al [35], 2021	• For predicting a newborn's BMI, linear regression (2.0744) and RF (2.1610) were better than ANN with 1, 2, and 3 hidden layers (150.7100, 154.7198, and 152.5843, respectively) in the mean squared error. • On the basis of variable importance from the RF, the major predictors of a newborn's BMI were the first abdominal circumference value and estimated fetal weight in week 36 or later, gestational age at delivery, the first abdominal circumference value during week 21 to week 35, maternal BMI at delivery, maternal weight at delivery, and the first biparietal diameter value in week 36 or later.	• ML approaches based on ultrasound measures would be a useful noninvasive tool for predicting a newborn's BMI. • Linear regression and RF were better models than ANNs for predicting a newborn's BMI.
Lin et al [34], 2021	• ML revealed the following 4 stable metabolically distinct obesity clusters in each cohort: • Metabolic healthy obesity (44% of the patients) was characterized by a relatively healthy metabolic status with the lowest incidents of comorbidity. • Hypermetabolic obesity–hyperuricemia (33% of the patients) was characterized by extremely high uric acid and an increased incidence of hyperuricemia (adjusted odds ratio 73.67 to metabolic healthy obesity, 95% CI 35.46-153.06). • Hypermetabolic obesity–hyperinsulinemia (8% of the patients) was distinguished by overcompensated insulin secretion and an increased incidence of polycystic ovary syndrome (adjusted odds ratio 14.44 to metabolic healthy obesity, 95% CI 1.75-118.99). • Hypometabolic obesity (15% of the patients) was characterized by extremely high glucose levels, decompensated insulin secretion, and the worst glucolipid metabolism (diabetes: adjusted odds ratio 105.85 to metabolic healthy obesity, 95% CI 42.00-266.74; metabolic syndrome: adjusted odds ratio 13.50 to metabolic healthy obesity, 95% CI 7.34-24.83). • The assignment of patients in the verification cohorts to the main model showed a mean accuracy of 0.941 in all clusters.	• ML automatically identified 4 subtypes of obesity in clinical characteristics in 4 independent patient cohorts. This proof-of-concept study provided evidence that a precise diagnosis of obesity can potentially guide therapeutic planning and decisions for different subtypes of obesity.
Pang et al [33], 2021	• XGB yielded a mean area under the curve value of 0.81 (SD 0.001), which outperformed all other models. It also achieved a statistically significant better performance than all other models on standard classifier metrics (sensitivity fixed at 80%): precision, mean 30.9% (SD 0.22%); F1-score, mean 44.6% (SD 0.26%); accuracy, mean 66.14% (SD 0.41%); and specificity, mean 63.27% (SD 0.41%).	• The presented ML model development workflow can be adapted to various EHR-based studies and is valuable for developing other clinical prediction models.
Park et al [32], 2021		

Authors, year	Estimated effects of AI[a] technologies on obesity prevention or treatment	Main findings
	• ML algorithms were used to determine the stances of tweets on Black Lives Matter. ML models showed better performance than lexicon-based sentiment analysis (accuracy: 61%). The NB[oo] model had an overall accuracy of 85%, slightly higher than that of the CNN model (83.8%); both had higher accuracy than the other models. • However, NB had the highest recall and F1-score for predicting the against stance, whereas CNN performed poorly on identifying the against stance.	• The study demonstrated the strengths of ML techniques in handling large data sets. Social scientists can use ML techniques to scale up traditional content analysis.
Rashmi et al [31], 2021	• The PCA[pp] method provides the best classification accuracy for SVM (98%), followed by NB and RF (97%).	• The regional thermography and computer-aided diagnostic tool with ML classifier could be used as a primary noninvasive prognostic tool for evaluating obesity in children.
Snekhalatha and Sangamithirai [30], 2021	• Among the region of interest studied, the abdomen region exhibited a high temperature difference of 4.703% between normal participants and participants who were obese compared with other regions. The proposed custom network-2 provided an overall accuracy of 92%, with an area under the curve value of 0.948. By contrast, the pretrained model VGG16 produced an accuracy of 79% and an area under the curve value of 0.90 for discrimination into obese and normal thermograms.	• The DL system based on custom CNN provided a reliable classification performance to identify the occurrence of obesity in test participants. • Custom CNN network-2 provided a commendable accuracy in classifying normal participants and participants who were obese from the thermal images. • The trained custom-2 CNN model can be used for computer-aided screening of test participants for obesity detection.
Thamrin et al [29], 2021	• Location, marital status, age group, education, sweet drinks, fatty or oily foods, grilled foods, preserved foods, seasoning powders, soft drinks or carbonated beverages, alcoholic beverages, mental or emotional disorders, diagnosed hypertension, physical activity, smoking, and fruit and vegetable consumption are significant in predicting obesity status in adults. • The classification prediction using the logistic regression method achieves the best performance based on the accuracy metric (72%), specificity (71%), precision (69%), kappa (44%), and Fβ-score (70%). Classification prediction by the classification and regression tree method achieves the highest sensitivity (82%) and the highest F1-score (72%). • With regard to the area under the receiver operating characteristic curve performance of the respective classification methods with 10-fold cross-validation, the logistic regression classifier has the highest average area under the receiver operating characteristic curve value (0.798).	• Logistic regression has a better performance than the classification and regression tree and NB methods. • Kappa coefficients show only moderate concordance between predicted and measured obesity. • The constructed obesity classification model can evaluate and predict the risk of obesity using ML methods for the population of Indonesia, which can then be applied to publicly available open data.
Zare et al [28], 2021	• The kindergarten BMI z score is the most important predictor of obesity by grade 4. • Including the kindergarten BMI z score of students in the model meaningfully increases the prediction accuracy. • Logistic regression, RF, and neural network algorithms performed similarly in terms of accuracy, sensitivity, specificity, and area under the curve values. The 95% CIs around the area under the curve overlap among these 3 algorithms. • The DT showed lower performance with an area under the curve value that was statistically lower than the area under the curve values from each of the other algorithms. Nevertheless, the performance of the DT algorithm was close to that of the others.	• Data from the Arkansas, United States, BMI screening program significantly improve the ability to identify children at a high risk of obesity to the extent that better prediction can be translated into more effective policy and better health outcomes. • The ability to predict obesity by grade 4 was robust across the ML algorithms and logistic regression with these data.

[a]AI: artificial intelligence.

[b]WHR: waist-to-hip ratio.

[c]AIM: abductory induction mechanism.

[d]CV: coefficient of variation.

[e]VAT: visceral adipose tissue.

[f]SAT: subcutaneous adipose tissue.
[g]SVM: support vector machine.
[h]HC: hip circumference.
[i]ANN: artificial neural network.
[j]BFP: body fat percentage.
[k]ELM: extreme learning machine.
[l]BPNN: back propagation neural network.
[m]ID3: iterative dichotomizer 3.
[n]CHICA: Child Health Improvement Through Computer Automation.
[o]ML: machine learning.
[p]CRF: conditional random forest.
[q]WHtR: waist-to-height ratio.
[r]CC: calf circumference.
[s]HC: hip circumference.
[t]RMSE: root mean square error.
[u]CCHMC: Cincinnati Children's Hospital and Medical Center.
[v]BCH: Boston Children's Hospital.
[w]BHS: Bogalusa Heart Study.
[x]WGRS: weighted genetic risk score.
[y]RF: random forest.
[z]CNN: convolutional neural network.
[aa]SNP: single-nucleotide polymorphism.
[bb]WC: waist circumference.
[cc]LASSO: least absolute shrinkage and selection operator.
[dd]EHR: electronic health record.
[ee]MIMIC: Multiparameter Intelligent Monitoring in Intensive Care.
[ff]NLP: natural language processing.
[gg]FHIR: Fast Healthcare Interoperability Resources.
[hh]KNN: k-nearest neighbor.
[ii]DL: deep learning.
[jj]PU: positive and unlabeled.
[kk]XGB: extreme gradient boosting.
[ll]DNN: deep neural network.
[mm]DT: decision tree.
[nn]VGI: Visible Green Index.
[oo]NB: naïve Bayes.
[pp]PCA: principal component analysis.

Methodological Review

AI Overview

AI symbolizes the effort to automate intellectual tasks usually performed by humans [72]. In general, AI consists of 2 domains or developmental periods: symbolic AI and modern AI [73]. Symbolic AI prevailed from the 1950s to the 1980s, characterized by the endeavors to achieve human-level intelligence by having programmers handcraft a sufficiently large set of explicit rules for manipulating knowledge [74]. Although symbolic AI proved suitable for solving well-defined, logical problems, such as a rule-based question-answer system, it became intractable when creating rules to solve more complex, fuzzy issues such as image classification, speech recognition, and language translation [74]. The definition of ML is "the field of study that gives computers the ability to learn without being explicitly programmed" [75]. Instead of hard coding all the rules in the symbolic AI, researchers provide examples (eg, images with labels that identify the objects in them) to *train* modern ML models to output rules [74]. As a subdomain of ML, DL is based on artificial neural networks in which multiple (*deep*) layers of artificial neurons are used to progressively extract higher-level features from data [76]. This layered representation enables the modeling of more complex, dynamic patterns compared with traditional ML (which sometimes is called *shallow learning* in contrast to DL), which finds its utility in analyzing *big data*: data massive in scale and *messy* to work with (eg, unstructured texts and images) [77]. The first ML and DL algorithms were developed in the 1950s, attracting initial excitement but then lying dormant for several decades [72]. Since the late 1980s, partly because of the rediscovery of backpropagation algorithms, the invention of CNNs, and the strong growth in computational capacity, ML and DL have regained their popularity vis-à-vis symbolic AI [72].

AI Versus Conventional Statistical Methods

Admittedly, the concept of conventional statistical methods is dubious at best because the development of statistical theories and algorithms is continual in time and intertwines at all levels [78]. Indeed, many *conventional* models fall into the ML domain, such as linear and logistic regressions. Despite the poorly defined domain and overlapping algorithms, at least 2 distinctions could be made between modern AI (ie, ML and DL) and other statistical methods. In terms of aims, the objective of AI models and their evaluation metrics predominantly concern prediction precision (often at the cost of compromising interpretability as models become complex) [78,79]. By contrast, conventional statistical approaches usually attempt to reveal relationships among variables (*statistical inference*) and focus on model interpretability [80]. In terms of procedures, it is standard practice to split data into training, validation, and test sets so that an AI model can be trained using the training set with the aim of achieving the optimal performance on some predefined evaluation metrics (eg, accuracy and mean squared error) when testing on the validation set [81,82]. The fine-tuned AI model is subsequently tested on the test set. The utility of the validation set is to prevent model overfitting (ie, too tailored to the training set while losing generalizability to new, unseen data) and fine-tune hyperparameters (ie, parameters external to the model, whose values cannot be automatically learned from data). The test set is preserved to test the final model's performance on unseen data. By contrast, conventional statistical methods do not usually fit and evaluate models using training, validation, and test sets but use other model selection criteria (eg, adjusted R-squared and Akaike and Bayesian information criteria) to evaluate model performance [83].

ML Subcategories

Overview

ML is classified into 2 subcategories: unsupervised ML and supervised ML [84]. Unsupervised ML analyzes and clusters unlabeled data sets, discovering hidden patterns or data groupings without the need for human intervention [85]. Its capability to reveal similarities and differences in information makes it ideal for exploratory data analysis. Unsupervised ML models are used for 3 main tasks: clustering, association, and dimensionality reduction [86]. Clustering algorithms (eg, k-means clustering, hierarchical clustering, and Gaussian mixture) group unlabeled data based on similarities [86]. Association algorithms (eg, Apriori, Eclat, and FP-Growth) identify rules and relations among variables in large databases [87]. Dimensionality reduction algorithms (eg, principal component analysis [PCA], singular value decomposition, and multidimensional scaling) deal with an excessive number of features during data preprocessing, reducing them to a manageable size while preserving the integrity of the data set as much as possible [88]. Supervised ML uses a training set consisting of input-output pairs to enable the algorithm to learn a function that maps input to output over time [89]. The algorithm measures its accuracy through the loss function, adjusting until the error is minimized sufficiently. The critical difference between supervised ML and unsupervised ML is that the former requires labeled data (ie, input-output pairs), whereas the latter only requires inputs (ie, unlabeled data) [84]. Supervised ML models are used for 2 main tasks: classification and regression [84]. Classification algorithms assign data to specific categories (eg, obese or nonobese). Regression algorithms learn the relationship between input features and continuously distributed outcomes and are commonly used for projections (eg, BMI in 5 years).

Unsupervised ML

K-means Clustering

K-means clustering is an iterative algorithm that tries to partition the data set into a total of *k* nonoverlapping groups (ie, clusters) [86,90]. Each data point belongs to only 1 group. The algorithm attempts to make the intracluster data points as similar as possible while keeping the clusters apart. In particular, it assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (ie, arithmetic mean of all the data points belonging to that cluster) is minimized. As the number of clusters *k* needs to be determined before implementing the algorithm, silhouette coefficients are commonly used to identify the optimal *k* value. Lin et al [34] used k-means clustering to classify patients with obesity into 4 groups based on 3 biomarkers concerning glucose, insulin, and uric acid.

Fuzzy C-means Clustering

In nonfuzzy clustering (also known as hard clustering; for example, k-means clustering), data are divided into distinct clusters, where each data point can only belong to 1 cluster [86]. In fuzzy clustering, data points can potentially belong to multiple clusters [91]. Fuzzy c-means clustering assigns each data point membership from 0% to 100% in each cluster center [92]. The fuzzy partition coefficient is often used to determine the optimal number of clusters with a value ranging from 0 (worst) to 1 (best) [93]. Positano et al [71] used the fuzzy c-means algorithm to classify MRI pixels into clusters to assess abdominal fat.

Group Factor Analysis

Factor analysis describes relationships among the individual variables of a data set [94]. Group factor analysis (GFA) extends this classical formulation into describing relationships among groups of variables, where each group represents either a set of related variables or a data set [95]. GFA is commonly formulated as a latent variable model consisting of 2 hierarchical levels: the higher level models the relationships among the groups, and the lower-level models the observed variables given the higher level [95]. Kibble et al [44] used GFA to jointly analyze 5 large multivariate data sets to understand the multimolecular-level interactions associated with obesity development.

PCA for Large Data Sets

Large data sets are increasingly common nowadays. PCA is a classic, widely adopted method to reduce the dimensionality of a large data set while preserving as much statistical information (ie, variability) as possible [86]. In particular, PCA attempts to find new variables, called principal components, that are linear functions of those in the original data set. The new variables are uncorrelated with each other (ie, orthogonal) and maximize the projected data variance. Rashmi et al [31] used PCA to

reduce the feature dimensions of a thermal imaging data set to classify children by their obesity severity level.

Supervised ML

Linear Regression

Linear regression is considered a conventional statistical model and a classical architecture to develop a predictive model [96], but it fulfills all criteria from an ML point of view and is widely used as an ML algorithm to predict continuous outcomes such as BMI or BFP [97]. Trainable weights (ie, coefficients) of linear regression are commonly estimated using ordinary least squares or gradient descent. Compared with many other ML models, linear regression has the advantages of simplicity and interpretability [98]. It is easy to understand how the model reaches its predictions. Wang et al [56] used linear regressions to identify features of single-nucleotide polymorphisms that predict obesity risk. Phan et al [42] used linear regressions to estimate the associations between built environment indicators and state-level obesity prevalence.

Regularized Linear Regression

The bias-variance tradeoff is a fundamental issue faced by all ML models [86,99]. Bias is an error from erroneous assumptions in a learning algorithm. High bias may cause the algorithm to miss the relevant relations between features and outputs (called underfitting). Variance is an error from a learning algorithm's sensitivity to small fluctuations in the training set. A high variance may result from the algorithm modeling the random noise in the training data, often leading to the algorithm's poor generalizability to new, unseen data (called overfitting). In general, decreasing variance increases bias and vice versa, and ML algorithms need to be fine-tuned to balance these 2 properties. Regularization is an essential technique to prevent model overfitting and improve generalizability (at the cost of increasing bias) by adding a penalty term of trainable weights to the loss function [86]. Optimization algorithms that minimize the loss function will learn to avoid extreme weight values and thus reduce variance. The penalty term with the sum of squared trainable weights is called L2 regularization, used in Ridge regression. The penalty term with the sum of the absolute values of trainable weights is called L1 regularization, used in the least absolute shrinkage and selection operator (LASSO) regression. Unlike Ridge regression, LASSO regression often shrinks some feature weights to absolute zero, making it useful for feature selection. Finally, ElasticNet regression uses a weighted sum of L1 and L2 regularizations. Gerl et al [54] used LASSO regression to estimate the relationship between human plasma lipidomes and body weight outcomes, including BMI, WC, WHR, and BFP.

Logistic Regression

In its simplest form, logistic regression uses a logistic function, called the sigmoid function, to model a binary outcome [100]. A sigmoid function is a continuous, smooth, differentiable S-shaped mathematical function that maps a real number to a value in the range of 0 and 1, making it ideal for modeling probabilities. The estimated probabilities are converted to predictions (0 or 1, denoting exclusive group membership) based on some predefined threshold (eg, >0.5). In ML, logistic regression often incorporates regularizations (L1, L2, or both) to prevent overfitting. Another common extension of logistic regression in ML is to solve multiclass classification problems when classification tasks involve >2 (exclusive) classes. A typical strategy uses the one-vs-rest method (also called one-vs-all) that fits 1 classifier (eg, a logistic regression) per class against all the other classes [101]. A data point is assigned to the class with the highest confidence score among all classifiers. Thamrin et al [29] used logistic regressions to assess the predictability of various obesity risk factors. Cheng et al [37] used logistic regressions to classify obesity status based on participants' physical activity levels.

NB Classifier

NB algorithms apply the Bayes theorem with the *naïve* assumption of conditional independence among each pair of features given the value of the class [102]. Despite this oversimplified assumption, NB classifiers have been widely used and have worked well in solving many real-world problems. The decoupling of conditional feature distributions allows each distribution to be independently estimated as 1D, making the training of NB classifiers much faster than more sophisticated ML models [86]. By contrast, the predicted probabilities of NB classifiers are less trustworthy owing to the algorithm's *naïve* assumption. Rashmi et al [31] used NB to classify childhood obesity based on thermogram images. Thamrin et al [29] adopted NB to predict adult obesity using Indonesian health survey data [29].

K-nearest Neighbor

K-nearest neighbor (KNN) is a nonparametric, supervised learning algorithm suitable for classification and regression tasks [103]. The input consists of the *k* closest training data points based on a prespecified distance measure (eg, Euclidean, Manhattan, or Minkowski distance). For classification tasks, the output is a class membership. A test data point is assigned to the class most common among its k-nearest neighbors (if k=1, the test data point is assigned to the class of the single nearest neighbor). For regression tasks, the output is the average value of its k-nearest neighbors. KNN should not be confused with k-means. The former is a supervised ML algorithm to determine the class or value of a data point based on its k-nearest neighbors, whereas the latter is an unsupervised ML algorithm to classify data points into *k* clusters that minimize the distances within clusters while maximizing those between clusters [90]. KNN is a memory-based learning algorithm that requires no training (called a lazy learner) but can become significantly slower when the sample size increases. Wang et al [56] used KNN to predict obesity risk based on features of single-nucleotide polymorphisms. Ramyaa et al [51] performed KNN to predict body weight using physical activity and dietary data.

Support Vector Machines

Support vector machines (SVMs), which are supervised learning models that construct a hyperplane in a high-dimensional space, can be used for classification and regression tasks [104]. SVMs attempt to identify the hyperplane separating different classes while maximizing the distance to any class's nearest training data point (ie, margin). Intuitively, the larger the margin, the

more likely the model's generalizability to new, unseen data. The choice of margin type can be critical for SVMs [86]. Hard-margin SVMs maximize the margin by minimizing the distance from the decision boundary to the training points. However, hard-margin SVMs may lead to overfitting and have no solution if the training data are linearly inseparable. Soft-margin SVMs modify the constraints of the hard-margin SVMs by allowing some data points to violate the margin (ie, misclassified). In practice, data are seldom linearly separable in the original feature space, and kernel methods are applied to map the input space of the data to a higher-dimensional feature space where linear models can be trained [105]. Many kernel functions, such as the Gaussian radial basis, sigmoid, and polynomial kernel, can be chosen. Wang et al [56] used SVM to predict obesity risk based on the features of single-nucleotide polymorphisms. Ramyaa et al [51] applied SVM to predict body weight using physical activity and diet data.

DT Algorithms

DTs are nonparametric supervised learning methods for classification and regression tasks [106]. In DT algorithms, a tree is built by splitting the source set that constitutes the tree's root node into subsets, which comprise the successor children [107]. The splitting is based on a set of rules applied to input features. Different splitting rules exist, such as variance reduction for regression tasks and Gini impurity or information gain for classification tasks. The splitting process is repeated on each derived subset recursively (ie, recursive partitioning). The recursion is completed when all subsets at a node share the same target value or when splitting no longer adds value to the predictions. DTs have several advantages over other ML algorithms, such as high transparency and interpretability and few requirements for data preprocessing [108]. However, DTs can be prone to overfitting (ie, too confident about the rules learned from the training set, which does not generalize well to the test set) and instability (minor variations in the data resulting in a very different tree). Using features extracted from electronic medical records, Hong et al [52] used DTs to predict obesity and 15 other comorbidities. Taghiyev et al [41] performed DTs to identify risk factors associated with obesity onset.

RF Models

Ensemble methods are approaches that aggregate the predictions of a group of models aiming for improved performance in classification or regression tasks [109]. Various ensemble methods exist, such as bagging, pasting, boosting, and stacking [86]. Bagging and pasting use the same training algorithm for every predictor included in the ensemble and train it on different random subsets of the training set. When sampling is performed with replacement, the method is called bagging; when sampling is performed without replacement, it is called pasting. RF is an ensemble of DTs commonly trained via the bagging or pasting method [110]. Specifically, RF fits many DTs on various subsets of the data and uses averaging to improve the predictive accuracy and prevent overfitting. For classification tasks, the RF output is the class selected by most trees; for regression tasks, the mean prediction of the individual trees is used. Some common hyperparameters of RF for fine-tuning include the number of trees in the forest, the maximum number of features considered for splitting a node, the maximum number of branches in each tree, the minimum number of data points placed in a node before the node is split, the minimum number of data points allowed in a leaf node, and the method for sampling data points (ie, with or without replacement) [86]. RF typically produces more accurate and robust predictions than DTs and is one of the most popular supervised ML algorithms [111]. Using RF models, Hinojosa et al [58] examined the relationship between social and physical school environments and childhood obesity in California, United States. Dunstan et al [46] performed RF to predict national obesity prevalence using food sales data from 79 countries.

Extreme Gradient Boosting

Boosting refers to any ensemble method that combines several weak models into a strong one [112]. The difference between boosting and bagging and pasting is that in boosting, different models are applied to the entire training set sequentially, the new model attempting to address the weaknesses (eg, misclassified targets and residual errors) of the previous model. By contrast, in bagging and pasting, the same models are trained on different random subsets of the training set. A popular boosting algorithm is gradient boosting, in which the new model is trained on the residual errors made by the previous model [113]. Extreme gradient boosting (XGBoost) implements an optimized, parallel-tree gradient boosting algorithm, aiming to be highly efficient, flexible, and portable [114]. XGBoost is considered one of the most powerful ML algorithms, often serving as an essential component of winning entries in ML competitions [86]. A few drawbacks of XGBoost include lacking interpretability and being prone to overfitting. Pang et al [33] used XGBoost to predict early childhood obesity based on electronic health records. Alkutbe et al [27] applied gradient boosting to predict BFP based on cross-sectional health survey data collected in Saudi Arabia.

Multivariate Adaptive Regression Splines

Multivariate adaptive regression splines (MARS) is a nonparametric regression technique that automatically models nonlinearities and interactions among variables by combining ≥2 linear regressions using hinge functions [115,116]. A hinge function is a function equal to its argument where that argument is >0 and 0 everywhere else. MARS builds a model using a 2-phase procedure [117]. The forward phase starts with a model consisting of only the intercept term (ie, mean of the target) and repeatedly adds basis functions (ie, constant or hinge function) in pairs to the model that minimizes the squared error loss of the training set. The backward (or pruning) phase usually starts with an overfitted model and removes its least effective term at each step until the best submodel is found. MARS requires little or no data preparation, is easy to understand and interpret, and can address classification and regression tasks. However, it often underperforms boosting ensemble methods. Shao [65] applied MARS to predict BFP using a small-scale health record data set.

DL Models

In the obesity literature reviewed, DL models were applied to 3 distinct data types: tabular data (eg, spreadsheet data), images, and texts. The model architectures differ systematically across these data types.

DL on Tabular Data

Although *shallow* ML models perform well on tabular data sets in most cases, some complex relationships between the features and the target could be more effectively learned by a deep neural network model [118]. A fully connected neural network consists of a series of fully connected layers, with each artificial neuron (ie, node) of a layer linking with all neurons in the following layer [76]. A multilayer perceptron (MLP) is a classic fully connected neural network consisting of at least 3 layers of neurons: an input layer, a hidden layer, and an output layer [119]. One advantage of fully connected neural networks is that they are *structure agnostic*, requiring no specific assumptions about the input. However, neural networks trained on tabular data can sometimes be prone to overfitting [120]. Park and Edington [121] used MLP to identify individuals at elevated diabetic risk. Heydari et al [67] performed MLP to predict obesity status using data from a cross-sectional study of military personnel in Iran.

DL on Images

CV is a field of AI that enables computers to learn from digital images, videos, or other visual inputs and derive meaningful information for decision-making and recommendations [122,123]. Nowadays, most CV applications use DL models, which prove more capable than their *shallow-learning* (ie, ML models) counterparts in representing and revealing high-dimensional, complex nonlinear patterns inherent in image data. Specifically, CNNs consistently outperform the traditional densely connected neural networks (eg, MLP) and achieve human-like or superhuman accuracy in many challenging CV tasks ranging from image classification to object detection and segmentation [124,125]. The main advantages of CNNs over densely connected neural networks are locality, translation invariance, and computational efficiency [126]. Locality refers to the repeated use of small-sized kernels (or filters) in CNNs to identify local patterns at an increasing level of complexity (eg, from basic shapes such as lines and edges to complex objects such as adipose tissue or brain tumor). Translation invariance refers to CNNs' capacity to detect an entity independent of its position in the image. The computational efficiency of CNNs is achieved by using kernels, global pooling, and other techniques, which typically make the models much smaller (ie, fewer learnable parameters) than their densely connected counterparts. Over the past decade, numerous CNN-based DL models were built and adopted to tackle domain-specific CV problems [76,127]. Some landmark models include, but are not limited to, LeNet, AlexNet, VGG, Inception, ResNet, Xception, ResNeXt, and U-Net.

Transfer learning plays a crucial role in modern AI, where a model developed for a task is reused as the starting point for a model on a different but related task [128]; for instance, the ResNet model trained on ImageNet data with >14 million images in approximately 1000 categories (eg, tables and horses) has stored many useful visual patterns in its weights, which can help solve other CV tasks (eg, identifying fat tissues in MRI scans) [129]. Transfer learning can substantially reduce the number of images required to train a model for a particular task and boost model performance compared with models trained from scratch [130].

Maharana and Nsoesie [57] adopted the VGG model architecture to examine the relationship between obesity prevalence and the built environment measured by Google Maps images (eg, parks, highways, green streets, crosswalks, and diverse housing types). Similarly, Phan et al [42] used the VGG model to assess the link between the statewide prevalence of obesity, physical activity, and chronic disease mortality and the built environment using images from Google Street View. Bhanu et al [38] applied the U-Net model to identify adipose tissues from MRI data. Snekhalatha and Sangamithirai [30] applied transfer learning on a pretrained CNN model to detect obesity based on thermal imaging data.

DL on Text

Besides CV, NLP is another field where DL dominates [131]. Early NLP models primarily adopted recurrent neural network (RNN) architecture, demonstrating broad applicability to various NLP tasks such as sentiment analysis, text summarization, language translation, and speech recognition [74,132]. RNN differs from feed-forward MLP in that it takes information from prior inputs (stored as *memories*) to influence the current input and output, which capitalizes on the structure of sequential data where order matters (eg, time series or natural languages) [133]. Some popular RNN models used in NLP tasks include gated recurrent unit and long short-term memory unit [74]. However, in today's NLP landscape, transformers, invented by a team at Google in 2017, have surpassed RNN models such as gated recurrent unit and long short-term memory unit [134-136]. Transformers are encoder-decoder models that use self-attention to process language sequences [137]. An encoder maps an input sequence into state representation vectors. A decoder decodes the state representation vector to generate the target output sequence. The self-attention mechanism is used repeatedly within the encoder and the decoder to help them contextualize the input data. Specifically, the mechanism compares every word in the sentence to every other word, including itself, and reweighs each word's embeddings to incorporate contextual relevance. Popular transformer models such as GPT-3, BERT, XLNet, RoBERTa, and T5 have been widely applied to various NLP tasks and achieved state-of-the-art results [137]. Stephens et al [48] tested the efficacy of pediatric obesity treatment support through Tess, a behavioral coaching chatbot built on NLP models. The study concluded that Tess demonstrated therapeutic values to pediatric patients with obesity and prediabetes, especially outside of office hours, and could be scaled up to serve a larger patient population.

Discussion

Overview

This study conducted a scoping review of the applications of AI to obesity research. A keyword search in digital bibliographic databases identified 46 studies that used diverse ML and DL models to study obesity-related outcomes. In general, the studies found AI models helpful in detecting clinically meaningful patterns of obesity or relationships between specific covariates and weight outcomes. The majority (18/22, 82%) of the studies comparing AI models with conventional statistical approaches found that the AI models achieved higher prediction accuracy

on test data. Some (5/46, 11%) of the studies comparing the performances of different AI models revealed mixed results, likely indicating the high contingency of model performance on the data set and task it was applied to. An accelerating trend of adopting state-of-the-art DL models over standard ML models was observed to address challenging CV and NLP tasks. We concisely introduced the popular ML and DL models and summarized their specific applications in the studies included in the review.

Despite the variety of ML and DL models used in obesity research, it could well be the beginning of the trend for using AI applications in the big data era. Future adoptions of AI in obesity research could be influenced by a broad spectrum of factors, with 3 prominent ones discussed in the following sections.

Artificial General Intelligence

The ML and DL models reviewed in this study were primarily unimodal and task specific: they were built on a single data type (eg, tabular, text, or image) to solve a specific problem such as obesity classification or BMI prediction. Recent advances in AI showcase the feasibility and possibly superior performance of multimodal, multitask ML and DL models that are trained on diverse data types (eg, tabular plus text, image, video, or audio) and can handle many domains of downstream tasks (eg, text generation, object detection, time series prediction, and speech recognition) simultaneously [138-140]. However, it should be noted that the predictive accuracy of AI models may vary across gender and age groups [27] and sex and age groups [59]. Different from BMI, BMI *z* scores adjust for sex and age differences [141]. Future research may evaluate the potential disparities in AI model performances in their applications to BMI versus BMI *z* scores as outcome measures. Artificial general intelligence (AGI) refers to the ability of an intelligent agent to understand or learn any intellectual task performed by a human being [142,143]. It is too early to tell whether these multimodal, multitask ML and DL models may lead to AGI (or whether we could ever achieve AGI through technological innovations) [144]. Nevertheless, we may soon witness increasing applications of these models in obesity-related research.

Synthetic Data Generation

Data access is fundamental to any AI model training. Two primary barriers with regard to data are limited sample size and confidentiality concerns [145-148]. ML and DL models are increasingly used to generate synthetic data as an alternative to data collected from the real world [149,150]. Synthetic data do not contain private information requiring human subject review and, therefore, can be shared with other parties or the public without confidentiality concerns [151]. By contrast, synthetic data preserve the original data's mathematical and statistical properties, ensuring that the AI model trained on them can be generalized to real-world data [152]. In addition, given the unrestrained availability of synthetic data (only limited by the computational power of data generation), AI models trained on synthetic data can be robust with regard to data variations [153]. Synthetic data of various types, such as tabular, text, and image, have been generated in massive quantities to train ML and DL models cost-effectively. Obesity-related data or, more generally, health-related data can be expensive to collect (eg, MRI scans) and contain confidential information (eg, patients' names or residential addresses), which could be addressed by synthetic data generation [154].

Human-in-the-Loop

There have been increasing concerns over AI-related data bias and ethical issues [155,156]. Fundamentally, AI models should facilitate but not replace human judgment and decision-making [157,158]. Human-in-the-loop (HITL) is an AI model that requires human interaction [159,160]. HITL ensures that algorithm biases and potentially destructive model outputs can be identified in a timely manner and corrected to prevent adverse consequences. However, such interactions between humans and machines require thoughtful designs in the data-processing pipeline, model architecture, and personnel management [159]. Data- and model-driven decision-making related to obesity, such as behavioral modifications (eg, diet or physical activity interventions) or medical treatment, can be complex [161]. AI-powered wearables and other digital health platforms can detect change in an individual's physical activity and provide actionable information to improve health outcomes [162-164]. Mobile chemical sensors could offer timely dietary information by monitoring real-time chemical variations upon food consumption, collecting dynamic data based on an individual's metabolic profile and environmental exposure, thus supporting dietary behavior decision-making to improve precise nutrition [165]. HITL may integrate AI model outputs with expert inputs to make informed decisions that capitalize on the strengths of both and maximize patients' chances of health restoration and improvement [166].

Limitations of the Scoping Review and Included Studies

To our knowledge, this study is the first to systematically review AI-related methodologies adopted in the obesity literature and project trends for future technological development and applications. However, several limitations should be noted concerning this review and the included studies. As our review focused on ML and DL methods, study-specific findings (eg, the effectiveness of an intervention and estimated associations between covariates and an outcome) were not synthesized in detail. The included studies were heterogeneous in terms of hypothesis and research question, study design, population sampled, data collection method, sample size, and data quality. The analytic approach chosen was endogenous to these study-specific parameters; therefore, across-study comparisons of model performances may not be reliable. Even within the same study, conclusions about relative model performances (eg, the prediction accuracy of logistic regression vs SVM) may lack generalizability because of the interdependency between data and ML and DL algorithms. AI technologies are rapidly advancing, with innovations and breakthroughs almost daily. A review such as this one will have a short shelf life and warrant periodic updates.

Conclusions

This study reviewed the AI-related methodologies adopted in the obesity literature, particularly ML and DL models applied to tabular, image, and text data for obesity measurement, prediction, and treatment. It aimed to provide researchers and practitioners with an overview of the AI applications to obesity research, familiarize them with popular ML and DL models, and facilitate their adoption of AI applications. The review also discussed emerging trends such as multimodal and multitask AI models, synthetic data generation, and HITL, which may witness increasing applications in obesity research.

Acknowledgments

This research was partially funded by the Fundamental Research Funds for the Central Universities, China University of Geosciences, Beijing (grant 2-9-2020-036).

Authors' Contributions

RA designed the study and wrote the manuscript. RA and JS jointly designed the search algorithm and screened articles. JS performed data extraction and constructed the summary tables. YX drafted part of the Discussion section. JS and YX revised the manuscript. The co–first authors RA and JS contributed equally.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search algorithm used in PubMed.
[DOC File , 12 KB - jmir_v24i12e40589_app1.doc]

References

1. The double burden of malnutrition. The Lancet. 2019 Dec 16. URL: https://www.thelancet.com/series/double-burden-malnutrition [accessed 2022-06-18]
2. An R, Ji M, Zhang S. Global warming and obesity: a systematic review. Obes Rev 2018 Mar;19(2):150-163. [doi: 10.1111/obr.12624] [Medline: 28977817]
3. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781 [FREE Full text] [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
4. Obesity and overweight. World Health Organization. URL: https://www.who.int/news- room /fact-sheets/detail/obesity-and-overweight [accessed 2022-06-18]
5. NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. Lancet 2016 Apr 02;387(10026):1377-1396 [FREE Full text] [doi: 10.1016/S0140-6736(16)30054-X] [Medline: 27115820]
6. Sivarajah U, Kamal M, Irani Z, Weerakkody V. Critical analysis of Big Data challenges and analytical methods. J Business Res 2017 Jan;70:263-286 [FREE Full text] [doi: 10.1016/j.jbusres.2016.08.001]
7. Dash S, Shakyawar S, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019 Jun 19;6(1):54 [FREE Full text] [doi: 10.1186/s40537-019-0217-0]
8. Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. Heredity (Edinb) 2020 Apr;124(4):525-534 [FREE Full text] [doi: 10.1038/s41437-020-0303-2] [Medline: 32139886]
9. Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, et al. Artificial intelligence: a powerful paradigm for scientific research. Innovation (Camb) 2021 Nov 28;2(4):100179 [FREE Full text] [doi: 10.1016/j.xinn.2021.100179] [Medline: 34877560]
10. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Humaniz Comput 2022 Jan 13:1-28 [FREE Full text] [doi: 10.1007/s12652-021-03612-z] [Medline: 35039756]
11. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
12. Syrowatka A, Kuznetsova M, Alsubai A, Beckman AL, Bain PA, Craig KJ, et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. NPJ Digit Med 2021 Jun 10;4(1):96 [FREE Full text] [doi: 10.1038/s41746-021-00459-8] [Medline: 34112939]
13. Bin Sawad A, Narayan B, Alnefaie A, Maqbool A, Mckie I, Smith J, et al. A systematic review on healthcare artificial intelligent conversational agents for chronic conditions. Sensors (Basel) 2022 Mar 29;22(7):2625 [FREE Full text] [doi: 10.3390/s22072625] [Medline: 35408238]

14. Goh YS, Ow Yong JQ, Chee BQ, Kuek JH, Ho CS. Machine learning in health promotion and behavioral change: scoping review. J Med Internet Res 2022 Jun 02;24(6):e35831 [FREE Full text] [doi: 10.2196/35831] [Medline: 35653177]
15. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare. Cambridge, Massachusetts, United States: Academic Press; 2020.
16. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022 Jan;28(1):31-38. [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
17. 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]
18. 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]
19. Aung YY, Wong DC, Ting DS. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. Br Med Bull 2021 Sep 10;139(1):4-15. [doi: 10.1093/bmb/ldab016] [Medline: 34405854]
20. Quinn TP, Senadeera M, Jacobs S, Coghlan S, Le V. Trust and medical AI: the challenges we face and the expertise needed to overcome them. J Am Med Inform Assoc 2021 Mar 18;28(4):890-894 [FREE Full text] [doi: 10.1093/jamia/ocaa268] [Medline: 33340404]
21. 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]
22. Chew HS, Achananuparp P. Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. J Med Internet Res 2022 Jan 14;24(1):e32939 [FREE Full text] [doi: 10.2196/32939] [Medline: 35029538]
23. Marmett B, Carvalho RB, Fortes MS, Cazella SC. Artificial Intelligence technologies to manage obesity. Vittalle J Health Sci 2018 Sep 27;30(2):73-79. [doi: 10.14295/vittalle.v30i2.7654]
24. Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: review of the literature. J Med Internet Res 2019 Apr 05;21(4):e12286 [FREE Full text] [doi: 10.2196/12286] [Medline: 30950797]
25. 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]
26. Abdel-Aal RE, Mangoud AM. Modeling obesity using abductive networks. Comput Biomed Res 1997 Dec;30(6):451-471. [doi: 10.1006/cbmr.1997.1460] [Medline: 9466835]
27. Alkutbe RB, Alruban A, Alturki H, Sattar A, Al-Hazzaa H, Rees G. Fat mass prediction equations and reference ranges for Saudi Arabian Children aged 8-12 years using machine technique method. PeerJ 2021;9:e10734 [FREE Full text] [doi: 10.7717/peerj.10734] [Medline: 33665006]
28. Zare S, Thomsen MR, Nayga RM, Goudie A. Use of machine learning to determine the information value of a BMI screening program. Am J Prev Med 2021 Mar;60(3):425-433 [FREE Full text] [doi: 10.1016/j.amepre.2020.10.016] [Medline: 33483154]
29. Thamrin SA, Arsyad DS, Kuswanto H, Lawi A, Nasir S. Predicting obesity in adults using machine learning techniques: an analysis of Indonesian basic health research 2018. Front Nutr 2021;8:669155 [FREE Full text] [doi: 10.3389/fnut.2021.669155] [Medline: 34235168]
30. U S, K. PT, K S. Computer aided diagnosis of obesity based on thermal imaging using various convolutional neural networks. Biomed Signal Process Control 2021 Jan;63:102233 [FREE Full text] [doi: 10.1016/j.bspc.2020.102233]
31. Rashmi R, Umapathy S, Krishnan P. Thermal imaging method to evaluate childhood obesity based on machine learning techniques. Int J Imaging Syst Technol 2021 Mar 20;31(3):1752-1768 [FREE Full text] [doi: 10.1002/ima.22572]
32. Park HJ, Francisco SC, Pang MR, Peng L, Chi G. Exposure to anti-black lives matter movement and obesity of the black population. Soc Sci Med 2021 Jul 28:114265. [doi: 10.1016/j.socscimed.2021.114265] [Medline: 34366168]
33. Pang X, Forrest CB, Lê-Scherban F, Masino AJ. Prediction of early childhood obesity with machine learning and electronic health record data. Int J Med Inform 2021 Jun;150:104454 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104454] [Medline: 33866231]
34. Lin Z, Feng W, Liu Y, Ma C, Arefan D, Zhou D, et al. Machine learning to identify metabolic subtypes of obesity: a multi-center study. Front Endocrinol (Lausanne) 2021;12:713592 [FREE Full text] [doi: 10.3389/fendo.2021.713592] [Medline: 34335479]
35. Lee K, Kim HY, Lee SJ, Kwon SO, Na S, Hwang HS, Korean Society of Ultrasound in ObstetricsGynecology Research Group. Prediction of newborn's body mass index using nationwide multicenter ultrasound data: a machine-learning study. BMC Pregnancy Childbirth 2021 Mar 02;21(1):172 [FREE Full text] [doi: 10.1186/s12884-021-03660-5] [Medline: 33653299]
36. Delnevo G, Mancini G, Roccetti M, Salomoni P, Trombini E, Andrei F. The prediction of body mass index from negative affectivity through machine learning: a confirmatory study. Sensors (Basel) 2021 Mar 29;21(7):2361 [FREE Full text] [doi: 10.3390/s21072361] [Medline: 33805257]
37. Cheng X, Lin S, Liu J, Liu S, Zhang J, Nie P, et al. Does physical activity predict obesity-a machine learning and statistical method-based analysis. Int J Environ Res Public Health 2021 Apr 09;18(8):3966 [FREE Full text] [doi: 10.3390/ijerph18083966] [Medline: 33918760]

38. Bhanu PK, Arvind CS, Yeow LY, Chen WX, Lim WS, Tan CH. CAFT: a deep learning-based comprehensive abdominal fat analysis tool for large cohort studies. MAGMA 2022 Apr;35(2):205-220. [doi: 10.1007/s10334-021-00946-9] [Medline: 34338926]
39. Yao Y, Song L, Ye J. Motion-to-BMI: using motion sensors to predict the body mass index of smartphone users. Sensors (Basel) 2020 Feb 19;20(4):1134 [FREE Full text] [doi: 10.3390/s20041134] [Medline: 32093013]
40. Xiao Y, Zhang Y, Sun Y, Tao P, Kuang X. Does green space really matter for residents' obesity? A new perspective from Baidu street view. Front Public Health 2020;8:332 [FREE Full text] [doi: 10.3389/fpubh.2020.00332] [Medline: 32850579]
41. Taghiyev A, Altun A, Caglar S. A hybrid approach based on machine learning to identify the causes of obesity. J Control Eng Applied Informatic 2020;22(2):56-66.
42. Phan L, Yu W, Keralis JM, Mukhija K, Dwivedi P, Brunisholz KD, et al. Google street view derived built environment indicators and associations with state-level obesity, physical activity, and chronic disease mortality in the united states. Int J Environ Res Public Health 2020 May 22;17(10):3659 [FREE Full text] [doi: 10.3390/ijerph17103659] [Medline: 32456114]
43. Park B, Chung C, Lee MJ, Park H. Accurate neuroimaging biomarkers to predict body mass index in adolescents: a longitudinal study. Brain Imaging Behav 2020 Oct;14(5):1682-1695. [doi: 10.1007/s11682-019-00101-y] [Medline: 31065926]
44. Kibble M, Khan SA, Ammad-Ud-Din M, Bollepalli S, Palviainen T, Kaprio J, et al. An integrative machine learning approach to discovering multi-level molecular mechanisms of obesity using data from monozygotic twin pairs. R Soc Open Sci 2020 Oct;7(10):200872 [FREE Full text] [doi: 10.1098/rsos.200872] [Medline: 33204460]
45. Fu Y, Gou W, Hu W, Mao Y, Tian Y, Liang X, et al. Integration of an interpretable machine learning algorithm to identify early life risk factors of childhood obesity among preterm infants: a prospective birth cohort. BMC Med 2020 Jul 10;18(1):184 [FREE Full text] [doi: 10.1186/s12916-020-01642-6] [Medline: 32646442]
46. Dunstan J, Aguirre M, Bastías M, Nau C, Glass TA, Tobar F. Predicting nationwide obesity from food sales using machine learning. Health Informatics J 2020 Mar;26(1):652-663 [FREE Full text] [doi: 10.1177/1460458219845959] [Medline: 31106648]
47. Blanes-Selva V, Tortajada S, Vilar R, Valdivieso B, García-Gómez JM. Machine learning-based identification of obesity from positive and unlabelled electronic health records. Stud Health Technol Inform 2020 Jun 16;270:864-868. [doi: 10.3233/SHTI200284] [Medline: 32570505]
48. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
49. Shin S, Lee J, Choe S, Yang HI, Min J, Ahn K, et al. Dry electrode-based body fat estimation system with anthropometric data for use in a wearable device. Sensors (Basel) 2019 May 10;19(9):2177 [FREE Full text] [doi: 10.3390/s19092177] [Medline: 31083445]
50. Scheinker D, Valencia A, Rodriguez F. Identification of factors associated with variation in US county-level obesity prevalence rates using epidemiologic vs machine learning models. JAMA Netw Open 2019 Apr 05;2(4):e192884 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2884] [Medline: 31026030]
51. Ramyaa R, Hosseini O, Krishnan GP, Krishnan S. Phenotyping women based on dietary macronutrients, physical activity, and body weight using machine learning tools. Nutrients 2019 Jul 22;11(7):1681 [FREE Full text] [doi: 10.3390/nu11071681] [Medline: 31336626]
52. Hong N, Wen A, Stone DJ, Tsuji S, Kingsbury PR, Rasmussen LV, et al. Developing a FHIR-based EHR phenotyping framework: a case study for identification of patients with obesity and multiple comorbidities from discharge summaries. J Biomed Inform 2019 Nov;99:103310 [FREE Full text] [doi: 10.1016/j.jbi.2019.103310] [Medline: 31622801]
53. Hammond R, Athanasiadou R, Curado S, Aphinyanaphongs Y, Abrams C, Messito MJ, et al. Predicting childhood obesity using electronic health records and publicly available data. PLoS ONE 2019 Apr 22;14(4):e0215571 [FREE Full text] [doi: 10.1371/journal.pone.0215571] [Medline: 31009509]
54. Gerl MJ, Klose C, Surma MA, Fernandez C, Melander O, Männistö S, et al. Machine learning of human plasma lipidomes for obesity estimation in a large population cohort. PLoS Biol 2019 Oct;17(10):e3000443 [FREE Full text] [doi: 10.1371/journal.pbio.3000443] [Medline: 31626640]
55. Duran I, Martakis K, Rehberg M, Semler O, Schoenau E. Diagnostic performance of an artificial neural network to predict excess body fat in children. Pediatr Obes 2019 Feb;14(2):e12494. [doi: 10.1111/ijpo.12494] [Medline: 30590878]
56. Wang H, Chang S, Lin W, Chen C, Chiang S, Huang K, et al. Machine learning-based method for obesity risk evaluation using single-nucleotide polymorphisms derived from next-generation sequencing. J Comput Biol 2018 Dec;25(12):1347-1360. [doi: 10.1089/cmb.2018.0002] [Medline: 30204480]
57. Maharana A, Nsoesie EO. Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity. JAMA Netw Open 2018 Aug 03;1(4):e181535 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1535] [Medline: 30646134]
58. Ortega Hinojosa AM, MacLeod KE, Balmes J, Jerrett M. Influence of school environments on childhood obesity in California. Environ Res 2018 Oct;166:100-107. [doi: 10.1016/j.envres.2018.04.022] [Medline: 29883903]

59. Seyednasrollah F, Mäkelä J, Pitkänen N, Juonala M, Hutri-Kähönen N, Lehtimäki T, et al. Prediction of adulthood obesity using genetic and childhood clinical risk factors in the cardiovascular risk in young finns study. Circ Cardiovasc Genet 2017 Jun;10(3):e001554 [FREE Full text] [doi: 10.1161/CIRCGENETICS.116.001554] [Medline: 28620069]
60. Lingren T, Thaker V, Brady C, Namjou B, Kennebeck S, Bickel J, et al. Developing an algorithm to detect early childhood obesity in two tertiary pediatric medical centers. Appl Clin Inform 2016 Jul 20;7(3):693-706 [FREE Full text] [doi: 10.4338/ACI-2016-01-RA-0015] [Medline: 27452794]
61. Almeida SM, Furtado JM, Mascarenhas P, Ferraz ME, Silva LR, Ferreira JC, et al. Anthropometric predictors of body fat in a large population of 9-year-old school-aged children. Obes Sci Pract 2016 Sep;2(3):272-281 [FREE Full text] [doi: 10.1002/osp4.51] [Medline: 27708844]
62. Nau C, Ellis H, Huang H, Schwartz BS, Hirsch A, Bailey-Davis L, et al. Exploring the forest instead of the trees: an innovative method for defining obesogenic and obesoprotective environments. Health Place 2015 Sep;35:136-146 [FREE Full text] [doi: 10.1016/j.healthplace.2015.08.002] [Medline: 26398219]
63. Dugan T, Mukhopadhyay S, Carroll A, Downs S. Machine learning techniques for prediction of early childhood obesity. Appl Clin Inform 2017 Dec 19;06(03):506-520 [FREE Full text] [doi: 10.4338/aci-2015-03-ra-0036]
64. Chen H, Yang B, Liu D, Liu W, Liu Y, Zhang X, et al. Using blood indexes to predict overweight statuses: an extreme learning machine-based approach. PLoS One 2015;10(11):e0143003 [FREE Full text] [doi: 10.1371/journal.pone.0143003] [Medline: 26600199]
65. Shao YE. Body fat percentage prediction using intelligent hybrid approaches. ScientificWorldJournal 2014;2014:383910 [FREE Full text] [doi: 10.1155/2014/383910] [Medline: 24723804]
66. Kupusinac A, Stokić E, Doroslovački R. Predicting body fat percentage based on gender, age and BMI by using artificial neural networks. Comput Methods Programs Biomed 2014 Feb;113(2):610-619. [doi: 10.1016/j.cmpb.2013.10.013] [Medline: 24275480]
67. Heydari ST, Ayatollahi SM, Zare N. Comparison of artificial neural networks with logistic regression for detection of obesity. J Med Syst 2012 Aug;36(4):2449-2454. [doi: 10.1007/s10916-011-9711-4] [Medline: 21556898]
68. Zhang S, Tjortjis C, Zeng X, Qiao H, Buchan I, Keane J. Comparing data mining methods with logistic regression in childhood obesity prediction. Inf Syst Front 2009 Feb 24;11(4):449-460 [FREE Full text] [doi: 10.1007/s10796-009-9157-0]
69. Yang H, Spasic I, Keane JA, Nenadic G. A text mining approach to the prediction of disease status from clinical discharge summaries. J Am Med Inform Assoc 2009;16(4):596-600 [FREE Full text] [doi: 10.1197/jamia.M3096] [Medline: 19390098]
70. Ergün U. The classification of obesity disease in logistic regression and neural network methods. J Med Syst 2009 Feb;33(1):67-72. [doi: 10.1007/s10916-008-9165-5] [Medline: 19238898]
71. Positano V, Cusi K, Santarelli MF, Sironi A, Petz R, Defronzo R, et al. Automatic correction of intensity inhomogeneities improves unsupervised assessment of abdominal fat by MRI. J Magn Reson Imaging 2008 Aug;28(2):403-410. [doi: 10.1002/jmri.21448] [Medline: 18666138]
72. Haenlein M, Kaplan A. A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. California Manag Rev 2019 Jul 17;61(4):5-14. [doi: 10.1177/0008125619864925]
73. Artificial Intelligence. Stanford Encyclopedia of Philosophy. URL: https://plato.stanford.edu/ [accessed 2022-06-18]
74. Chollet F. Deep Learning with Python, Second Edition. Shelter Island, New York, United States: Manning; 2021.
75. Samuel A. Some studies in machine learning using the game of checkers. IBM J Res Dev 1959 Jul;3(3):210-229 [FREE Full text] [doi: 10.1147/rd.33.0210]
76. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 2021;8(1):53 [FREE Full text] [doi: 10.1186/s40537-021-00444-8] [Medline: 33816053]
77. Chauhan N, Singh K. A review on conventional machine learning vs deep learning. In: Proceedings of the International Conference on Computing, Power and Communication Technologies (GUCON). 2018 Presented at: International Conference on Computing, Power and Communication Technologies (GUCON); Sep 28-29, 2018; Greater Noida, India. [doi: 10.1109/gucon.2018.8675097]
78. Rajula HS, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina (Kaunas) 2020 Sep 08;56(9):455 [FREE Full text] [doi: 10.3390/medicina56090455] [Medline: 32911665]
79. Bennett M, Hayes K, Kleczyk E, Mehta R. Similarities and differences between machine learning and traditional advanced statistical modeling in healthcare analytics. arXiv 2022 [FREE Full text] [doi: 10.5772/intechopen.105116]
80. Ley C, Martin RK, Pareek A, Groll A, Seil R, Tischer T. Machine learning and conventional statistics: making sense of the differences. Knee Surg Sports Traumatol Arthrosc 2022 Mar;30(3):753-757. [doi: 10.1007/s00167-022-06896-6] [Medline: 35106604]
81. KhosrowHassibi. Machine learning vs. traditional statistics: different philosophies, different approaches. Data Science Central. 2016 Oct 28. URL: https://www.datasciencecentral.com/machine-learning-vs-traditional-statistics-different- [accessed 2022-06-18]
82. Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. arXiv. 2018 Nov. URL: https://arxiv.org/pdf/1811.12808.pdf [accessed 2022-06-18]

83. Ding J, Tarokh V, Yang Y. Model selection techniques: an overview. arXiv. URL: https://arxiv.org/pdf/ [accessed 2022-06-18]
84. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf A. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. Cham: Springer; 2019.
85. Wittek P. Quantum Machine Learning What Quantum Computing Means to Data Mining. Boston: Academic Press; 2014.
86. Géron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems. Sebastopol, California, United States: O'Reilly Media; 2017.
87. Addi A, Tarik A, Fatima G. Comparative survey of association rule mining algorithms based on multiple-criteria decision analysis approach. In: Proceedings of the 3rd International Conference on Control, Engineering & Information Technology (CEIT). 2015 Presented at: 3rd International Conference on Control, Engineering & Information Technology (CEIT); May 25-27, 2015; Tlemcen, Algeria. [doi: 10.1109/ceit.2015.7233078]
88. Velliangiri S, Alagumuthukrishnan S, Thankumar joseph SI. A review of dimensionality reduction techniques for efficient computation. Procedia Comput Sci 2019;165:104-111. [doi: 10.1016/j.procs.2020.01.079]
89. Singh A, Thakur N, Sharma A. A review of supervised machine learning algorithms. In: Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). 2016 Presented at: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom); Mar 16-18, 2016; New Delhi, India.
90. Ashabi A, Sahibuddin S, Haghighi MS. The systematic review of K-means clustering algorithm. In: Proceedings of the 2020 The 9th International Conference on Networks, Communication and Computing. 2020 Presented at: ICNCC 2020: 2020 The 9th International Conference on Networks, Communication and Computing; Dec 18 - 20, 2020; Tokyo Japan. [doi: 10.1145/3447654.3447657]
91. Gosain A, Dahiya S. Performance analysis of various fuzzy clustering algorithms: a review. Procedia Comput Sci 2016;79:100-111. [doi: 10.1016/j.procs.2016.03.014]
92. Arora J, Khatter K, Tushir M. Fuzzy c-Means Clustering Strategies: A Review of Distance Measures. Singapore: Springer; 2018.
93. Zhang M, Zhang W, Sicotte H, Yang P. A new validity measure for a correlation-based fuzzy c-means clustering algorithm. In: Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2009 Presented at: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Sep 03-06, 2009; Minneapolis, MN, USA. [doi: 10.1109/iembs.2009.5332582]
94. Tavakol M, Wetzel A. Factor analysis: a means for theory and instrument development in support of construct validity. Int J Med Educ 2020 Nov 06;11:245-247 [FREE Full text] [doi: 10.5116/ijme.5f96.0f4a] [Medline: 33170146]
95. Klami A, Virtanen S, Leppäaho E, Kaski S. Group factor analysis. IEEE Trans Neural Netw Learn Syst 2015 Sep;26(9):2136-2147. [doi: 10.1109/TNNLS.2014.2376974] [Medline: 25532193]
96. Steyerberg E. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.
97. Dasgupta A, Sun YV, König IR, Bailey-Wilson JE, Malley JD. Brief review of regression-based and machine learning methods in genetic epidemiology: the Genetic Analysis Workshop 17 experience. Genet Epidemiol 2011;35 Suppl 1:S5-11 [FREE Full text] [doi: 10.1002/gepi.20642] [Medline: 22128059]
98. Gosiewska A, Kozak A, Biecek P. Simpler is better: lifting interpretability-performance trade-off via automated feature engineering. Decision Support Syst 2021 Nov;150:113556. [doi: 10.1016/j.dss.2021.113556]
99. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. Proc Natl Acad Sci U S A 2019 Aug 06;116(32):15849-15854 [FREE Full text] [doi: 10.1073/pnas.1903070116] [Medline: 31341078]
100. Bewick V, Cheek L, Ball J. Statistics review 14: logistic regression. Crit Care 2005 Feb;9(1):112-118 [FREE Full text] [doi: 10.1186/cc3045] [Medline: 15693993]
101. Rifkin R, Klautau A. In defense of one-vs-all classification. J Mach Learn Res 2004;5:101-141.
102. Wickramasinghe I, Kalutarage H. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Comput 2020 Sep 09;25(3):2277-2293. [doi: 10.1007/s00500-020-05297-6]
103. Taunk K, De S, Verma S, Swetapadma A. A brief review of nearest neighbor algorithm for learning and classification. In: Proceedings of the 2019 International Conference on Intelligent Computing and Control Systems (ICCS). 2019 Presented at: 2019 International Conference on Intelligent Computing and Control Systems (ICCS); May 15-17, 2019; Madurai, India URL: https://doi.org/10.1109/ICCS45141.2019.9065747 [doi: 10.1109/iccs45141.2019.9065747]
104. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 2020 Sep;408:189-215. [doi: 10.1016/j.neucom.2019.10.118]
105. Ponte P, Melko RG. Kernel methods for interpretable machine learning of order parameters. Phys Rev B 2017 Nov 27;96(20). [doi: 10.1103/physrevb.96.205146]
106. Kotsiantis SB. Decision trees: a recent overview. Artif Intell Rev 2011 Jun 29;39(4):261-283. [doi: 10.1007/s10462-011-9272-4]

107. Podgorelec V, Zorman M. Decision tree learning. In: Encyclopedia of Complexity and Systems Science. Berlin, Heidelberg: Springer; 2015.
108. Somvanshi M, Chavan P, Tambade S, Shinde S. A review of machine learning techniques using decision tree and support vector machine. In: Proceedings of the 2016 International Conference on Computing Communication Control and automation (ICCUBEA). 2016 Presented at: 2016 International Conference on Computing Communication Control and automation (ICCUBEA); Aug 12-13, 2016; Pune, India. [doi: 10.1109/iccubea.2016.7860040]
109. Re M, Valentini G. Ensemble methods: a review. In: Advances in Machine Learning and Data Mining for Astronomy. London, United Kingdom: Chapman & Hall; 2012.
110. Parmar A, Katariya R, Patel V. A review on random forest: an ensemble classifier. In: International Conference on Intelligent Data Communication Technologies and Internet of Things. Cham: Springer; 2018.
111. Talekar B. A detailed review on decision tree and random forest. Biosci Biotech Res Comm 2020 Dec 28;13(14):245-248. [doi: 10.21786/bbrc/13.14/57]
112. Ferreira A, Figueiredo M. Boosting algorithms: a review of methods, theory, and applications. In: Ensemble Machine Learning. Boston, MA: Springer; 2012.
113. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev 2020 Aug 24;54(3):1937-1967. [doi: 10.1007/s10462-020-09896-5]
114. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13 - 17, 2016; San Francisco California USA. [doi: 10.1145/2939672.2939785]
115. Prihastuti Yasmirullah SD, Otok BW, Trijoyo Purnomo JD, Prastyo DD. Modification of Multivariate Adaptive Regression Spline (MARS). J Phys Conf Ser 2021 Mar 01;1863(1):012078. [doi: 10.1088/1742-6596/1863/1/012078]
116. Friedman J. Multivariate adaptive regression splines. Ann Statist 1991 Mar 1;19(1):1-67 [FREE Full text] [doi: 10.1214/aos/1176347963]
117. Zhang W, Goh A. Multivariate adaptive regression splines for analysis of geotechnical engineering systems. Comput Geotechnics 2013 Mar;48:82-95. [doi: 10.1016/j.compgeo.2012.09.016]
118. Zhong G, Ling X, Wang L. From shallow feature learning to deep learning: benefits from the width and depth of deep architectures. WIREs Data Min Knowl 2018 Mar 28;9(1):e1255 [FREE Full text] [doi: 10.1002/widm.1255]
119. Gardner M, Dorling S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric Environ 1998 Aug;32(14-15):2627-2636. [doi: 10.1016/s1352-2310(97)00447-0]
120. Rynkiewicz J. On overfitting of multilayer perceptrons for classification. In: ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2019 Presented at: ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learnin; Apr 24-26, 2019; Bruges, Belgium URL: https://www.esann.org/sites/default/files/proceedings/legacy/es2019-80.pdf [doi: 10.1007/978-1-4615-3210-1_4]
121. Park J, Edington DW. Application of a prediction model for identification of individuals at diabetic risk. Methods Inf Med 2004;43(3):273-281. [Medline: 15227557]
122. Voulodimos A, Doulamis N, Bebis G, Stathaki T. Recent developments in deep learning for engineering applications. Comput Intell Neurosci 2018;2018:8141259-8142018 [FREE Full text] [doi: 10.1155/2018/8141259] [Medline: 29861713]
123. Chai J, Zeng H, Li A, Ngai EW. Deep learning in computer vision: a critical review of emerging techniques and application scenarios. Mach Learn Application 2021 Dec;6:100134. [doi: 10.1016/j.mlwa.2021.100134]
124. Dhillon A, Verma GK. Convolutional neural network: a review of models, methodologies and applications to object detection. Prog Artif Intell 2019 Dec 20;9(2):85-112. [doi: 10.1007/s13748-019-00203-0]
125. Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 2020 Apr 21;53(8):5455-5516. [doi: 10.1007/s10462-020-09825-6]
126. Stevens E, Antiga L, Viehmann T. Deep Learning with PyTorch Build, Train, and Tune Neural Networks Using Python Tools. Shelter Island, New York, United States: Manning; 2020.
127. Aloysius N, Geetha M. A review on deep convolutional neural networks. In: Proceedings of the 2017 International Conference on Communication and Signal Processing (ICCSP). 2017 Presented at: 2017 International Conference on Communication and Signal Processing (ICCSP); Apr 06-08, 2017; Chennai, India. [doi: 10.1109/iccsp.2017.8286426]
128. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proc IEEE 2021 Jan;109(1):43-76. [doi: 10.1109/jproc.2020.3004555]
129. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Jun 27-30, 2016; Las Vegas, NV, USA. [doi: 10.1109/cvpr.2016.90]
130. Weiss K, Khoshgoftaar T, Wang D. A survey of transfer learning. J Big Data 2016 May 28;3(1):1-40 [FREE Full text] [doi: 10.1186/s40537-016-0043-6]
131. Li H. Deep learning for natural language processing: advantages and challenges. National Sci Rev 2017;5(1):24-26 [FREE Full text]

132. Le Glaz A, Haralambous Y, Kim-Dufor D, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021 May 04;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
133. Lipton ZC, Berkowitz J, Elkan C. A critical review of recurrent neural networks for sequence learning. arXiv 2015 Jun 5 [FREE Full text]
134. Chernyavskiy A, Ilvovsky D, Nakov P. Transformers: "the end of history" for natural language processing? In: Machine Learning and Knowledge Discovery in Databases. Cham: Springer; 2021.
135. Lin T, Wang Y, Liu X, Qiu X. A survey of transformers. AI Open 2022;3:111-132 [FREE Full text] [doi: 10.1016/j.aiopen.2022.10.001]
136. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN. Attention is all you need. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017). 2017 Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4 - 9, 2017; Long Beach, CA, USA URL: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
137. Tunstall L, von WL, Wolf T. Natural Language Processing with Transformers, Revised Edition. Sebastopol, California, United States: O'Reilly Media; 2022.
138. Summaira J, Li X, Shoib AM, Bourahla O, Songyuan L, Abdul J. Recent advances and trends in multimodal deep learning: a review. arXiv 2021 May [FREE Full text]
139. Bayoudh K, Knani R, Hamdaoui F, Mtibaa A. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. Vis Comput 2022;38(8):2939-2970 [FREE Full text] [doi: 10.1007/s00371-021-02166-7] [Medline: 34131356]
140. Zhang Y, Yang Q. An overview of multi-task learning. National Sci Rev 2018;5(1):30-43. [doi: 10.1093/nsr/nwx105]
141. Fagerberg P, Charmandari E, Diou C, Heimeier R, Karavidopoulou Y, Kassari P, et al. Fast eating is associated with increased BMI among high-school students. Nutrients 2021 Mar 09;13(3):880 [FREE Full text] [doi: 10.3390/nu13030880] [Medline: 33803093]
142. Nyalapelli V, Gandhi M, Bhargava S, Dhanare R, Bothe S. Review of progress in artificial general intelligence and human brain inspired cognitive architecture. In: Proceedings of the 2021 International Conference on Computer Communication and Informatics (ICCCI). 2021 Presented at: 2021 International Conference on Computer Communication and Informatics (ICCCI); Jan 27-29, 2021; Coimbatore, India. [doi: 10.1109/iccci50826.2021.9402446]
143. Long L, Cotner C. A review and proposed framework for artificial general intelligence. In: Proceedings of the 2019 IEEE Aerospace Conference. 2019 Presented at: 2019 IEEE Aerospace Conference; Mar 02-09, 2019; Big Sky, MT, USA. [doi: 10.1109/aero.2019.8742125]
144. Fjelland R. Why general artificial intelligence will not be realized. Humanit Soc Sci Commun 2020 Jun 17;7(1). [doi: 10.1057/s41599-020-0494-4]
145. Bae H, Jang J, Jung D, Jang H, Ha H, Lee H, et al. Security and privacy issues in deep learning. arXiv 2021 Mar [FREE Full text]
146. Ha T, Dang TK, Le H, Truong TA. Security and privacy issues in deep learning: a brief review. Sn Comput Sci 2020 Aug 06;1(5). [doi: 10.1007/s42979-020-00254-4]
147. Keshari R, Ghosh S, Chhabra S, Vatsa M, Singh R. Unravelling small sample size problems in the deep learning world. In: Proceedings of the 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM). 2020 Presented at: 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM); Sep 24-26, 2020; New Delhi, India. [doi: 10.1109/bigmm50055.2020.00028]
148. Liu B, Wei Y, Zhang Y, Yang Q. Deep neural networks for high dimension, low sample size data. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). 2017 Presented at: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17); Aug 19-25, 2017; Melbourne, Australia URL: https://www.ijcai.org/Proceedings/2017/0318.pdf [doi: 10.24963/ijcai.2017/318]
149. Goncalves A, Ray P, Soper B, Stevens J, Coyle L, Sales AP. Generation and evaluation of synthetic patient data. BMC Med Res Methodol 2020 May 07;20(1):108 [FREE Full text] [doi: 10.1186/s12874-020-00977-1] [Medline: 32381039]
150. Chen RJ, Lu MY, Chen TY, Williamson DF, Mahmood F. Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng 2021 Jun;5(6):493-497 [FREE Full text] [doi: 10.1038/s41551-021-00751-8] [Medline: 34131324]
151. Rankin D, Black M, Bond R, Wallace J, Mulvenna M, Epelde G. Reliability of supervised machine learning using synthetic data in health care: model to preserve privacy for data sharing. JMIR Med Inform 2020 Jul 20;8(7):e18910 [FREE Full text] [doi: 10.2196/18910] [Medline: 32501278]
152. El Emam K, Mosquera L, Hoptroff R. Practical Synthetic Data Generation Balancing Privacy and the Broad Availability of Data. Sebastopol, California, United States: O'Reilly Media; 2020.
153. Sergey I N. Synthetic Data for Deep Learning. Cham: Springer International Publishing; 2021.
154. Moya-Sáez E, Peña-Nogales Ó, Luis-García R, Alberola-López C. A deep learning approach for synthetic MRI based on two routine sequences and training with synthetic data. Comput Methods Programs Biomed 2021 Oct;210:106371 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106371] [Medline: 34525411]

155. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput Surv 2021 Jul;54(6):1-35. [doi: 10.1145/3457607]
156. Zhou N, Zhang Z, Nair V, Singhal H, Chen J. Bias, fairness and accountability with artificial intelligence and machine learning algorithms. Int Statistical Rev 2022 Apr 10;90(3):468-480 [FREE Full text] [doi: 10.1111/insr.12492]
157. Zerilli J, Knott A, Maclaurin J, Gavaghan C. Algorithmic decision-making and the control problem. Minds Mach 2019 Dec 11;29(4):555-578. [doi: 10.1007/s11023-019-09513-7]
158. Lepri B, Oliver N, Pentland A. Ethical machines: the human-centric use of artificial intelligence. iScience 2021 Mar 19;24(3):102249 [FREE Full text] [doi: 10.1016/j.isci.2021.102249] [Medline: 33763636]
159. Monarch R. Human-in-the-Loop Machine Learning Active Learning and Annotation for Human-centered AI. Shelter Island, New York, United States: Manning Publications; 2021.
160. Wu X, Xiao L, Sun Y, Zhang J, Ma T, He L. A survey of human-in-the-loop for machine learning. Future Gen Comput Syst 2022 Oct;135:364-381 [FREE Full text] [doi: 10.1016/j.future.2022.05.014]
161. Timmins KA, Green MA, Radley D, Morris MA, Pearce J. How has big data contributed to obesity research? A review of the literature. Int J Obes (Lond) 2018 Dec;42(12):1951-1962 [FREE Full text] [doi: 10.1038/s41366-018-0153-7] [Medline: 30022056]
162. Sapci AH, Sapci HA. Innovative assisted living tools, remote monitoring technologies, artificial intelligence-driven solutions, and robotic systems for aging societies: systematic review. JMIR Aging 2019 Nov 29;2(2):e15429 [FREE Full text] [doi: 10.2196/15429] [Medline: 31782740]
163. Guisado-Fernández E, Giunti G, Mackey LM, Blake C, Caulfield BM. Factors influencing the adoption of smart health technologies for people with dementia and their informal caregivers: scoping review and design framework. JMIR Aging 2019 Apr 30;2(1):e12192 [FREE Full text] [doi: 10.2196/12192] [Medline: 31518262]
164. Wilmink G, Dupey K, Alkire S, Grote J, Zobel G, Fillit HM, et al. Artificial intelligence-powered digital health platform and wearable devices improve outcomes for older adults in assisted living communities: pilot intervention study. JMIR Aging 2020 Sep 10;3(2):e19554 [FREE Full text] [doi: 10.2196/19554] [Medline: 32723711]
165. Sempionatto J, Montiel V, Vargas E, Teymourian H, Wang J. Wearable and mobile sensors for personalized nutrition. ACS Sens 2021 May 28;6(5):1745-1760 [FREE Full text] [doi: 10.1021/acssensors.1c00553] [Medline: 34008960]
166. Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, et al. Human–machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med 2019 Nov 18;2(1):111 [FREE Full text] [doi: 10.1038/s41746-019-0189-7] [Medline: 31754637]

Abbreviations

AGI: artificial general intelligence
AI: artificial intelligence
BFP: body fat percentage
CNN: convolutional neural network
CV: computer vision
DL: deep learning
DT: decision tree
GFA: group factor analysis
HITL: human-in-the-loop
KNN: k-nearest neighbor
LASSO: least absolute shrinkage and selection operator
MARS: multivariate adaptive regression splines
ML: machine learning
MLP: multilayer perceptron
MRI: magnetic resonance imaging
NB: naïve Bayes
NLP: natural language processing
PCA: principal component analysis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RF: random forest
RNN: recurrent neural network
SVM: support vector machine
WC: waist circumference
WHR: waist-to-hip ratio
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 28.06.22; peer-reviewed by N Maglaveras, B Puladi; comments to author 30.08.22; revised version received 05.10.22; accepted 01.11.22; published 07.12.22.

Please cite as:
An R, Shen J, Xiao Y
Applications of Artificial Intelligence to Obesity Research: Scoping Review of Methodologies
J Med Internet Res 2022;24(12):e40589
URL: https://www.jmir.org/2022/12/e40589
doi:10.2196/40589
PMID:36476515

Review

Desperately Seeking Intersectionality in Digital Health Disparity Research: Narrative Review to Inform a Richer Theorization of Multiple Disadvantage

Laiba Husain[1], MPH; Trisha Greenhalgh[1], PhD; Gemma Hughes[1], DPhil; Teresa Finlay[1], PhD; Joseph Wherton[1], PhD

Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Laiba Husain, MPH
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Rd
Oxford, OX2 6GG
United Kingdom
Phone: 44 7398850015
Email: laiba.husain@phc.ox.ac.uk

Abstract

Background: Digital consultations between patients and clinicians increased markedly during the COVID-19 pandemic, raising questions about equity.

Objective: This study aimed to review the literature on how multiple disadvantage—specifically, older age, lower socioeconomic status, and limited English proficiency—has been conceptualized, theorized, and studied empirically in relation to digital consultations. We focused mainly on video consultations as they have wider disparities than telephone consultations and relevant data on e-consultations are sparse.

Methods: Using keyword and snowball searching, we identified relevant papers published between 2012 and 2022 using Ovid MEDLINE, Web of Science, Google Scholar, and PubMed. The first search was completed in July 2022. Papers meeting the inclusion criteria were analyzed thematically and summarized, and their key findings were tabulated using the Grading of Recommendations Assessment, Development, and Evaluation Confidence in the Evidence from Reviews of Qualitative Research criteria. Explanations for digital disparities were critically examined, and a search was undertaken in October 2022 to identify theoretical lenses on multiple disadvantage.

Results: Of 663 articles from the initial search, 27 (4.1%) met our inclusion criteria. In total, 37% (10/27) were commentaries, and 63% (17/27) were peer-reviewed empirical studies (11/27, 41% quantitative; 5/27, 19% qualitative; 1/27, 4% mixed methods; 1/27, 4% systematic reviews; and 1/27, 4% narrative reviews). Empirical studies were mostly small, rapidly conducted, and briefly reported. Most studies (25/27, 93%) identified marked digital disparities but lacked a strong theoretical lens. Proposed solutions focused on identifying and removing barriers, but the authors generally overlooked the pervasive impact of multiple layers of disadvantage. The data set included no theoretically informed studies that examined how different dimensions of disadvantage combined to affect digital health disparities. In our subsequent search, we identified 3 theoretical approaches that might help account for these digital disparities. Fundamental cause theory by Link and Phelan addresses why the association between socioeconomic status and health is pervasive and persists over time. Digital capital theory by Ragnedda and Ruiu explains how people mobilize resources to participate in digitally mediated activities and services. Intersectionality theory by Crenshaw states that systems of oppression are inherently bound together, creating singular social experiences for people who bear the force of multiple adverse social structures.

Conclusions: A limitation of our initial sample was the sparse and undertheorized nature of the primary literature. The lack of attention to how digital health disparities emerge and play out both within and across categories of disadvantage means that solutions proposed to date may be oversimplistic and insufficient. Theories of multiple disadvantage have bearing on digital health, and there may be others of relevance besides those discussed in this paper. We call for greater interdisciplinary dialogue between theoretical research on multiple disadvantage and empirical studies on digital health disparities.

(J Med Internet Res 2022;24(12):e42358) doi:10.2196/42358

KEYWORDS

digital health disparities; video consultations; intersectionality; health inequity; narrative review; digital capital; fundamental cause theory; mobile phone

Introduction

Background

COVID-19 has thrown a spotlight on digital health disparities. Before the pandemic, patients with poorer self-reported health, of older age and lower incomes, and from certain minority ethnic groups were less likely to access health care by technological means [1,2]. The proportion of health care consultations conducted remotely (telephone, video, and web-based) increased dramatically during the pandemic [3], chiefly because many face-to-face consultations were canceled owing to the risk of transmitting the virus. Therefore, the shift to remote consultations affected a greater number of patients with increased access burdens because of being disadvantaged through poverty, low health literacy, limited English proficiency (LEP; in countries where English is the main language), or lacking digital skills or devices [4,5].

There is a vast amount of research on health disparities in general; this literature falls into 3 broad categories corresponding to 3 longitudinal phases. The first phase, detection, involves defining health disparities, identifying vulnerable populations, and developing valid measures for studying both. The second phase, understanding why disparities exist, involves identifying factors that explain gaps in health care between vulnerable and less vulnerable groups. The third phase involves the development, implementation, and evaluation of interventions that reduce or eliminate health disparities. These different kinds of research are all relevant to the study of digital disparities as well, although the literature on the latter is currently sparse in both volume and depth.

Digital health is sometimes presented vaguely and futuristically as having the potential to strengthen health systems and public health, improve efficiency, and increase equity in access to health services [6,7]. Video consultations in particular have been extensively researched (often in randomized controlled trials in comparison with face-to-face consultations) and depicted with promissory claims of delivering efficient care without compromising safety or patient satisfaction [8-10]. However, as the pandemic showed, digital *solutions* intended to reduce inequalities may actually widen them [11-13]. Broadly speaking, and with some notable exceptions [14-16], technologies are least frequently and least readily used by limited English–speaking communities, those with low income, and older adults—and especially by those in the *triple jeopardy* of all 3 groups.

Uptake of video consultations, for example, is known to be low among various disadvantaged groups [17]. This and other digital disparities have been explained by multiple factors, including lack of access to technology, low digital literacy, suboptimal internet coverage, and power differentials within the home in terms of who has access to digital devices [18-20].

Several recent publications have proposed strategies to ensure that the emergence of digital services does not exacerbate disparities in access to health care and health outcomes [11,21-23]. In total, 2 broad approaches have been taken. One approach speculates that digitally driven efficiency savings could *free up* staff to attend to the disadvantaged, who would continue to consult in traditional ways. However, there is limited evidence that such savings occur even with telephone consultation services and—to our knowledge—no evidence that they occur with video services (which have higher setup costs and require staff and patients to learn new skills) [24].

Another approach centers on identifying and removing barriers to video access among disadvantaged groups—for example, ensuring that people are equipped, competent, and confident in using the video modality where appropriate. This strategy is founded on an individual deficit model that depicts the disadvantaged as deficient in certain things (eg, knowledge, confidence, and bandwidth) and assumes that these deficiencies can be rectified by specific inputs (eg, training, practice, and digital upgrades). Thus, it tends to overlook the pervasive impact of multiple layers of structural disadvantage. Of particular interest is how key risk factors for digital exclusion such as LEP, poverty, and older age are often mutually reinforcing, an effect that some have called intersectionality [25].

These approaches are discussed throughout this paper, which starts by outlining the aim, scope, and research questions of the review along with definitions of important terms and concepts. We then explain our methodological approach to the review and the details of our methods. Our findings show that, in the relatively sparse literature uncovered on the topic, substantial digital disparities are reported and that this research to date in relation to video consultations has been almost entirely descriptive rather than explanatory. We also describe how we identified 3 candidate theories of multiple disadvantage that could further enhance our understanding of digital disparities. We conclude by proposing 3 candidate theories that may have particular relevance in explaining and helping address digital health disparities in people with multiple disadvantage.

Aim, Scope, and Research Questions

In this narrative review, we sought to explore how published studies of disparities in digital health consultations have defined, theorized, and empirically tested the concept of multiple disadvantage.

To sharpen our focus in a potentially vast field, we chose to restrict our sample of empirical digital health research to studies of video consultations between patients and health care professionals as this is where the most dramatic differences in digital access have been documented in the literature [8]. In contrast, there has been little research on digital disparities in electronic consultations [26], and research on telephone consultations suggests that digital disparities are less marked

[27]. On the basis of findings from our previous work on remote consulting [28], we decided to focus particularly on studies that provided insights relevant to an underresearched group: older adults with low income and LEP.

Our research questions were as follows: (1) How have the intersecting effects of age, socioeconomic status, and LEP been conceptualized, theorized, and studied empirically in relation to digital consultations (especially video)? (2) What interventions have been developed and tested within the context of digital consultations to try to overcome the effects of multiple disadvantage? (3) What were the findings of these studies and how can they help us extend theory and inform future research? and (4) What are the implications for policy and practice?

Definitions

This review covers a number of closely related terms and concepts, which we define and discuss in Textbox 1.

Textbox 1. Concepts and definitions.

Health disparities

- Defined by the Centers for Disease Control and Prevention of the United States (as this term has almost exclusively been used in the United States) as differences in health status or health outcomes among population groups as a result of—for example—social, economic, racial, or ethnic characteristics [29].

Health equity

- Defined by the World Health Organization (WHO) as "the absence of avoidable, unfair, or remediable differences among groups of people, whether those groups are defined socially, economically, demographically or geographically or by other means of stratification" [30]. Others have defined health equity in positive terms as "the attainment of the highest level of health for all people...valuing everyone equally with focused and ongoing societal efforts to address avoidable inequalities, historical and contemporary injustices, and the elimination of health and healthcare disparities" [31]. Thus, reducing health disparities is one aspect of achieving health equity.

Health inequity

- Refers to the presence of these avoidable, unfair, or remediable differences [31]. Some authors have distinguished health inequities from *health inequalities*, the latter being disparities that are explained by differences that are not avoidable or remediable (eg, because of age) [32] but, in practice, these terms tend to be used interchangeably.

Digital health

- An interdisciplinary field linking technologies (software, hardware, and underpinning infrastructure) and the service models in which they are used [33]. It includes mobile health apps, electronic health records, electronic medical records, wearable devices, remote consultations (by telephone, video, or the web), and remote monitoring of various kinds.
- Digital health includes *telemedicine*, which the WHO set a standardized definition for in 2007 as the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment, and prevention of disease and injuries; research and evaluation; and the continuing education of health care providers, all in the interest of advancing the health of individuals and their communities [34].
- The terms *telemedicine* and *telehealth* are often used interchangeably, but telehealth has evolved to encapsulate a broader array of digital health care activities and services, including patient and professional health-related education, public health, and health administration [34].
- *Video consultations* are a specific type of telemedicine involving a video connection.

Digital divide

- This constitutes a societal division between those who have the means and capability to make full use of digital technology and those who lack those means for reasons relating to (for example) income, education, or age [35].

Digital health disparities

- A concept that emerged recently; refers to inequalities that may be widened when technologies are required for accessing and receiving care. One author has coined the expression "digital inverse care law" to depict how people who are most in need of care (in particular, older people and those experiencing social deprivation) are least likely to access or receive it through digital platforms [36].
- For consistency in this paper, we have chosen to use the term *disparities* rather than *inequities* or *inequalities*.

Disadvantage

- Defined as those for whom the social conditions in which they are born, live, and age do not ensure opportunities for them to be healthy and flourish [37]. Disadvantaged people are disproportionately affected by disease, dysfunction, and ill health.
- *Underserved* and *marginalized* populations include people who experience discrimination of any kind and encounter barriers (eg, racial, ethnic, gender, sexual orientation, economic, cultural, or linguistic) to accessing health care goods and services [38]. They tend to receive fewer and lower-quality health care and public health goods and services, have a lack of familiarity with the health care delivery system, face a shortage of readily available providers, and lack access to quality systems of care.

Intersectionality

- Refers to the idea that systems of oppression are inherently bound together, thus creating singular social experiences for people who bear the force of multiple systems [25]. It has been defined as "the relationships among dimensions and modalities of social relations and subject formations" [39]. A more specific definition in relation to health disparities is the "intersections of individuals' multiple identities within social systems of power that compound and exacerbate experiences of ill health" [40], thus recognizing that health is shaped by a multidimensional overlapping of factors such as race, class, income, education, age, ability, sexual orientation, immigration status, ethnicity, indigeneity, and geography.

Methods

We undertook a narrative review of the literature published from 2012 to 2022 focusing on digital health disparities in disadvantaged groups, with a specific focus on older, low-income, limited English–speaking individuals and on video consultations. Various combinations of search terms, including those regarding age, language, and income, were trialed as part of the initial search strategy but, because of the unique combination of terms being used, they yielded no relevant results. Following discussion with an expert librarian and coauthors, the search strategy shown in Textbox 2 was applied and updated with further terms as the study progressed to reflect the developing field. Although the search range extended back to 2012, only one study in our final sample was published before 2020. Earlier studies were clearly superseded by later work because of technologies having been developed at pace and the evolving field of digital health care. An evaluation of a video consulting service in Scotland by Wherton et al [41] reported that, as late as 2017, the platforms used for video consultations were designed for videoconferencing rather than video consulting and were expensive, clunky, unreliable, and poorly aligned with clinical workflows. A few years later, bespoke video technologies for health care encounters had been developed; they were cheaper, more agile, and better designed around key workflows. Accordingly, studies undertaken before the development of mature, fit-for-purpose technologies were less relevant. Similarly, when looking to the literature for explanations of digital health disparities, little had been published before 2012, and earlier studies reflected challenges that are no longer relevant today, such as website provision and public health dissemination through digital television [42].

We drew in particular on 3 methodological sources. First, we aligned with Greenhalgh et al [43], who highlighted the purpose of narrative review (to achieve clarification and understanding across a broader field of inquiry) and distinguished this from that of quantitative systematic review (to identify, summarize, and synthesize data on a narrowly focused topic, typically through meta-analysis). The latter relies on largely technical processes (eg, data extraction and the use of risk-of-bias tools), whereas the former requires the progressive development and refinement of an argument through interpretative methods.

Second, we engaged with the methodology by Boell and Cecez-Kecmanovic [44] for hermeneutic review, which applies the hermeneutic circle (progressively adding parts to the whole) to secondary research. The hermeneutic review begins with a close reading of key texts—those known to the research team and those identified on an initial scoping search. These texts are mapped and classified according to coherence, adequacy, and relevance (with relevant extracts as quotes) in interim summaries. Some summaries are *deep dives* on specific themes (eg, how a particular theory has been applied). Some are broader but less deep (eg, an early draft of the review findings to be progressively refined as the study unfolds). The researchers move back and forth between further searches and a progressively richer overall summary.

Third, we took note of methodological guidance from a group of journal editors in the health care field [45], who developed a structured critical appraisal tool (SANRA [Scale for the Assessment of Narrative Review Articles]). SANRA defines quality in narrative reviews in terms of strong justification for the importance and aims of the review, a well-described and well-justified literature search, claims backed up by referencing relevant primary studies, the quality of scientific reasoning (including a nonselective approach to inclusion of studies and study designs appropriate to the research question), and appropriate presentation of data.

The following search terms were developed by author 1 and discussed with the coauthors: *remote consultations* (or *virtual consultations* or *video consultations* or *video visits* or *telemedicine* or *telehealth*) and *digital health inequality* (or *digital divide* or *inequity/ies* or *inequality/ies* or *health disparity/ies* or *disadvantage*). These terms were then checked by the expert university librarian. Following some pilot searches, these were developed into formal search strings (Textbox 2).

The search was first completed in July 2022 and then supplemented with an additional search with revised search terms in October 2022. The results of all these searches were combined into a single data set, and irrelevant studies were excluded by screening titles and abstracts. All papers were reviewed by author 1, and each paper was second reviewed by at least one coauthor. Rigor was strengthened by reflexivity and the application of the Grading of Recommendations Assessment, Development, and Evaluation Confidence in the Evidence from Reviews of Qualitative Research checklist [46] to determine confidence in the findings.

To gain familiarity and aid data management, author 1 extracted and tabulated the following information from the included sources: (1) study design, setting, and sample; (2) key findings in relation to health inequalities and web-based care; (3) key issues raised by the authors; (4) authors' recommendations for how to promote health equity in the context of remote care; and (5) any theories or frameworks used by the authors to study digital disparities.

Following the hermeneutic circle, we worked back and forth between individual papers and our overview of the findings, progressively adding detail. The process of interpretive synthesis was aided by discussions among the coauthors. When we realized that the initial data set of 27 papers included very little theoretical analysis, we conducted a further search for theories of multiple disadvantage that were potentially relevant to digital health disparities and had the potential to add a theoretical depth to our original sample of papers. This second search was deliberately not exhaustive; it included asking experts in the field, using key sources known to the authors, and searching the PubMed database for the term *theor** along with the selected terms listed in Textbox 2.

A near-final version of the synthesis was shared with a wider group of peers, including experts in various aspects of health disparity or digital health, and with laypeople with lived experience of accessing care. The synthesis was refined in response to their feedback.

Textbox 2. Search strategy with updated terms.

Ovid MEDLINE

- Search in title, abstract, keywords, and subject headings
- Search limited to the years 2012 to 2022
- Search string: ("remote consultations" or "virtual consultations" or "video consultations" or "video visits" or telehealth or telemedicine) AND ("digital health inequalit*" or "digital divide" or inequit* or inequalit* or "health disparit*" or disadvantag*)

Web of Science

- Search in title, abstract, keywords, and subject headings
- Search limited to the years 2012 to 2022
- Search string: ("remote consultations" or "virtual consultations" or "video visits" or telehealth or telemedicine) AND ("digital health inequalit*" or "digital divide" or inequit* or inequalit* or "health disparit*" or disadvantag*)

Google Scholar

- General search in Google Scholar
- Search limited to the years 2012 to 2022
- Search string: ("remote consultations" or "virtual consultations" or "video visits" or telehealth or telemedicine) AND ("digital health inequalit*" or "digital divide" or inequit* or inequalit* or "health disparit*" or disadvantag*)

Sources known to the research team and their networks

- Key studies already on file
- Asking expert colleagues to recommend sources

Forward and backward reference searching

- Identifying highly relevant papers cited by the included papers
- Identifying highly relevant papers that cited the included papers

Results

Description of the Data Set

In this section, we describe the main findings of the empirical studies included in this narrative review (Table 1) as well as key points from commentaries on the theme of inequity of access (Table 2). All the authors of both commentaries and empirical studies offered a list of proposed solutions to digital disparities. These are further summarized and categorized (Textbox 3). Figure 1 presents a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart demonstrating the number of papers identified, included, and excluded.

The 17 peer-reviewed research papers (summarized in Table 1) comprised 1 (6%) systematic review [48], 1 (6%) narrative synthesis [49], 7 (41%) retrospective audits of medical records [56,57,59-62,70], 3 (18%) quantitative surveys [5,50,58], and 5 (29%) qualitative studies based on semistructured interviews [5,51,52,54,55]. A further 37% (10/27) of the articles (Table 2) commented on others' research, reflected on findings from clinical practice, or proposed measures to reduce digital health disparities [11,21,22,63-69].

Table 1. Characteristics of the empirical studies.

Author, year, and country	Study design	Setting	Sample	Aim of research
Parker et al [48], 2021, United Kingdom	Systematic review	Primary care (any country)	Studies that compared remote and face-to-face consultations (until June 2020)	To explore the impact of remote consultations on use and clinical outcomes in disadvantaged groups
Litchfield et al [49], 2021, United Kingdom	Rapid review and narrative synthesis	UK primary care	Studies that explored various constructs within the 3 domains of the digital divide framework	To identify how this "digital divide" was manifested during the first wave of the pandemic and highlight any areas that might be usefully addressed for practice beyond the pandemic
Chang et al [50], 2020, United States	Quantitative study using rapid response surveys	Small primary care practices in low-income, minority, or migrant areas of New York City	5372 primary care providers contacted in 5 waves	To understand how primary care practices were responding to the COVID-19 pandemic and examine whether telemedicine use and barriers differed based on the socioeconomic characteristics of the communities served
Eberly et al [23], 2020, United States	Quantitative retrospective electronic record audit	Large academic health system	2940 patients who had scheduled a remote consultation	To compare patients who completed telemedicine encounters with patients who were scheduled but did not complete a visit early in the COVID-19 pandemic
Fu et al [5], 2022, United Kingdom	Mixed methods—quantitative survey before and during the pandemic and qualitative data from free-text notes	DOTW[a] United Kingdom—a third-sector organization serving migrants with drop-in clinics	Migrant patients (survey: n=6268; free-text analysis: n=96)	To understand the living conditions, changes in the service user profile, and needs of vulnerable migrants trying to access health care both before and during the pandemic (when DOTW services moved to telephone)
Stevens et al [51], 2021, United Kingdom	Qualitative interview study ("rapid health needs assessment")	Across England	42 interviews with people who experienced social vulnerability and faced barriers to accessing health care	To assess the experiences of socially vulnerable people during the COVID-19 pandemic
Kaihlanen et al [52], 2022, Finland	Qualitative semistructured interview study	National study, mostly via third-sector organizations working with vulnerable groups	N=74, including older adults, migrants, mental health service users, high users of health services, and unemployed individuals	To examine the challenges experienced by vulnerable groups using digital health services during the COVID-19 pandemic
Knights et al [53], 2021, United Kingdom	Qualitative semistructured interview study	Urban, suburban, and rural settings in England	48 clinicians, 16 administrative staff, and 17 migrant patients	To understand the pandemic's impact on recent immigrants and their access to primary health care and implications for vaccine rollout
Alkureishi et al [54], 2021, United States	Qualitative semistructured interview study	Primary care clinics linked to a Chicago academic research center	54 interviews with adult patients and parents of child patients who had web-based visits (March 2020-September 2020)	To understand service users' perspectives on (1) the definition, causes, and impact of the digital divide; (2) whose responsibility it is to address this divide; and (3) potential solutions to mitigate this
Donaghy et al [55], 2019, United Kingdom	Qualitative semistructured interview study	Scotland primary care	Patients (n=21) and primary care clinicians (n=13)	To explore the views of physicians, nurses, and patients who have experienced a web-based consultation
Rodriguez et al [56], 2021, United States	Quantitative cross-sectional study using electronic health record data	Across primary and specialty care in Boston, Massachusetts	Data from 162,102 patients across 1652 primary and specialty care practices	To use data from a large, integrated health system to determine patient, clinician, clinic, and neighborhood characteristics associated with visit modality

Author, year, and country	Study design	Setting	Sample	Aim of research
Hsueh et al [57], 2021, United States	Quantitative retrospective cross-sectional study	Primary care across Kaiser Permanente, Northern California	955,352 patient portal self-scheduled primary care telemedicine visits	To test the hypothesis that limited English proficiency would be associated with lower video use compared with telephone, especially among patients without previous video visit experience
Yu and Hagens [58], 2022, Canada	Quantitative cross-sectional web survey	Across Canada	2303 older adult Canadians	To investigate socioeconomic disparities in the demand for and use of web-based visits during the COVID-19 pandemic among older adults in Canada
Schifeling et al [59], 2020, United States	Quantitative retrospective, cross-sectional analysis	2 primary clinics in Colorado	192 appointments reviewed	To determine (1) whether video visits had longer duration, more visit diagnoses, and more discussions than telephone visits in the rapid implementation of telemedicine during the pandemic and (2) whether disparities in visit type existed based on patient characteristics
Sachs et al [60], 2021, United States	Quantitative cross-sectional analysis	Oregon Health and Science University	134,274 ambulatory patients	To evaluate for demographic disparities in the use of telehealth modalities
Broffman et al [61], 2022, United States	Quantitative cross-sectional analysis	Electronic platform across the United States	2847 men	To examine the complex relationship between individual and environmental characteristics, broadband access, device type, and telehealth use as it relates to the digital divide
Zachrison et al [62], 2021, United States	Quantitative retrospective analysis	Electronic health record data across the Northeastern United States	N=1,241,313 individual health records	To describe patient characteristics associated with successful transition from in-person to web-based care and video vs audio-only participation

[a]DOTW: Doctors of the World.

Table 2. Characteristics of the commentaries and editorials.

Author, year, and country	Context	Aim of paper
Nouri et al [22], 2020, United States	Commentary emerging from practice, by clinicians at an academic medical center in San Francisco with 3 clinics including an urban "safety-net" service	Discussion of challenges encountered in ensuring equitable access to telemedicine in the early weeks of the pandemic
Ramsetty and Adams [63], 2020, United States	Editorial by directors of free CARES[a] clinics in South Carolina	To discuss disparities in access to telemedicine among vulnerable patients and evaluate why patients could not access the web-based system at CARES clinics
Mehmi et al [64], 2020, United Kingdom	Commentary on 2 articles, one arguing for the benefits of video consultations and one about health inequalities exposed by the pandemic	Authors criticize a BMJ article for failing to address digital disparities, notably relating to lack of effective internet access in some geographical areas, digital poverty (inability to afford a device or adequate data package), poor digital skills and confidence, refugee and other uncertain citizenship status, and lack of space and privacy at home
Thronson et al [65], 2020, United States	Commentary on an empirical audit showing that the pandemic led to fewer primary care encounters overall and many more conducted remotely	Authors comment that the empirical study failed to capture a key finding from their own clinical experience (supported by audit data): that web-based visits (by video) were rarely taken up by the homeless, limited-English speakers, and those in a "racially diverse safety-net population"
Ramasawmy et al [66], 2021, United Kingdom	Commentary on how the move to digital could increase many well-documented inequities	Summarizes the literature on health inequities, including key reports from the past; warns that these inequities could increase with "digital first" policies; and highlights areas in which existing knowledge and evidence might be translated into cross-sectoral action
Gray et al [67], 2020, United States	Editorial from the Department of Internal Medicine at Ohio State University	To explore the strategies for digital care of vulnerable patients in a COVID-19 world; recommends 5 key strategies to prevent losing touch with vulnerable patients who are alienated by the digital divide
Crawford and Serhal [21], 2020, United States	Commentary summarizing existing literature and offering a new framework	Authors introduce the Digital Health Equity Framework to identify the digital determinants of health and their links to digital health equity; aim is to establish systematic ways to ensure that health inequities are identified and addressed in digital health policies and programs
Rodriguez et al [11], 2020, United States	Opinion piece on digital health equity	To discuss views on how the digital divide should be considered in the implementation of recent policy (21st Century Cures Act)
Eruchalu et al [68], 2021, United States	Commentary emerging from practice (New York City)	To discuss concerns about inequities in digital health access
Gallegos-Rejas et al [69], 2022, Australia	Article proposing a series of practical steps to improve access to telehealth services	Summarizes selected strategies to improve equity of access to telehealth for stakeholder groups: consumers (patients and carers), consumer advocacy groups, health service staff (clinicians), health services (providers), policy makers, funders, and researchers

[a]CARES: Community Aid, Relief, Education, and Support.

Textbox 3. Summary of solutions to digital disparities proposed by authors of primary studies and commentaries.

Policy and government

- Finance and governance
  - Extension of temporary waivers by private medical programs for telemedicine beyond the end of the public health emergency declaration
  - Develop targeted payment mechanisms to reimburse providers for helping patients adapt to video-enabled telemedicine
  - Payment parity between insurers for video and audio visits
  - Clarify standards for design of digital health innovations
  - Secure funding for projects that address equity and access to health services (including telehealth) with focus on patient experience and acceptance
- Internet access
  - Improve distribution of video-enabling devices to those unable to afford them
  - Expand device and broadband internet access
  - Install free, bookable, soundproofed video booths in community centers, libraries, or physicians' offices to ensure privacy, with staff available to help with using technology
- Evaluation
  - Identify and monitor disparities in access
  - Incentivize quality improvement programs based on equity-related outcomes

Organization and health system

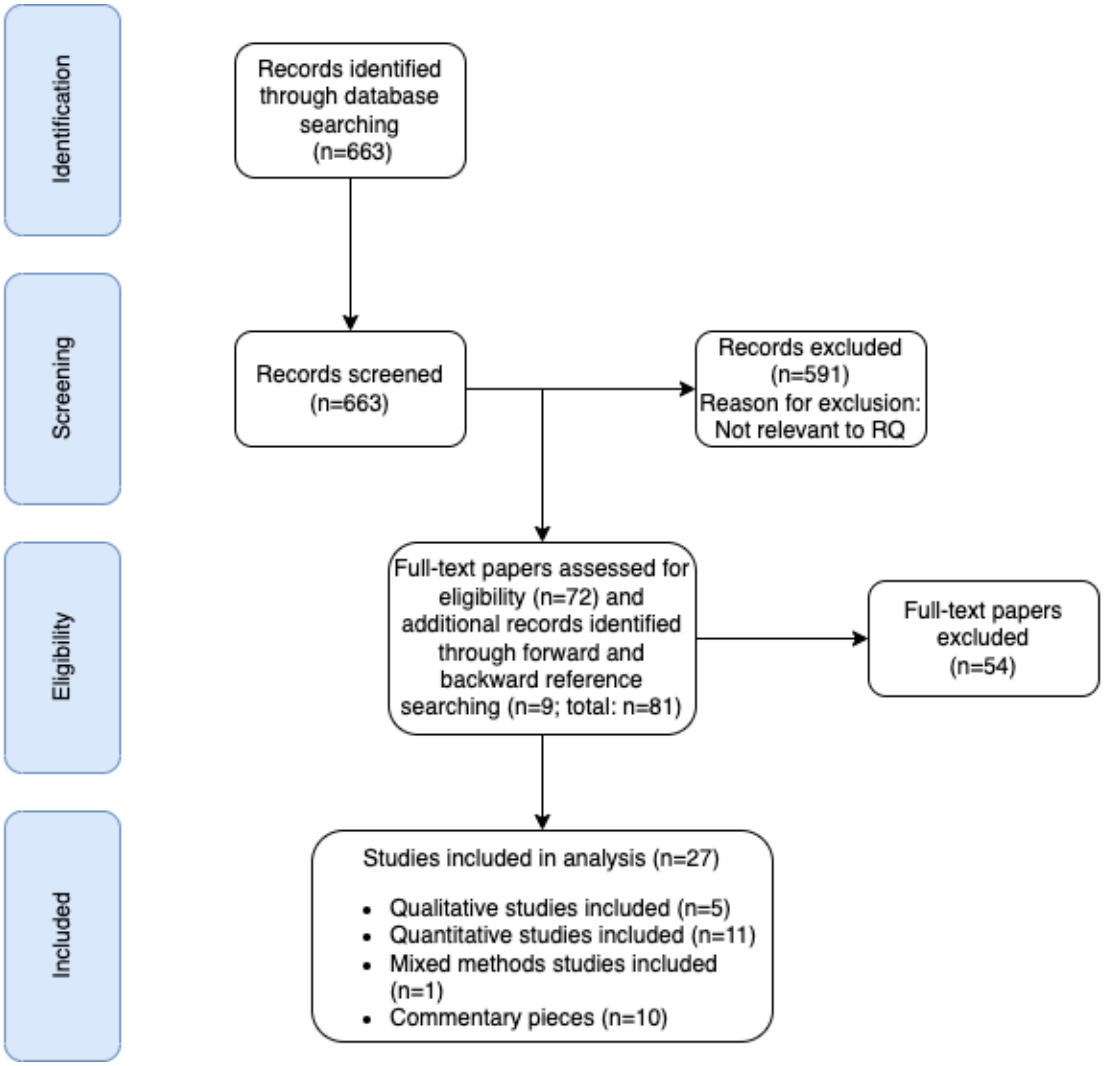
- Staff training
  - Increase system leadership awareness of barriers to telemedicine
  - Engage community health workers
  - Promote empathy and bedside manner
  - Provide clinical telehealth training to all staff
- Service delivery and choice
  - Offer digital services to all patients
  - Targeted access slots
  - On-the-day appointments reserved for marginalized patients
  - New models of care such as web-based group consultations
- Technology
  - Provide different modality options, including high- and low-technology forms
  - Explore technologies that supplement or simulate face-to-face interactions during web-based consultations
  - Develop device loan schemes to support those who would benefit from telehealth interventions but do not have access to equipment

Patients and citizens

- Translation and communication
  - Provide interpreting services
  - Translation of relevant documents
  - Information and guidance that are more inclusive and relevant to those living in challenging and vulnerable circumstances
  - Use of tailored translated texts and text templates to encourage access
- Education and awareness
  - Develop programs to improve general and health technology literacy
  - Increase public awareness of available resources

- Community engagement and co-design
  - Involve marginalized people in co-design and data stewardship
  - Develop and evaluate evidence-based health care communication protocols for telemedicine practice to help providers create a patient-centered experience

Figure 1. Flow diagram of the search process (adapted from the PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] guidelines [47]). RQ: research question.



Descriptive Findings

Formal research studies on this topic in the early months of the pandemic were few in number, mostly small in size, and conducted rapidly. As such, our data set included no in-depth, theoretically informed empirical studies that had set out to explain how different dimensions of disadvantage combined to affect digital health disparities. During 2020, several clinical authors were moved to write urgent editorials and commentaries on the theme of inequity of access as frontline services shifted from face-to-face to remote modes. The summary of the review findings using the Grading of Recommendations Assessment, Development, and Evaluation Confidence in the Evidence from Reviews of Qualitative Research checklist [46] to determine confidence can be found in Table 3.

Table 3. Grading of Recommendations Assessment, Development, and Evaluation Confidence in the Evidence from Reviews of Qualitative Research (GRADE-CERQual) summary of findings.

Summary of review findings	Studies contributing to the review findings	Methodological limitations	Coherence	Adequacy	Relevance	CERQual assessment of confidence	Explanation of CERQual assessment
Low household income, older age, and ethnic minority background (especially limited-English speakers) were all independently associated with lower uptake of video consultations.	• Fu et al [5] • Moher et al [47] • Parker et al [48] • Litchfield et al [49] • Eberly et al [70] • Rodriguez et al [56] • Zachrison et al [62] • Stevens et al [51] • Donaghy et al [55] • Ramsetty et al [63] • Thronson et al [65] • Gray et al [67] • Eruchalu et al [68] • Mehmi et al [64] • Ramasawmy et al [66]	Some concerns about reflexivity [47-49,51,55,63-66,70], recruitment [5,49,55,65-68,70], and analytical rigor [5,47-49,56,63,64,67,70]	• No concerns—strongly evidenced with qualitative or quantitative data	• No concerns [49,51,62,64,67,68] • Minor concerns [5,47,48,55,56,63,65] • Moderate concerns [66,70]	• No concerns [5,47,49,51,55,56, 63-68] • Minor concerns [48,51,62,70]	High confidence	15 studies contributing data with good coherence and few other concerns
Research into digital health disparities, at least in relation to video consultations, has to date been almost entirely descriptive rather than explanatory.	• All studies	Some concerns about reflexivity [47-49,64,65,67,68,70], recruitment [5,49,56,57, 70], analytical rigor [5,47,48,56,61,63,66,70], and ethical considerations [56]	• No concerns	• No concerns [49,51,57,62,64,65] • Minor concerns [5,47,48,55,56,61,66,68] • Moderate concerns [63,67,70]	• No concerns	High confidence	All studies contributing data with good coherence and few other concerns
The higher the patient's social vulnerability index, the more likely the consultation occurred by phone instead of video.	• Eberly et al [70] • Thronson et al [65] • Eruchalu et al [68] • Mehmi et al [64] • Ramasawmy et al [66]	Some concerns about analytical rigor, recruitment, and reflexivity [64,66,68]	• No concerns [64-66,68] • Minor concerns—limited sample size and not explored in detail in the study [70]	• No concerns [64-66,68] • Moderate concerns—limited sample size, not sufficiently "rich" data [70]	• No concerns [64-66,68] • Minor concerns [70]	High confidence	Only one study contributing data with moderate concerns about sample size

Summary of review findings	Studies contributing to the review findings	Methodological limitations	Coherence	Adequacy	Relevance	CERQual assessment of confidence	Explanation of CERQual assessment
Access to digital health services is hampered by insufficient digital or local language skills.	• Parker et al [48] • Rodriguez et al [56] • Hsueh et al [57] • Zachrison et al [62] • Donaghy et al [55] • Ramsetty et al [63] • Thronson et al [65] • Gray et al [67] • Mehmi et al [64] • Ramasawmy et al [66]	Some concerns about reflexivity [48,56,57,64,65,67], analytical rigor [48,56,63,66], and recruitment [57]	• No concerns—strongly evidenced with qualitative or quantitative data	• No concerns [50-52,54,55,57-67] • Minor concerns [48,56]	• No concerns [55,56,63,65-67] • Minor concerns [48,61,62,64]	High confidence	Few concerns
Digitization and web-based consultations amplified existing inequalities in access to health care for many migrants because of a lack of digital literacy and access to technology compounded by language barriers and indirect discrimination.	• Fu et al [5] • Rodriguez et al [56] • Stevens et al [51]	Some concerns about recruitment [5,57], analytical rigor [5,56], and reflexivity [56,57]	• Minor concerns—strongly evidenced with qualitative data but varying experiences in some studies	• No concerns [51] • Minor concerns [5,56]	• No concerns [5,56] • Minor concerns [51]	High confidence	3 studies contributing data with good coherence and few other concerns
Although demand for video consultation services in primary care is likely to rise, for complex or sensitive problems, face-to-face consultations remain preferable.	• Broffman et al [61]	Some concerns about analytical rigor	• Minor concerns given methodological limitations	• Minor concerns	• No concerns	Moderate confidence	Only one study contributing data with minor concerns about analytical rigor

Systematic and Narrative Reviews

Studies published in late 2020 and early 2021 included rapid systematic reviews comparing the impact of face-to-face and remote consultations on a range of predefined variables. Parker et al [48], for example, reviewed quantitative studies from before the pandemic to mid-2020 across socioeconomic and disadvantaged groups in the United Kingdom. A total of 13 studies met their inclusion criteria, and they found that phone consultations were used more by young people of working age,

nonimmigrants, and women. Asynchronous web-based consultations were used more by affluent and educated people. Findings in relation to socioeconomic status and ethnicity were inconsistent across primary studies, and none of the included studies reported on quality of care or clinical outcomes.

In a rapid review and narrative synthesis of UK primary care by Litchfield et al [49], 9 studies were identified that explored various constructs within three domains of the digital divide: (1) digital access, within which one study described continuing issues with internet connectivity among vulnerable patients in the United Kingdom; (2) digital literacy, where 7 studies described how ethnic minorities and older adults were less likely to use digital technologies for accessing care; and (3) digital assimilation, where one study described how video technologies can reduce feelings of isolation and another described how older Black men were the most likely group to share information about COVID-19 on social media platforms. This review also found a large number of opinion pieces and editorials on digital disparities.

Quantitative and Mixed Methods Studies

Nouri et al [22] audited the uptake of remote consultations by demographic group in 3 clinics at an academic medical center in San Francisco. Their paper included a literature review and empirical data from the authors' own 2 general practices. They showed that the shift to remote consulting had been associated with a decrease in the number of consultations with older adults, those of low socioeconomic status, low–health-literacy groups, limited-English speakers, and Black and Asian minority ethnic groups. The authors offer a framework for addressing inequities, which includes goals such as improving digital literacy and resource barriers, removing health system–created barriers, and advocating changes to support sustained and equitable access.

Rodriguez et al [56] also found lower use of video versus telephone visits among older Black, Hispanic, and Spanish-speaking patients, which extends previous findings that showed decreased telemedicine use among patients with LEP [70]. This study also found that clinicians and practices largely drove this variation in the use of video versus telephone visits, suggesting an important target for intervention.

Hsueh et al [57] similarly found that patients with LEP chose video consultation options less often than those without LEP. However, they also found that, once patients with LEP had video visit use experience, they were not different from patients without LEP in likelihood to reuse video visits.

The survey of small primary care providers in New York City by Chang et al [50] uncovered the percentage of encounters undertaken by telephone, video, or web-based patient portal or face-to-face as well as barriers to remote consultations. A key finding was that the higher a patient's social vulnerability index, the more likely the consultation occurred by phone instead of video. This was similarly echoed by Broffman et al [61] and Zachrison et al [62], who also found that patients of color, who have historically experienced the greatest access disparities, are significantly more likely to use smartphones to access telehealth compared with a computer. This provides insight into the extent to which low-bandwidth telehealth is accessible under certain conditions.

Although Chang et al [50] conducted a relatively small survey with a low response rate (exact figures not given), their findings align with those of other studies. For example, a retrospective case note review of 80,000 American patients from March 2020 to May 2020 who had completed a telemedicine visit (from a total of 140,000 who had scheduled such a visit) found that, overall, 46% of teleconsultations were by video and 54% were by phone [70]. Factors independently associated with fewer completed telemedicine visits included older age, preference for languages other than English, Asian ethnicity, and Medicaid insurance (indicating low income). Higher rates of telephone consultations were associated with female sex, Black or Latino ethnicity, older age, and lower household income. This was also the case with the study by Schifeling et al [59], where more than half of the older patients did not use video visits, especially if they were from racial or ethnic minority backgrounds or Medicaid beneficiaries, and with the cross-sectional analysis by Sachs et al [60] showcasing seniors, non-English speakers, and Black patients to be more reliant on telephone than video for care. Although these large studies provided important quantitative information about individual risk factors for a failed telemedicine visit (and for a telephone visit over a video visit), they were not designed to explore the interaction between different independent variables. Further theoretical explanations of *how* different risk factors, individually and in combination, contributed to digital inequities in certain vulnerable population groups are necessary to fully understand these data.

Similarly, a mixed methods study based in the United Kingdom by Fu et al [5] found that there was a reduction in video consultations in older users, undocumented migrants, and individuals with poor health, which could mean that those in the greatest need were being excluded, perhaps because of the digital divide evidenced in some groups of migrants before the pandemic. Yu and Hagens [58] echoed similar findings in older adults in Canada, with results highlighting socioeconomic disparities among older adults that could potentially explain this trend, including lower income and education levels that may act as barriers for older adults to acquiring the skills and technologies necessary to use more complex solutions such as video. As Fu et al [5] stated, "those in the greatest need of health care appeared to be less able to access remote services."

Qualitative Interview Studies

The qualitative rapid health needs assessment by Steven et al [51] across the United Kingdom found that all groups studied experienced challenges in accessing and following COVID-19 information and government guidance, attributed variously to lack of access to digital technology, lack of translated resources, absent or inadequate tailored support, and inadequate housing. Changes in the organization and delivery of health care services, including closure of outreach and drop-in services, remote consultations, and web-based patient registration, worsened existing barriers to accessing health care.

The semistructured interview study by Kaihlanen et al [52] sought to explore and explain the challenges related to the use of digital health services in Finland through the lens of a digital

health equity framework. They found that access to digital health services was hampered by insufficient digital or local language skills.

Knights et al [53] found that digitization and web-based consultations appeared to have amplified existing inequalities in access to health care for many migrants in Finland because of a lack of digital literacy and access to technology compounded by language barriers and indirect discrimination (eg, telephone-only booking services become inaccessible to those without a phone). Health care professionals perceived low digital literacy among migrants and were concerned that web-based consultations resulted in difficulties building trust and risked missing safeguarding cues. These semistructured interviews were conducted by phone, and the sample of migrants was small and skewed; the study was not designed to capture individual, contextualized narratives, and findings were largely impressionistic.

The qualitative study by Alkureishi et al [54] of patient perspectives on the digital divide in US primary care settings explained the concept to participants as follows: "there are people that have and can use technology like computers and the internet. But there are also people that do not have or cannot use this kind of technology. So there is a split or a divide, between people that have and know how to use technology, and those that do not" [61]. These authors found that patients were very aware of the digital divide and described the impacts beyond health care, including employment, education, community and social contexts, and personal economic stability. These participants viewed access to technology and digital skills as important influencers of health disparities.

The clinician-patient relationship was a theme in a qualitative study of patients' and clinicians' experiences with video consultations in general practice in Lothian, Scotland [55]. Although the sample size of this study was comparable with that of the study by Knights et al [53], purposive sampling was used in this study to include both sexes, a range of ages and socioeconomic statuses, and those with and without technical problems during their video consultations. The study stated that although the demand for video consultation services in primary care is likely to rise, for complex or sensitive problems that require *touch*, face-to-face consultations remain preferable for patients and clinicians.

Commentaries

A commentary from 2 directors of free clinics in the United States described how their patients were unable to access their web-based system and offered the solution of a combination of technology and face-to-face services to help address some of these disparities [63]. Their discussion highlighted various upstream societal and social factors (such as mistrust of technology, internet availability regionally, and housing insecurity, to name a few) that were being exposed across hospital systems in the country at a critical time in a public health crisis with no measures in place to address them.

Mehmi et al [64] further discussed how they expected video consultations to increase health inequality if the correct infrastructure is not put in place. Thronson et al [65] considered the underlying cause of this "pandemic of health care inequity" [50] to be access—the disadvantaged simply cannot access telemedicine or home monitoring tools. Ramasawmy et al [66], by contrast, highlighted 3 areas in which existing knowledge and evidence can be translated into cross-sectoral action to avoid further ethnic and digital health inequalities: data and measurement, improved communication, and embedded equality impact.

Gray et al [67] offered 5 strategies to prevent the exacerbation of health disparities for low-income, rural, disabled, ethnic minority, and older adult populations in the United States. They considered sociocultural barriers to digital inclusion, including limited digital skills, low health literacy, disability, low income, and LEP, and structural barriers such as geographic isolation, broadband capacity, and technical hardware.

Crawford and Serhal [21] offered a new digital health equity framework to identify the digital determinants of health and their links to digital health equity, which requires additional evidence and empirical application. Gallegos-Rejas et al [69] proposed practical steps to reduce the digital divide and encourage equitable access to telehealth through improvements in digital health literacy, workforce training in clinical telehealth, co-design of new telehealth-enabled models of care, change management, advocacy for culturally appropriate services, and sustainable funding models.

Another 20% (2/10) of the commentaries, by Rodriguez et al [11] and Eruchalu et al [68], also explored strategies for the digital care of vulnerable patients during the pandemic and further discussed concerns about inequities in digital health access. They reiterated the finding that ethnic minority patients had significantly lower chances of attending telemedicine visits because of inequities in broadband access, lack of available technology, and mistrust of health care professionals.

All the authors of the aforementioned studies offered a list of proposed solutions to digital disparities. These are summarized and categorized in Textbox 3.

Although the changes proposed in Textbox 3 have some face validity, they remain largely untested.

A reviewer of a previous draft of this paper suggested that the taxonomy offered in Textbox 3 (which was our own way of making sense of the data we extracted from papers in our sample) reflected a particular theoretical perspective, namely, the social-ecological framework, which "considers the complex interplay between individual, relationship, community and societal factors" [71]. We agree that this lens could potentially provide an overarching framework within which to synthesize middle-range theories in a future paper.

Our second search identified 3 candidate theories that helped explain the effects of multiple disadvantage identified in our data set: fundamental cause theory, digital capital theory, and intersectionality theory, all of which are discussed in the next section.

The fundamental cause theory and intersectionality theory were the most common theories cited in general health disparity research; digital capital theory was mentioned in studies and

commentaries in digital disparity research. Various other theories recurred in the literature but were unhelpful in analyzing our data set. We found the aforementioned theories helpful as (1) intersectionality worked as an overall guiding principle for understanding how people's lives and characteristics stem from and lead to multiple axes of disadvantage; (2) digital capital theory helped us understand how these axes of disadvantage played out in terms of access to and use of digital resources; and (3) fundamental cause theory sensitized us to the pervasive impact of poverty, which operates through multiple intermediate mechanisms.

Discussion

Principal Findings

Findings from our narrative review of digital health disparities in relation to video consultations highlighted that the available literature reports substantial digital disparities. Formal research studies on this topic in the early months of the pandemic were few, mostly small, and rapidly conducted. Research in relation to video consultations to date has been almost entirely descriptive, and our data set included no in-depth, theoretically informed empirical studies that were able to explain how different dimensions of disadvantage combined to affect digital health disparities.

Our narrative review, which focused on video consultations, produced 3 key findings in particular. First, the literature was sparse, comprising only 7% (2/27) of reviews and 63% (17/27) of empirical studies, most of which were published since the start of the COVID-19 pandemic in 2020. Of these studies, most (25/27, 93%) were relatively small and undertaken quickly and under pressure during the pandemic, for example, qualitative studies that comprised one-shot semistructured interviews on convenience samples.

The second finding is that, despite the limitations of the literature, substantial digital disparities were reported [5,22,50-53,56,57,59]. Low household income, older age, and ethnic minority background (especially limited-English speakers) were all independently associated with lower uptake of video consultations [5,50,52,53,58,60,61]. Proposed explanations include lack of digital devices and infrastructure, low health and digital literacy, and inability to understand written resource materials [11,65,66,68]. The disparities found were sometimes dramatic and contrasted strikingly with studies of video consulting undertaken before the pandemic. These studies had framed this as an innovative service model that might increase service efficiency.

The third major finding was that research on digital health disparities, at least in relation to video consultations, has to date been almost entirely descriptive rather than explanatory. All the quantitative studies (audits and surveys) in our data set were designed to generate knowledge of the *association* between particular patient characteristics and the uptake and outcome of video consultations. Although such knowledge is essential in identifying a problem, there is a risk that such studies reduce the complex and interacting aspects of disadvantage to simple variables. Variable-focused research has been criticized by social scientists for oversimplifying context, removing key content (eg, unmeasured variables), overlooking historical path dependencies, and failing to explore how different variables combine and unfold over time to produce complex and sometimes unpredictable outcomes for individuals [72,73]. Explanations generated from such studies and from the superficial and qualitative designs included in our data set tended to couch findings in terms of *barriers* and *enablers*, which were depicted as having more or less fixed effects (negative and positive, respectively) on outcomes. The result is a body of literature that is desperately in need of theorization; for example, none of the 63% (17/27) of empirical studies cited any theory of disadvantage, and none of the well-intentioned ideas listed in Textbox 3 are couched in a well-developed theory of change.

Descriptive, variable-centered research is common when studying health disparities as it allows the so-called social determinants of health (eg, income, education, and gender) to be manipulated by quantitative techniques such as aggregation and correlation. However, such approaches are inherently problematic as they require research participants to be placed into categories that can then be manipulated as variables (*older adults*, *those with low income*, and *ethnic group X*), and findings tend to be presented in terms of what has been called "single-axis" analyses [25].

As Zheng and Walsham [74] have argued, these notions and categorizations do not consider the multifaceted and complex interplay of factors that contribute to digital disparities, nor how a characteristic that disadvantages an individual in one setting may have little adverse impact or even a positive impact in another. In terms of multiple disadvantage, for example, limited-English speakers are more likely to be older and lack basic digital devices and skills [75]. Although these factors may in some cases be mutually reinforcing, *some* older adults from *some* minority ethnic groups may be more likely than White British older adults to live in intergenerational households with good internet connection and family member support—hence, a characteristic (non-White ethnicity) that acts as a *barrier* in one setting may act as an *enabler* in another.

In sum, the current literature on digital health disparities is not only sparse but also in need of richer theorization to generate *explanations* of how different dimensions of disadvantage interact. We have argued elsewhere that the overemphasis in evidence-based medicine on empirical research at the expense of explanatory theory on the *causes* of phenomena (what some have called *EBM+*) may produce impoverished findings [76]. This builds on earlier work emphasizing the crucial importance of theory in selecting *which* hypotheses to test and how when studying disparities [77,78].

We have begun to explore the wider literature to identify theories of multiple disadvantage that have a potential bearing on digital health. In the following sections, we consider the 3 most relevant theories that emerged in our search to date: fundamental cause theory, digital capital theory, and intersectionality theory.

Fundamental Cause Theory

Link and Phelan [79] define a fundamental cause of health disparity as anything that involves resources that influence the extent to which people are able to avoid risks of mortality and morbidity. Socioeconomic status operates as a *fundamental cause* as it (1) involves access to resources (in particular, wealth, income, education, and racial privilege) that allow individuals to avoid diseases and their consequences and (2) affects multiple risk factors (eg, health literacy, quality of medical care, and diet) and disease outcomes that change over time. In short, those with financial resources and high social status can use these resources to avoid disease, seek treatment, and adopt healthy behaviors. The higher risk of heart disease in those of lower socioeconomic status, for example, can be explained by a combination of less access to money, health care, healthy food options, opportunities to exercise safely, and social support. These fundamental *upstream* disparities operate through multiple mediating factors at both the individual level (eg, diet, physical activity, and attendance at screening programs) and the metabolic level (eg, cholesterol, blood glucose, and stress hormones). An intervention that successfully changes one *risk factor* (such as BMI) will have limited impact as the fundamental cause will still operate through other mediating factors. This theory is often invoked when authors talk of the structural determinants of health disparities [80-82]. Although human behavior, lifestyle *choices*, knowledge, and beliefs may *mediate* the link between social determinants and adverse health outcomes, these factors are inadequate as explanatory *causes* of disease.

Fundamental cause theory offers a plausible explanation for the powerful and persistent link between multiple disadvantage and digital disparities. Applied to digital health disparities, it would depict the fundamental cause of these disparities as low socioeconomic status and that this cause operates through *flexible resources* such as access (or lack thereof) to money, knowledge, power, prestige, and beneficial social connections. A key hypothesis based on this theory is that addressing any one proximal cause—for example, by supplying a person with low income with a digital device—will not solve the fundamental problem as, although this *intervening mechanism* may change one *risk factor* (money in this case), it may have limited impact overall as the fundamental cause will continue to cause disparity through other mediating factors such as lack of knowledge (ie, not knowing how to use the digital device given).

Digital Capital Theory

Bourdieu [83] applied the idea of capital to signify the internal (eg, abilities and attitudes) and external (possessions and attributes) resources that people mobilize to achieve their goals in social life. He highlighted cultural capital as a form of capital that can be accumulated and transformed into other capitals. Digital capital is an extension by Ragnedda and Ruiu [84] of the theory of cultural capital by Bourdieu [83], made up of "both digital competencies and digital technologies." They argue that digital capital is a form of capital in its own right and is essential for building social, economic, and cultural resources in the digital world that we live in today. Disparities involving digital skills originate in inequalities of access but are mediated by orientations that can only be understood in relation to total life contexts (eg, education, income bracket, age, location, and social support all influence a person's access to digital technologies and the level of digital skills that they can acquire) [85]. Digital capital is a relatively new concept that scholars have begun to explore empirically using various methodological approaches [86,87]. Digital capital may be estimated, for example, at the individual level by assessing a person's digital literacy and skills, at the organizational level by measures of digital infrastructure (including the digital competence of personnel), and at the locality level in terms of the quality of the area's IT infrastructure.

Digital capital theory points us to the hypothesis that traditional forms of capital (such as economic, cultural, and social capital) are converted into digital capital and vice versa and provides the conceptual tools to examine how and to what extent this occurs, thereby illuminating how social inequality relates to digital inequality. If digital spaces—because of social inequality and underlying power structures—become increasingly stratified, there will be significant impacts on how individuals from differing backgrounds gain accumulated forms of capital through the digital realm. In other words, digital capital theory seems to offer an explanation as to why people who already experience health and other disparities find that these disparities widen when services are digitized.

Intersectionality Theory

The Black feminist scholar Kimberle Crenshaw [25] critiqued traditional studies of Black women's oppression for offering what she called "single-axis" analyses focusing on either race or gender but failing to integrate the 2 categories. Subsequent authors have extended the original concept of race-gender intersectionality by Crenshaw by adding categories, including nationality, class, age, sexual orientation, and disability [88,89], revealing "crosscutting and mutually reinforcing systems of domination and subordination" that "may construct multiple, uneven and contradictory social patterns" according to Anthias [90]. Intersectionality has been invoked to explain disparities in outcomes within minority ethnic groups in the context of the pandemic [91]. Intersectionality has been studied in many different ways [39]. Most relevant to our own data set is what we call *lived-experience* intersectionality research, which seeks to elucidate (through qualitative methods such as narrative interviews, ethnography, and arts-based methods) the complex and unique experiences of individuals whose identity crosses the boundaries of traditionally constructed groups [40,89].

Intersectionality theory applied to digital health disparities suggests the hypothesis that each individual's identity and lived experience is unique and multifaceted and that individuals will use (or will not use) digital services based on their own unique identity and circumstances—rather than as members of a single category such as *asylum seeker*, *Black individuals*, or *older adults*. This theoretical approach would support detailed small-scale case studies to see how different aspects of disadvantage interact in individual lives.

Strengths and Limitations

Although our narrative approach to this review allowed for a comprehensive overview of the wide range of literature spanning digital health disparities, there was no evaluation of selected articles for validity. Nonetheless, quality was not jeopardized as the methodology by Boell and Cecez-Kecmanovic [44] for hermeneutic review, an explicit methodology and accepted standard, was used alongside other methodological guidance for quality judgment. Narrative reviews are also often criticized for the subjective weighing of the studies chosen for the review. We sought to mitigate this through discussion between coauthors and the study team for an investigation focused on using remote care as part of a wider study [28]. To account for any additional selection bias on the authors' part, all the included studies were second reviewed for relevance.

A major limitation of the review was the sparse and undertheorized nature of the primary literature. We tried to remedy this by beginning to explore the wider literature to identify theories of multiple disadvantage that have a potential bearing on digital health. It should be noted that the 3 theories that we describe in the discussion are not a result of an exhaustive search, and there may be others of relevance. We plan to develop this stream of theoretical research in a future paper.

Conclusions: Suggested New Research Directions

Studies published since the COVID-19 pandemic began have shown that the move to digital forms of access and care provision has widened health disparities.

This recent literature contrasts strikingly with research on digital health services undertaken before the pandemic, which was largely focused on demonstrating noninferiority of digital modalities in terms of acceptability, safety, and subsequent use of services, often using randomized controlled trials with highly selected samples (stable, compliant, digitally equipped, and digitally confident patients recruited mostly from outpatient settings). The denominator population for digital health research expanded rapidly when such services became the default option for *everyone* for infection control reasons [92], revealing the (previously largely hidden) problem of wide digital disparities linked to multiple aspects of disadvantage. However, as this review has shown, in-pandemic research to date has been descriptive and superficial and has revealed few insights into how digital disparities emerge and play out both within and across categories of disadvantage.

This lack of attention to multiple disadvantage in digital health research to date represents both a unique opportunity and an important challenge for us to engage more curiously and theoretically with this core subject matter. We now need to turn our attention to specifically developing and using interdisciplinary theories of health disparities and technological innovation to inform our studies. Theories, including (but not limited to) fundamental cause theory, digital capital theory, and intersectionality theory, can provide a distinctive foundation for digital health inequality research and serve to guide ongoing research on this topic area. We suggest 3 complementary empirical approaches, as evidenced by previous studies that have conceptualized the aforementioned theories of health disparities.

First, quantitative studies using electronic patient record data should move beyond the current focus on single-axis analyses framed around a reified notion of the *digital divide* to produce category-focused intersectionality research. Such studies would need to be large, prospective, and hypothesis-driven to explore questions about the interaction between different categories of disadvantage; for example, how do gender, education, and ethnicity influence digital disparities in older people? The Digital Health Equity Framework [21] or the eHealth Equity Framework may provide a frame to think comprehensively about multifaceted approaches [93]. Using the eHealth Equity Framework approach in an initial scoping exercise, for example, can illuminate the proximal factors that need to be incorporated to address health inequities while also drawing attention to possible unintended consequences through distal interactions.

Second, qualitative studies of disadvantaged patients' digital access and experiences should move beyond the one-shot, theory-free semistructured interview on a convenience sample to achieve richly theorized, in-depth longitudinal studies of lived experience in maximum-diversity samples within particular categories of intersectionality. As noted previously, extended narrative interviews and ethnographic techniques should focus on the unique and particular experiences of individuals. Through rich description and the use of literary devices (eg, metaphor and dramatization), such lived-experience studies will reveal how multiple intersections play out over time in members of a particular broad category of intersectionality (eg, people who have low income, are older, and with limited English and multiple health needs) and also, importantly, illustrate the wide diversity of experiences within that group.

Third, based on the findings of this review as well as on the solutions to digital disparities proposed by the authors of primary studies and commentaries, we suggest adding a co-design component to research. Heard et al [40] proposed adapting lived-experience studies to inform the design of inclusive policies and interventions that can take account of the multiple social, cultural, and political contexts within which individual lives are lived and choices are contemplated. They cited the Intersectionality-Based Policy Analysis Framework as a tool to inform the design of such approaches [94]. The framework provides guidance and direction to address the challenges of health inequities across diverse populations in 3 ways. First, it provides an innovative structure for critical policy analysis. Second, it captures the different dimensions of policy contexts, including history, politics, everyday lived experiences, diverse knowledge, and intersecting social locations. Finally, it generates insights, knowledge, policy solutions, and actions that may not be fully captured through equity-focused policy frameworks. This systematic approach will help in designing policy responses that mitigate instead of increase the potential unequal effect of this phenomenon.

Health disparities are already wide and (in many countries) increasing. As society becomes increasingly digitized, the problem is likely to escalate if intersectional disparities are overlooked. This paper, which is intended as the starting point

for a wider debate, has outlined a novel and ambitious research agenda. We invite others to help address and extend it.

Acknowledgments

This review was undertaken to inform an empirical doctoral study on digital disadvantage by LH based on UK primary care and linked to the Remote by Default 2 study [24]. The PhD is supported by the Health Foundation grant to the University of Cambridge for The Healthcare Improvement Studies Institute. Ethics approval was not needed for the desk research component of the study reported in this paper. The study was overseen by the independent advisory group of the Remote by Default 2 study.

Conflicts of Interest

TG is a member of Independent Sage. All other authors declare they have no conflicts of interest.

References

1. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
2. Walker DM, Hefner JL, Fareed N, Huerta TR, McAlearney AS. Exploring the digital divide: age and race disparities in use of an inpatient portal. Telemed J E Health 2020 May;26(5):603-613 [FREE Full text] [doi: 10.1089/tmj.2019.0065] [Medline: 31313977]
3. Alexander GC, Tajanlangit M, Heyward J, Mansour O, Qato DM, Stafford RS. Use and content of primary care office-based vs telemedicine care visits during the COVID-19 pandemic in the US. JAMA Netw Open 2020 Oct 01;3(10):e2021476 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21476] [Medline: 33006622]
4. Nadkarni A, Hasler V, AhnAllen CG, Amonoo HL, Green DW, Levy-Carrick NC, et al. Telehealth during COVID-19-does everyone have equal access? Am J Psychiatry 2020 Nov 01;177(11):1093-1094. [doi: 10.1176/appi.ajp.2020.20060867] [Medline: 33135470]
5. Fu L, Lindenmeyer A, Phillimore J, Lessard-Phillips L. Vulnerable migrants' access to healthcare in the early stages of the COVID-19 pandemic in the UK. Public Health 2022 Feb;203:36-42 [FREE Full text] [doi: 10.1016/j.puhe.2021.12.008] [Medline: 35026578]
6. Kim HS, Kwon IH, Cha WC. Future and development direction of digital healthcare. Healthc Inform Res 2021 Apr;27(2):95-101 [FREE Full text] [doi: 10.4258/hir.2021.27.2.95] [Medline: 34015874]
7. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012 Jul 19;2(4):e001574 [FREE Full text] [doi: 10.1136/bmjopen-2012-001574] [Medline: 22815469]
8. Salisbury C, Quigley A, Hex N, Aznar C. Private video consultation services and the future of primary care. J Med Internet Res 2020 Oct 01;22(10):e19415 [FREE Full text] [doi: 10.2196/19415] [Medline: 32812887]
9. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Sep;69(686):e595-e604 [FREE Full text] [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
10. Huxley CJ, Atherton H, Watkins JA, Griffiths F. Digital communication between clinician and patient and the impact on marginalised groups: a realist review in general practice. Br J Gen Pract 2015 Dec;65(641):e813-e821 [FREE Full text] [doi: 10.3399/bjgp15X687853] [Medline: 26622034]
11. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st century cures act era. JAMA 2020 Jun 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
12. Shaw J, Brewer LC, Veinot T. Recommendations for health equity and virtual care arising from the COVID-19 pandemic: narrative review. JMIR Form Res 2021 Apr 05;5(4):e23233 [FREE Full text] [doi: 10.2196/23233] [Medline: 33739931]
13. Ibrahim FA, Malathesh BC, Gajera G, Pandey P, Nirisha PL, Jyrwa S, et al. Chhattisgarh community mental healthcare tele-mentoring program (CHaMP): digitally driven initiative to reach the unreached. Int J Soc Psychiatry 2022 Aug;68(5):954-957. [doi: 10.1177/00207640211011191] [Medline: 33860714]
14. Anderson M, Perrin A. Tech Adoption Climbs Among Older Adults. Pew Research Center. 2017 May 17. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2022-05-10]
15. Anaya YB, Hernandez GD, Hernandez SA, Hayes-Bautista DE. Meeting them where they are on the web: addressing structural barriers for Latinos in telehealth care. J Am Med Inform Assoc 2021 Sep 18;28(10):2301-2305 [FREE Full text] [doi: 10.1093/jamia/ocab155] [Medline: 34313774]
16. Shah DA, Sall D, Peng W, Sharer R, Essary AC, Radhakrishnan P. Exploring the role of telehealth in providing equitable healthcare to the vulnerable patient population during COVID-19. J Telemed Telecare (forthcoming) 2022 Jul 14:1357633X221113711 [FREE Full text] [doi: 10.1177/1357633X221113711] [Medline: 35833345]
17. Sutherland AE, Stickland J, Wee B. Can video consultations replace face-to-face interviews? Palliative medicine and the Covid-19 pandemic: rapid review. BMJ Support Palliat Care 2020 Sep;10(3):271-275 [FREE Full text] [doi: 10.1136/bmjspcare-2020-002326] [Medline: 32457086]

18. Wherton J, Shaw S, Papoutsi C, Seuren L, Greenhalgh T. Guidance on the introduction and use of video consultations during COVID-19: important lessons from qualitative research. BMJ Leader 2020 May 18;4(3):120-123. [doi: 10.1136/leader-2020-000262]
19. Schou J, Pors AS. Digital by default? A qualitative study of exclusion in digitalised welfare. Soc Policy Admin 2018 Nov 22;53(3):464-477. [doi: 10.1111/spol.12470]
20. Roberts ET, Mehrotra A. Assessment of disparities in digital access among Medicare beneficiaries and implications for telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
21. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
22. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst 2020 May 4;1(3) [FREE Full text]
23. Eberly LA, Khatana SA, Nathan AS, Snider C, Julien HM, Deleener ME, et al. Telemedicine outpatient cardiovascular care during the COVID-19 pandemic: bridging or opening the digital divide? Circulation 2020 Aug 04;142(5):510-512 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.120.048185] [Medline: 32510987]
24. Greenhalgh T, Shaw SE, Alvarez Nishio A, Booth A, Byng R, Clarke A, et al. Protocol: remote care as the 'new normal'? Multi-site case study in UK general practice. NIHR Open Res 2022 Aug 8;2:46. [doi: 10.3310/nihropenres.13289.1]
25. Crenshaw K. Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ Chicago Legal Forum 1989;1989(1):8 [FREE Full text] [doi: 10.4324/9780429500480-5]
26. Vimalananda VG, Gupte G, Seraj SM, Orlander J, Berlowitz D, Fincke BG, et al. Electronic consultations (e-consults) to improve access to specialty care: a systematic review and narrative synthesis. J Telemed Telecare 2015 Sep;21(6):323-330 [FREE Full text] [doi: 10.1177/1357633X15582108] [Medline: 25995331]
27. DeGuzman PB, Jain N, Loureiro CG. Public libraries as partners in telemedicine delivery: a review and research agenda. Public Library Q 2022;41(3):294-304. [doi: 10.1080/01616846.2021.1877080]
28. Greenhalgh T, Rosen R. Remote by default general practice: must we, should we, dare we? Br J Gen Pract 2021 Apr;71(705):149-150 [FREE Full text] [doi: 10.3399/bjgp21X715313] [Medline: 33771790]
29. Carter-Pokras O, Baquet C. What is a "health disparity"? Public Health Rep 2002;117(5):426-434 [FREE Full text] [doi: 10.1093/phr/117.5.426] [Medline: 12500958]
30. Amri MM, Siddiqi A, O'Campo P, Enright T, Di Ruggiero E. Underlying equity discourses of the World Health Organization. Soc Sci Protocols 2020 Jun 21;3:1-6. [doi: 10.7565/ssp.2020.2812]
31. Goldman DP, Lakdawalla DN. A theory of health disparities and medical technology. Contrib Econ Anal Policy 2005 Sep 16;4(1):1-30. [doi: 10.2202/1538-0645.1395]
32. McCartney G, Popham F, McMaster R, Cumbers A. Defining health and health inequalities. Public Health 2019 Jul;172:22-30 [FREE Full text] [doi: 10.1016/j.puhe.2019.03.023] [Medline: 31154234]
33. Global strategy on digital health 2020-2025. World Health Organization. 2021 Aug 18. URL: http://apps.who.int/bookorders [accessed 2022-06-25]
34. Ryu S. Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). Healthc Inform Res 2012 Jun;18(2):153-155. [doi: 10.4258/hir.2012.18.2.153]
35. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014 Dec;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]
36. Davies AR, Honeyman M, Gann B. Addressing the digital inverse care law in the time of COVID-19: potential for digital technology to exacerbate or mitigate health inequalities. J Med Internet Res 2021 Apr 07;23(4):e21726 [FREE Full text] [doi: 10.2196/21726] [Medline: 33735096]
37. Stone E, Nuckley P, Shapiro R. Digital Inclusion in Health and Care: Lessons learned from the NHS Widening Digital Participation Programme (2017-2020). Good Things Foundation. 2020. URL: https://tinyurl.com/3uc37nek [accessed 2022-07-16]
38. Lazarus JV, Cascio M, Halford R, Onyango D, Schatz E, Smith A, et al. Nobody Left Outside (NLO) Checklist: improving access to healthcare for vulnerable and underserved groups. Int J Integr Care 2019 Aug 08;19(4):507. [doi: 10.5334/ijic.s3507]
39. McCall L. The complexity of intersectionality. Signs 2005 Mar;30(3):1771-1800. [doi: 10.1086/426800]
40. Heard E, Fitzgerald L, Wigginton B, Mutch A. Applying intersectionality theory in health promotion research and practice. Health Promot Int 2020 Aug 01;35(4):866-876. [doi: 10.1093/heapro/daz080] [Medline: 31390472]
41. Wherton J, Greenhalgh T, Shaw SE. Expanding video consultation services at pace and scale in Scotland during the COVID-19 pandemic: national mixed methods case study. J Med Internet Res 2021 Oct 07;23(10):e31374 [FREE Full text] [doi: 10.2196/31374] [Medline: 34516389]
42. Jones R. Inequalities in general practice website provision. Br J Gen Pract 2011 Nov;61(592):663 [FREE Full text] [doi: 10.3399/bjgp11x606555] [Medline: 22167899]

43. Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? Eur J Clin Invest 2018 Jun;48(6):e12931 [FREE Full text] [doi: 10.1111/eci.12931] [Medline: 29578574]
44. Boell SK, Cecez-Kecmanovic D. A hermeneutic approach for conducting literature reviews and literature searches. Commun Assoc Inf Syst 2014;34:12. [doi: 10.17705/1cais.03412]
45. Baethge C, Goldbeck-Wood S, Mertens S. SANRA-a scale for the quality assessment of narrative review articles. Res Integr Peer Rev 2019 Mar 26;4:5 [FREE Full text] [doi: 10.1186/s41073-019-0064-8] [Medline: 30962953]
46. Lewin S, Bohren M, Rashidian A, Munthe-Kaas H, Glenton C, Colvin CJ, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 2: how to make an overall CERQual assessment of confidence and create a Summary of Qualitative Findings table. Implement Sci 2018 Jan 25;13(Suppl 1):10 [FREE Full text] [doi: 10.1186/s13012-017-0689-2] [Medline: 29384082]
47. 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]
48. Parker RF, Figures EL, Paddison CA, Matheson JI, Blane DN, Ford JA. Inequalities in general practice remote consultations: a systematic review. BJGP Open 2021 Jun;5(3):BJGPO.2021.0040 [FREE Full text] [doi: 10.3399/BJGPO.2021.0040] [Medline: 33712502]
49. Litchfield I, Shukla D, Greenfield S. Impact of COVID-19 on the digital divide: a rapid review. BMJ Open 2021 Oct 12;11(10):e053440 [FREE Full text] [doi: 10.1136/bmjopen-2021-053440] [Medline: 34642200]
50. Chang JE, Lai AY, Gupta A, Nguyen AM, Berry CA, Shelley DR. Rapid transition to telehealth and the digital divide: implications for primary care access and equity in a post-COVID era. Milbank Q 2021 Jun;99(2):340-368 [FREE Full text] [doi: 10.1111/1468-0009.12509] [Medline: 34075622]
51. Stevens AJ, Ray AM, Thirunavukarasu A, Johnson E, Jones L, Miller A, et al. The experiences of socially vulnerable groups in England during the COVID-19 pandemic: a rapid health needs assessment. Public Health Pract (Oxf) 2021 Nov;2:100192 [FREE Full text] [doi: 10.1016/j.puhip.2021.100192] [Medline: 34608460]
52. Kaihlanen AM, Virtanen L, Buchert U, Safarov N, Valkonen P, Hietapakka L, et al. Towards digital health equity - a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188 [FREE Full text] [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
53. Knights F, Carter J, Deal A, Crawshaw AF, Hayward SE, Jones L, et al. Impact of COVID-19 on migrants' access to primary care and implications for vaccine roll-out: a national qualitative study. Br J Gen Pract 2021 Jul 29;71(709):e583-e595 [FREE Full text] [doi: 10.3399/BJGP.2021.0028] [Medline: 33875420]
54. Alkureishi MA, Choo ZY, Rahman A, Ho K, Benning-Shorb J, Lenti G, et al. Digitally disconnected: qualitative study of patient perspectives on the digital divide and potential solutions. JMIR Hum Factors 2021 Dec 15;8(4):e33364 [FREE Full text] [doi: 10.2196/33364] [Medline: 34705664]
55. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
56. 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;27(1):21-26 [FREE Full text] [doi: 10.37765/ajmc.2021.88573] [Medline: 33471458]
57. Hsueh L, Huang J, Millman AK, Gopalan A, Parikh RK, Teran S, et al. Disparities in use of video telemedicine among patients with limited English proficiency during the COVID-19 pandemic. JAMA Netw Open 2021 Nov 01;4(11):e2133129 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.33129] [Medline: 34735015]
58. Yu E, Hagens S. Socioeconomic disparities in the demand for and use of virtual visits among senior adults during the COVID-19 pandemic: cross-sectional study. JMIR Aging 2022 Mar 22;5(1):e35221 [FREE Full text] [doi: 10.2196/35221] [Medline: 35134746]
59. Schifeling CH, Shanbhag P, Johnson A, Atwater RC, Koljack C, Parnes BL, et al. Disparities in video and telephone visits among older adults during the COVID-19 pandemic: cross-sectional analysis. JMIR Aging 2020 Nov 10;3(2):e23176 [FREE Full text] [doi: 10.2196/23176] [Medline: 33048821]
60. Sachs JW, Graven P, Gold JA, Kassakian SZ. Disparities in telephone and video telehealth engagement during the COVID-19 pandemic. JAMIA Open 2021 Jul;4(3):ooab056 [FREE Full text] [doi: 10.1093/jamiaopen/ooab056] [Medline: 34632322]
61. Broffman L, Harrison S, Zhao M, Goldman A, Patnaik I, Zhou M. The relationship between broadband speeds, device type, demographic characteristics, and care-seeking via telehealth. Telemed J E Health (forthcoming) 2022 Jul 22. [doi: 10.1089/tmj.2022.0058] [Medline: 35867048]
62. Zachrison KS, Yan Z, Sequist T, Licurse A, Tan-McGrory A, Erskine A, et al. Patient characteristics associated with the successful transition to virtual care: lessons learned from the first million patients. J Telemed Telecare (forthcoming) 2021 Jun 13:1357633X211015547. [doi: 10.1177/1357633X211015547] [Medline: 34120506]
63. 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]

64. Mehmi AK, Winters D, Green JS. Video consultations will increase health inequality without the correct infrastructure put in place. BMJ 2020 Nov 19;371:m4498. [doi: 10.1136/bmj.m4498] [Medline: 33214144]
65. Thronson LR, Jackson SL, Chew LD. The pandemic of health care inequity. JAMA Netw Open 2020 Oct 01;3(10):e2021767 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21767] [Medline: 33006616]
66. Ramasawmy M, Poole L, Banerjee A. Learning our lesson: using past policies to improve digital and ethnic inequalities beyond the pandemic. Arch Public Health 2021 Dec 01;79(1):218 [FREE Full text] [doi: 10.1186/s13690-021-00744-8] [Medline: 34847923]
67. Gray 2nd DM, Joseph JJ, Olayiwola JN. Strategies for digital care of vulnerable patients in a COVID-19 world-keeping in touch. JAMA Health Forum 2020 Jun 01;1(6):e200734 [FREE Full text] [doi: 10.1001/jamahealthforum.2020.0734] [Medline: 36218529]
68. Eruchalu CN, Pichardo MS, Bharadwaj M, Rodriguez CB, Rodriguez JA, Bergmark RW, 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]
69. Gallegos-Rejas VM, Thomas EE, Kelly JT, Smith AC. A multi-stakeholder approach is needed to reduce the digital divide and encourage equitable access to telehealth. J Telemed Telecare (forthcoming) 2022 Jun 22:1357633X221107995. [doi: 10.1177/1357633X221107995] [Medline: 35733379]
70. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SA, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
71. The Social-Ecological Model: A Framework for Prevention - Injury Prevention. Centers for Disease Control and Prevention. 2020. URL: https://tinyurl.com/577h55k2 [accessed 2022-11-15]
72. Ramiller NC, Pentland BT. Management implications in information systems research: the untold story. J Assoc Inf Syst 2009 Jun 18;10(6):474-494. [doi: 10.17705/1jais.00202]
73. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010 Nov 02;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
74. Zheng Y, Walsham G. Inequality of what? An intersectional approach to digital inequality under Covid-19. Inf Organ 2021 Mar;31(1):100341. [doi: 10.1016/j.infoandorg.2021.100341]
75. Elahi F. Digital Inclusion: Bridging Divides. Cumberland Lodge. 2020. URL: https://www.cumberlandlodge.ac.uk/sites/default/files/cumberland_lodge_digital_inclusion_-_bridging_divides_august_2020_for_web_0.pdf [accessed 2022-05-18]
76. Greenhalgh T, Fisman D, Cane DJ, Oliver M, Macintyre CR. Adapt or die: how the pandemic made the shift from EBM to EBM+ more urgent. BMJ Evid Based Med (forthcoming) 2022 Jul 19:253-260 [FREE Full text] [doi: 10.1136/bmjebm-2022-111952] [Medline: 35853682]
77. Krieger N. Methods for the scientific study of discrimination and health: an ecosocial approach. Am J Public Health 2012 May;102(5):936-944. [doi: 10.2105/AJPH.2011.300544] [Medline: 22420803]
78. Lynch EA, Mudge A, Knowles S, Kitson AL, Hunter SC, Harvey G. "There is nothing so practical as a good theory": a pragmatic guide for selecting theoretical approaches for implementation projects. BMC Health Serv Res 2018 Nov 14;18(1):857 [FREE Full text] [doi: 10.1186/s12913-018-3671-z] [Medline: 30428882]
79. Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav 1995;35:80-94. [doi: 10.2307/2626958]
80. Marmot M, Bell R. Fair society, healthy lives. Public Health 2012 Sep;126 Suppl 1:S4-10. [doi: 10.1016/j.puhe.2012.05.014] [Medline: 22784581]
81. Equally Well: Report of the Ministerial Task Force on Health Inequalities. Scottish Government. 2008 Jun 19. URL: https://www.gov.scot/publications/equally-well-report-ministerial-task-force-health-inequalities/ [accessed 2022-05-18]
82. McCartney G, Collins C, Mackenzie M. What (or who) causes health inequalities: theories, evidence and implications? Health Policy 2013 Dec;113(3):221-227. [doi: 10.1016/j.healthpol.2013.05.021] [Medline: 23810172]
83. Bourdieu P. The forms of capital. In: Granovetter M, Swedberg R, editors. The Sociology of Economic Life. 3rd edition. Oxfordshire, UK: Routledge; 2011.
84. Ragnedda M, Ruiu ML. Digital Capital: A Bourdieusian Perspective on the Digital Divide. Bingley, UK: Emerald Group Publishing; 2020.
85. Robinson L. A taste for the necessary: a Bourdieuian approach to digital inequality. Inf Commun Soc 2009 Jun;12(4):488-507. [doi: 10.1080/13691180902857678]
86. Ragnedda M, Ruiu ML, Addeo F. Measuring Digital Capital: an empirical investigation. New Media Soc 2019 Aug 27;22(5):793-816. [doi: 10.1177/1461444819869604]
87. Bannykh G, Kostina S. Measuring digital capital: methodological approaches. KnE Soc Sci 2022 Feb 04:96-103. [doi: 10.18502/kss.v7i2.10282]
88. Gkiouleka A, Huijts T, Beckfield J, Bambra C. Understanding the micro and macro politics of health: inequalities, intersectionality and institutions - a research agenda. Soc Sci Med 2018 Mar;200:92-98 [FREE Full text] [doi: 10.1016/j.socscimed.2018.01.025] [Medline: 29421476]

89. Bowleg L. The problem with the phrase women and minorities: intersectionality-an important theoretical framework for public health. Am J Public Health 2012 Jul;102(7):1267-1273. [doi: 10.2105/AJPH.2012.300750] [Medline: 22594719]
90. Anthias F. Social stratification and social inequality: models of intersectionality and identity. In: Devine F, Savage M, Scott J, Crompton R, editors. Rethinking Class: Cultures, Identities and Lifestyles. London, UK: Bloomsbury Publishing; Jan 2005:24-45.
91. Berkhout SG, Richardson L. Identity, politics, and the pandemic: why is COVID-19 a disaster for feminism(s)? Hist Philos Life Sci 2020 Oct 13;42(4):49 [FREE Full text] [doi: 10.1007/s40656-020-00346-7] [Medline: 33048278]
92. Bidmead E, Marshall A. Covid-19 and the 'new normal': are remote video consultations here to stay? Br Med Bull 2020 Oct 14;135(1):16-22 [FREE Full text] [doi: 10.1093/bmb/ldaa025] [Medline: 32827250]
93. Antonio MG, Petrovskaya O. Towards developing an eHealth equity conceptual framework. Stud Health Technol Inform 2019;257:24-30. [Medline: 30741167]
94. Hankivsky O, Grace D, Hunting G, Giesbrecht M, Fridkin A, Rudrum S, et al. An intersectionality-based policy analysis framework: critical reflections on a methodology for advancing equity. Int J Equity Health 2014 Dec 10;13:119 [FREE Full text] [doi: 10.1186/s12939-014-0119-x] [Medline: 25492385]

Abbreviations

LEP: limited English proficiency
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SANRA: Scale for the Assessment of Narrative Review Articles

Edited by T Leung; submitted 10.09.22; peer-reviewed by M Ramasawmy, J Shaw, E Khoong, H Veldandi, R Kakumanu; comments to author 30.09.22; revised version received 20.10.22; accepted 16.11.22; published 07.12.22.

Please cite as:
Husain L, Greenhalgh T, Hughes G, Finlay T, Wherton J
Desperately Seeking Intersectionality in Digital Health Disparity Research: Narrative Review to Inform a Richer Theorization of Multiple Disadvantage
J Med Internet Res 2022;24(12):e42358
URL: https://www.jmir.org/2022/12/e42358
doi:10.2196/42358
PMID:36383632

Review

Efficacy of Internet-Based Acceptance and Commitment Therapy for Depressive Symptoms, Anxiety, Stress, Psychological Distress, and Quality of Life: Systematic Review and Meta-analysis

Areum Han[1*], PhD; Tae Hui Kim[2,3*], MD

[1]Department of Occupational Therapy, University of Alabama at Birmingham, Birmingham, AL, United States

[2]Department of Psychiatry, Yonsei Wonju Medical College, Wonju, Republic of Korea

[3]Department of Psychiatry, Wonju Severance Christian Hospital, Wonju, Republic of Korea

[*]all authors contributed equally

Corresponding Author:
Tae Hui Kim, MD
Department of Psychiatry
Yonsei Wonju Medical College
Ilsan-ro 20
Wonju, 26426
Republic of Korea
Phone: 82 337410534
Email: gooddr@yonsei.ac.kr

Abstract

Background: Acceptance and commitment therapy (ACT) is an empirically supported transdiagnostic approach that involves mindfulness processes and behavior change processes for valued living.

Objective: This systematic review and meta-analysis of randomized controlled trials (RCTs) aimed to assess the efficacy of internet-based ACT (iACT) for depressive symptoms, anxiety, stress, psychological distress, and quality of life (QoL).

Methods: PubMed, CINAHL, PsycINFO, and SCOPUS databases were searched to identify relevant RCTs published up to June 5, 2021. The included RCTs were assessed using the Cochrane Collaboration risk-of-bias tool. The use of either a random effects model or fixed effects model was determined using I2 statistic values for heterogeneity. Subgroup analyses were conducted according to the type of control group, the use of therapist guidance, delivery modes, and the use of targeted participants, when applicable.

Results: A total of 39 RCTs met the inclusion criteria. Meta-analyses found small effects of iACT on depressive symptoms, anxiety, stress, psychological distress, and QoL at the immediate posttest and follow-up. There was no significant effect of iACT on stress at follow-up. Subgroup analyses showed small to medium effects of iACT on all the outcomes at the immediate posttest and follow-up compared with the passive control groups. In contrast, subgroup analyses that compared iACT with active control groups found no differences between groups on stress, psychological distress, and QoL at the immediate posttest or on depressive symptoms, anxiety, and stress at follow-up. In addition, subgroup analyses conducted according to the use of therapist guidance, delivery modes, and the use of targeted participants found no statistically significant subgroup differences among studies in all the outcomes, except for the subgroup difference among studies according to the use of targeted participants for depressive symptoms at the immediate posttest (ie, a statistically significant, larger effect of iACT when studies targeted people with depressive symptoms). The overall risk of bias across the studies was unclear.

Conclusions: The findings of this study contribute to the body of evidence regarding the effects of iACT on depressive symptoms, anxiety, stress, psychological distress, and QoL and may be applicable in any population, as ACT is a transdiagnostic approach. Few studies have compared iACT with active control conditions, especially for stress and psychological distress at the immediate posttest and follow-up. In addition, the active control conditions varied among the included studies. Further high-quality studies are needed to better understand whether iACT is comparable or superior to other evidence-based interventions, such as cognitive behavioral therapy, in decreasing depressive symptoms, anxiety, stress, and psychological distress and improving QoL.

(J Med Internet Res 2022;24(12):e39727) doi:10.2196/39727

KEYWORDS

acceptance and commitment therapy; anxiety; depression; internet-based intervention; meta-analysis; psychological distress; quality of life; stress; systematic review

Introduction

Background

Acceptance and commitment therapy (ACT) is an empirically supported transdiagnostic approach that involves mindfulness and acceptance processes and behavior change processes for valued living [1,2]. ACT aims to develop greater psychological flexibility, or the ability to face challenging experiences in an open manner and change one's behaviors to participate in valued activities, rather than avoiding uncomfortable or painful experiences, emotions, and thoughts [1,2]. ACT is based on a psychological flexibility model involving six processes [2]: (1) acceptance (ie, being open to unwanted thoughts and emotions as they are), (2) cognitive defusion (ie, stepping back from unhelpful thoughts and emotions to reduce their dominance over behaviors), (3) being present (ie, maintaining contact with the present moment), (4) observing self (ie, flexible self-conceptualization and perspective taking), (5) values (ie, clarifying personal values), and (6) committed action (ie, establishing patterns of behaviors for valued living) [2].

A growing body of evidence shows that ACT can reduce depressive symptoms, anxiety, stress, and psychological distress and improve quality of life (QoL) in various populations [3,4]. For example, previous meta-analyses found that ACT had small to medium effects on reducing depressive symptoms and anxiety and improving QoL in family caregivers [4] and a medium effect on reducing depressive symptoms in people diagnosed with depression [3]. A majority of previous meta-analyses regarding ACT were limited to specific populations, such as people with chronic pain [5] and people with psychosis [6], which led to a small number of included studies for meta-analysis. In addition, subgroup analyses were not conducted according to the delivery method (eg, face-to-face ACT vs internet-based ACT [iACT]) in the previous meta-analyses because of the limited number of studies included for ACT [4,7].

Internet-based psychological interventions are easy to access and inexpensive; therefore, it is important to determine whether iACT is an effective alternative option for reducing depressive symptoms, anxiety, stress, and overall psychological distress and improving QoL [8]. Brown et al [9] conducted a meta-analysis to measure the effects of iACT on outcomes related to mental health and well-being in any population and found a small effect of iACT on depressive symptoms only. This could be because of the limited number of included studies for each outcome at that time, which included 10 studies for depressive symptoms, 7 studies for anxiety, and 8 studies for QoL. More recently, Thompson et al [10] conducted a meta-analysis to measure the effects of iACT on depression, anxiety, and QoL and found small effects of iACT on all these outcomes. However, stress and psychological distress were not included in the meta-analysis by Thompson et al [10]. It is also possible that more studies have been published since the search by Thompson et al [10], which was conducted in June 2019. More importantly, subgroup analyses for each outcome were not conducted according to the type of control group in any of the previous meta-analyses to determine whether the effects of iACT differed compared with active control groups provided with other comparable interventions and passive control groups provided with no intervention. In addition, other subgroup analyses (eg, subgroup analyses according to the use of therapist guidance in iACT) may be possible and may provide useful information.

Objectives

This systematic review and meta-analysis aimed to assess the efficacy of iACT for depressive symptoms, anxiety, stress, overall psychological distress, and QoL in any population, with subgroup analyses according to the type of control group and other possible subgroup analyses depending on the characteristics of the included studies, when applicable.

Methods

Overview

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline [11] and the *Cochrane Handbook for Systematic Reviews of Interventions* (version 5.1.0) [12]. However, this study was not preregistered.

Inclusion and Exclusion Criteria

Studies were selected based on the following inclusion criteria: (1) the study must be a randomized controlled trial (RCT); (2) ACT must be mainly delivered on the web (ie, iACT); (3) the study must have pre- and posttest results in measures of depressive symptoms, anxiety, stress, psychological distress, or QoL; (4) the study must compare iACT with a non-ACT comparison or a control condition; and (5) the study must be written in English. Studies were excluded if they compared only different delivery modes among the ACT groups (eg, ACT delivered on the web vs ACT delivered in person) without any other comparison or a control condition.

Search Strategy

A comprehensive search was conducted using 4 electronic databases from the date of inception of each database to June 5, 2021: PubMed (1966-2021), CINAHL (1981-2021), PsycINFO (1935-2021), and SCOPUS (1966-2021). Key search terms were combined using keywords for iACT to identify the relevant literature. To broaden the database search, keywords for the outcomes were not entered as search terms. The search terms used in PubMed were as follows: ("acceptance and commitment therapy"[tiab] OR "Acceptance and Commitment Therapy"[MeSH]) AND (online[tiab] OR e-health[tiab] OR Internet*[tiab] OR web[tiab] OR webs[tiab] OR "web-based"[tiab] OR "web-delivered"[tiab] OR computer*[tiab] OR app[tiab] OR apps[tiab] OR mobile[tiab] OR technolog*[tiab] OR "Computers"[Mesh] OR

"Internet-Based Intervention"[MeSH] OR "Telemedicine"[MeSH] OR "Distance Counseling"[MeSH] OR "Mobile Applications"[MeSH]). The full search strategies for all databases can be found in Multimedia Appendix 1. Articles were also manually searched using the reference lists of the identified articles and the related article features in the databases.

Data Extraction and Quality Assessment

The characteristics of the included RCTs (eg, sample size, characteristics of participants, brief description of intervention and control groups, outcome measures, and results) were extracted into a table. Data regarding the means, SD, and sample sizes of each group were entered into an Excel (Microsoft) file. The risk of bias in the included RCTs was assessed using the Cochrane Collaboration risk-of-bias tool [12]. The domains in the tool include random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective reporting. The risk of bias in each of the domains was judged as *low risk* of bias, *high risk* of bias, or *unclear risk* of bias according to the criteria provided in the Cochrane Collaboration handbook [12]. Summary assessments of the risk of bias within a study and across studies were also determined based on the handbook's criteria [12]. One author completed the process for data extraction and quality assessment.

Meta-analysis

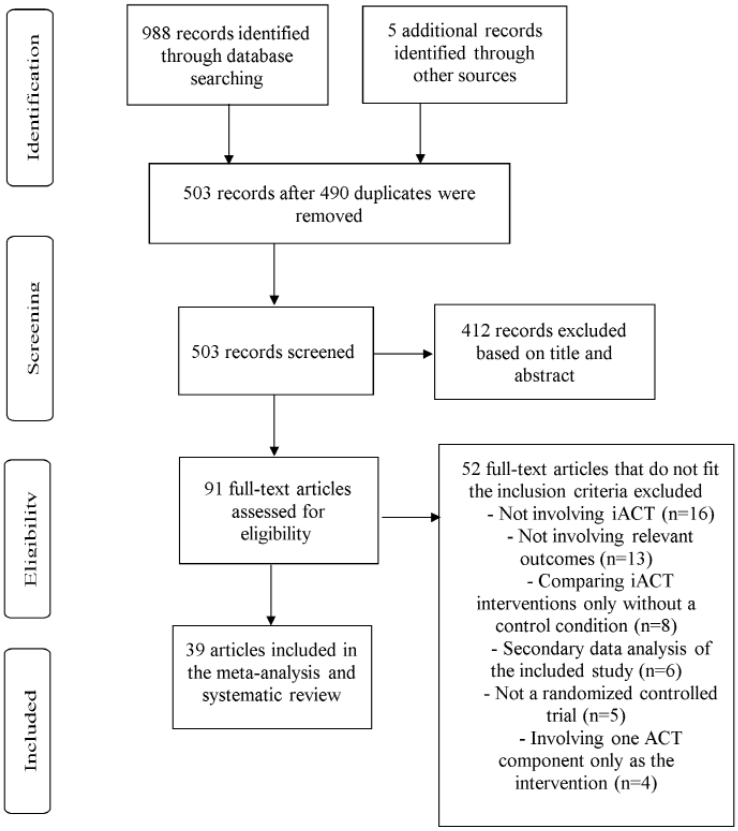
Means, SDs, and sample sizes of the intervention and control groups in the included studies were entered into RevMan (version 5.4; Cochrane Collaboration) for meta-analysis and pooled for each outcome at the immediate posttest and at follow-up. The I^2 statistic was used to measure the statistical heterogeneity across studies, and an I^2>60% was interpreted as substantial heterogeneity [12]. The decision to use either a random effects model or a fixed effects model with the inverse variance method was determined by the I^2 statistic values for each outcome, that is, a random effects model was used when the I^2 statistic for each variable was >60%; otherwise, a fixed effects model was used. A significance level (*P* value) of .05 was used. The standardized mean difference (SMD) with 95% CIs was used as a summary statistic for the size of the intervention effect to account for outcomes measured using different assessment tools [12]. SMDs <0.4 indicate a small effect, SMDs between 0.4 and 0.7 indicate a medium effect, and SMDs >0.7 indicate a large effect [12]. Subgroup analyses for each outcome were conducted according to the type of control group, if applicable, to see whether the effects of iACT differed compared with active control groups provided with other comparable interventions and passive control groups provided with no intervention (ie, treatment-as-usual control groups and waitlist control groups).

Results

Selection of Studies

Figure 1 illustrates the study selection process. A total of 988 articles were identified through database searching, and 5 additional articles were identified through manual searching. After removing 490 duplicates, 503 articles were screened based on title and abstract screening. A total of 412 articles were excluded based on title and abstract screening, and 91 articles were assessed for eligibility by reading the full text. A total of 52 articles were excluded after reading the full text because of the following reasons: not involving iACT (16 studies), not involving relevant outcomes (13 studies), comparing different delivery methods of iACT interventions without a control condition (8 studies), secondary data analysis of the included study (6 studies), not an RCT (5 studies), and involving 1 ACT component only (4 studies). A total of 39 articles met the eligibility criteria [13-51].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection process. ACT: acceptance and commitment therapy; iACT: internet-based acceptance and commitment therapy.



Characteristics of the Included Studies

The main characteristics of the included 39 RCTs are summarized in Multimedia Appendix 2 [13-51]. The average number of ACT modules (sessions) in the included studies was 6.7 (SD 2.5), ranging from 2 modules to 13 modules. ACT was delivered on the web with therapist guidance in 32 studies (eg, via videoconferencing, phone calls, written feedback, and a mobile app) and without therapist guidance in 7 studies [13-19]. A total of 8 studies used a blended ACT program involving both iACT and in-person ACT sessions [20-27]. In addition, 2 studies involved videoconferencing ACT [28,29]. The remaining 29 RCTs involved web-based ACT modules. Of the 39 RCTs, 10 studies involved active control groups, including web-based discussion forums [35,37,41], web-based expressive writing [32,51], web-based mental health education [17], a web-based smoking cessation intervention [14], web-based cognitive behavioral therapy (CBT) [13], in-person CBT [27], and in-person documentary discussions [25]. Moreover, 18 studies directly targeted people with depressive symptoms [14,21,23-25,27,30-40]. Additional subgroup analyses were conducted because studies could be categorized according to the following 3 characteristics: use of therapist guidance, delivery modes (ie, web-based ACT modules, iACT accompanied by in-person ACT sessions, and videoconferencing ACT), and use of targeted participants (eg, studies that directly targeted participants with depressive symptoms vs studies that involved participants regardless of the depressive symptoms used for subgroup analysis of depressive symptoms).

The average sample size of participants in the included RCTs was 139 (SD 185), ranging from 24 to 1162. The mean age of the participants was 40.1 (SD 13.3) years, ranging from 15.3 to 63.1 years, and the average percentage of female participants was 69.14% (3791/5414; SD 19.5%), ranging from 0% to 98.5%. The included RCTs were conducted in Sweden (10 studies), the United States (8 studies), the Netherlands (6 studies), Finland (5 studies), the United Kingdom (2 studies), Australia (2 studies), Canada (1 study), Belgium (1 study), Ireland (1 study), France (1 study), Denmark (1 study), and Germany (1 study). Of the 39 included studies, 30 (77%) were published between 2016 and 2021, and the remaining 9 (23%) were published between 2012 and 2015.

The following section describes the results of the meta-analyses regarding the efficacy of iACT for depressive symptoms, anxiety, stress, psychological distress, and QoL at the immediate posttest and follow-up. Subgroup analyses for each outcome were performed according to the type of control groups (ie, iACT vs active control groups and iACT vs passive control groups) when applicable.

Effects of iACT on Reducing Depressive Symptoms

A meta-analysis of 31 RCTs (N=4124 participants) showed that iACT had a small effect on reducing depressive symptoms at the immediate posttest compared with control groups overall (SMD 0.35, 95% CI 0.27-0.44; Figure 2). There was no statistically significant subgroup difference at the immediate posttest (χ^2_1=0.8; *P*=.37), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and

subgroup 2: iACT vs passive control groups) at the immediate posttest were not statistically different from one another. Small effects of iACT on depressive symptoms were found at the immediate posttest regardless of the control group conditions in 8 studies (N=1155 participants) that compared iACT with active control groups (SMD 0.29, 95% CI 0.08-0.49) and 23 studies (N=2969) that compared iACT with passive control conditions (SMD 0.39, 95% CI 0.31-0.46).

A meta-analysis of 13 RCTs with follow-up data (N=1645 participants) revealed that iACT had a small effect on reducing depressive symptoms at follow-up compared with control groups overall (SMD 0.39, 95% CI 0.20-0.57; Figure 3). There was no statistically significant subgroup difference at follow-up (χ^2_1=0.3; *P*=.58), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at follow-up were not statistically different from one another. The iACT had a medium effect compared with passive control groups at follow-up (8 studies that involved 732 participants, SMD 0.44, 95% CI 0.29-0.59), but iACT was not significantly different from active control groups (5 studies that involved 913 participants, SMD 0.33, 95% CI −0.03 to 0.69).

Figure 2. Forest plots showing effects of internet-based acceptance and commitment therapy on depressive symptoms at the immediate posttest. iACT: internet-based acceptance and commitment therapy [15-19,21-25,27-38,40-43,46,48-51].

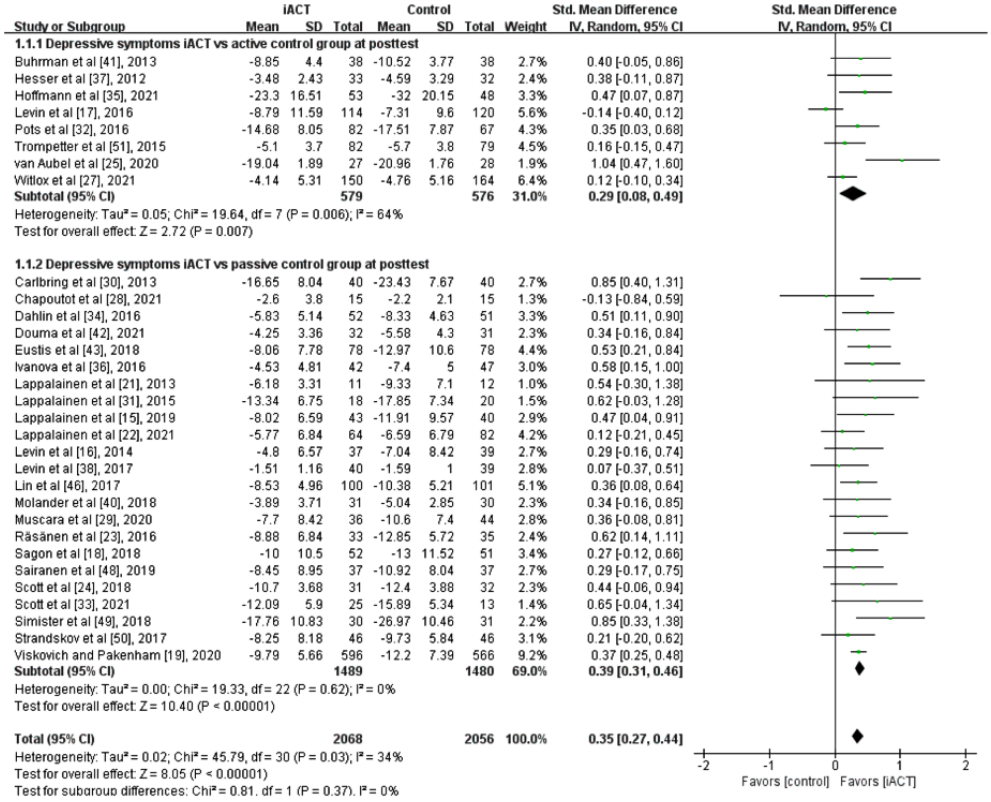
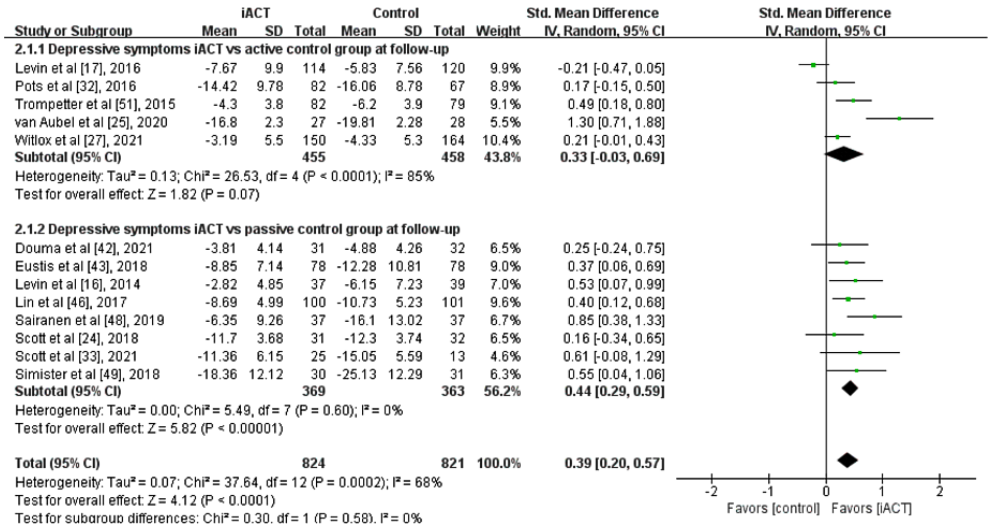


Figure 3. Forest plots showing effects of internet-based acceptance and commitment therapy on depressive symptoms at follow-up. iACT: internet-based acceptance and commitment therapy [16,17,24,25,27,32,33,42,43,46,48,49,51].



Effects of iACT on Reducing Anxiety

A meta-analysis of 22 RCTs (N=2407 participants) found that iACT had a small effect on reducing anxiety at the immediate posttest compared with control groups overall (SMD 0.30, 95% CI 0.18-0.42; Figure 4). There was no statistically significant subgroup difference at the immediate posttest (χ^2_1=0.2; *P*=.65), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at the immediate posttest were not statistically different from one another. Small effects of iACT on anxiety were found at the immediate posttest regardless of the control group conditions in 8 studies (N=1155 participants) that compared iACT with active control groups (SMD 0.36, 95% CI 0.11-0.62) and 14 studies (N=1251 participants) that compared iACT with passive control conditions (SMD 0.30, 95% CI 0.19-0.41).

A meta-analysis of 10 RCTs with follow-up data (N=1483 participants) showed that iACT had a small effect on reducing anxiety at follow-up compared with control groups overall (SMD 0.23, 95% CI 0.13-0.33; Figure 5). There was a statistically significant subgroup difference at follow-up (χ^2_1=5.9; *P*=.01), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at follow-up were statistically different from one another. At follow-up, iACT had small effects compared with both passive control groups (5 studies that involved 570 participants, SMD 0.39, 95% CI 0.23-0.56) and active control groups (5 studies that involved 913 participants, SMD 0.13, 95% CI 0.00-0.26), but the effect size of iACT was larger compared with passive control groups.

Figure 4. Forest plots showing effects of internet-based acceptance and commitment therapy on anxiety at the immediate posttest. iACT: internet-based acceptance and commitment therapy [16,17,19,23,25,27-30,32,34-38,40-43,48,50,51].

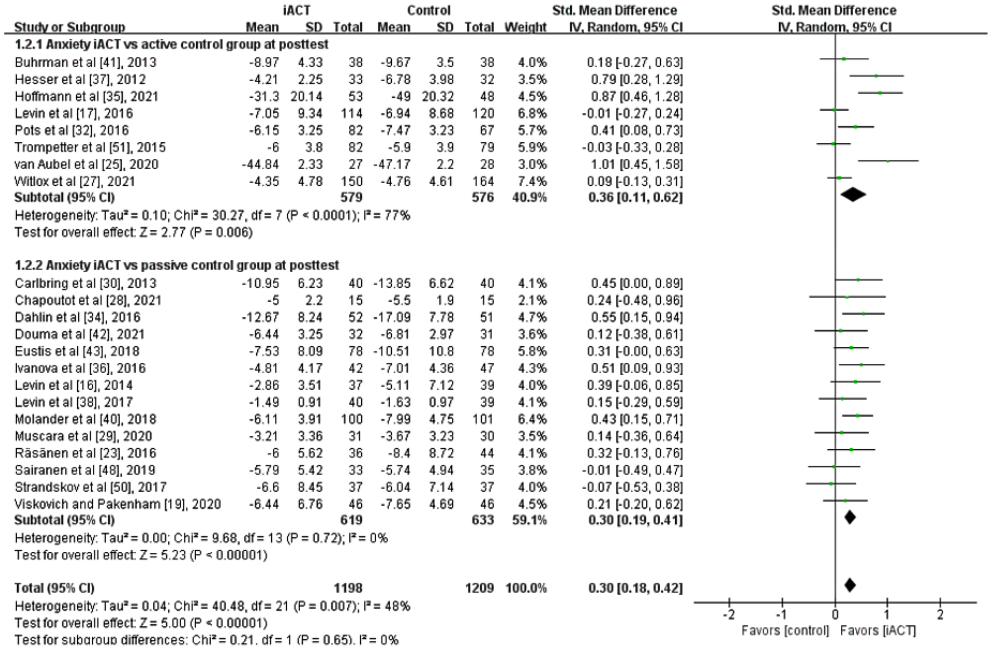
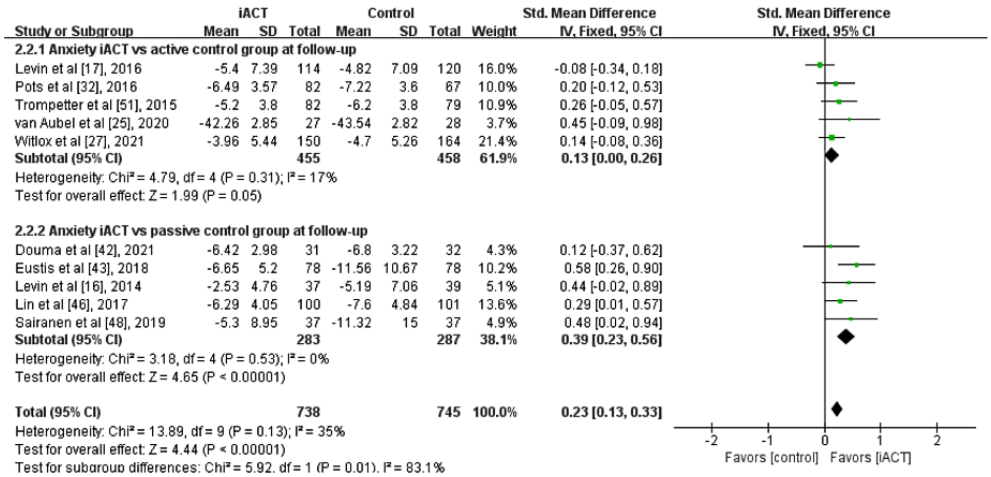


Figure 5. Forest plots showing effects of internet-based acceptance and commitment therapy on anxiety at follow-up. iACT: internet-based acceptance and commitment therapy [16,17,25,27,32,42,43,46,48,51].



Effects of iACT on Reducing Stress

A meta-analysis of 12 RCTs (N=2260 participants) revealed that iACT had a small effect on reducing stress at the immediate posttest compared with control groups overall (SMD 0.25, 95% CI 0.11-0.40; Figure 6). There was no statistically significant subgroup difference at the immediate posttest (χ^2_1=0.1; *P*=.71), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at the immediate posttest were not statistically different from one another. The iACT had a small effect on reducing stress compared with passive control groups at the immediate posttest (10 studies that involved 1961 participants, SMD 0.32, 95% CI 0.25-0.43), but iACT was not significantly different from active control groups (2 studies that involved 299 participants, SMD 0.17, 95% CI −0.61 to 0.96).

A meta-analysis of 5 RCTs with follow-up data (N=646 participants) found that iACT did not differ from control groups in reducing stress at follow-up overall (SMD 0.30, 95% CI −0.01 to 0.61; Figure 7). There was a statistically significant subgroup difference at follow-up (χ^2_1=6.7; *P*=.01), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at follow-up were statistically different from one another. The iACT had a medium effect on reducing stress compared with passive control groups at follow-up (4 studies that involved 412 participants, SMD 0.41, 95% CI 0.13-0.70), but iACT was not significantly different from active control groups (1 study that involved 234 participants, SMD −0.10, 95% CI −0.35 to 0.16).

Figure 6. Forest plots showing effects of internet-based acceptance and commitment therapy on stress at the immediate posttest. iACT: internet-based acceptance and commitment therapy [16,17,19,20,21,23,29,37,43,44,47,48].

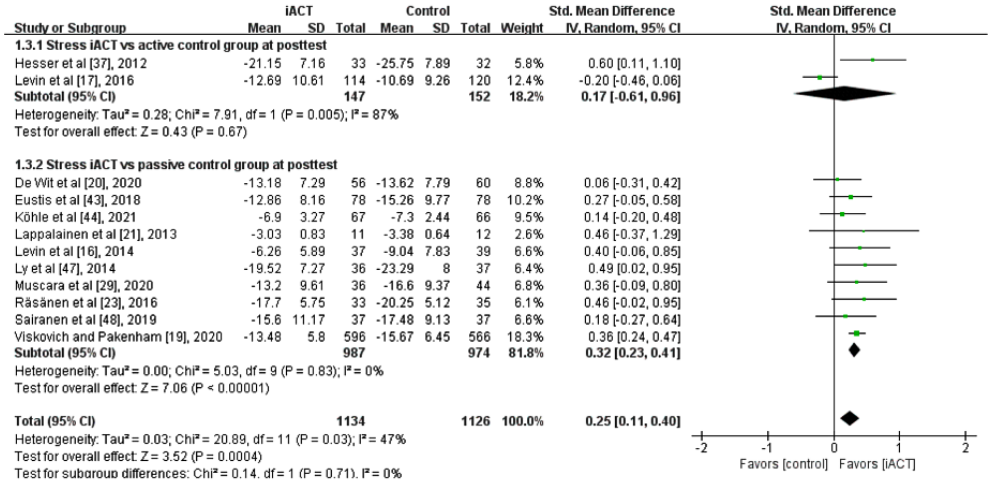
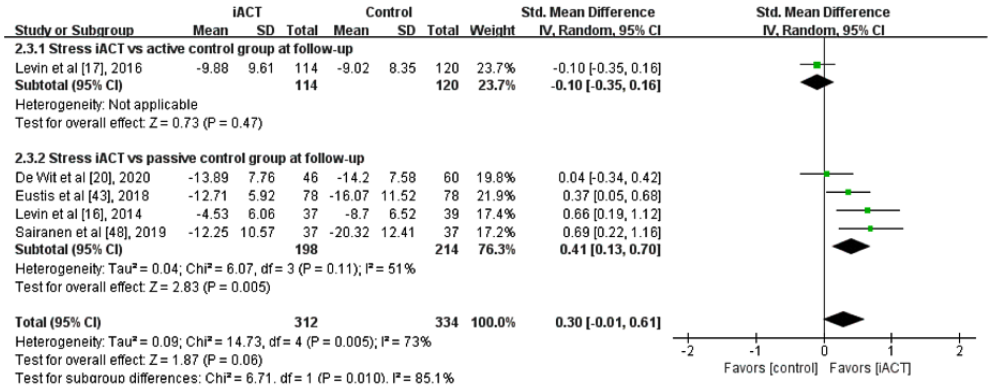


Figure 7. Forest plots showing effects of internet-based acceptance and commitment therapy on stress at follow-up. iACT: internet-based acceptance and commitment therapy [16,17,20,43,48].



Effects of iACT on Reducing Psychological Distress

A meta-analysis of 12 RCTs (N=890 participants) found that iACT had a small effect on reducing psychological distress at the immediate posttest compared with control groups overall (SMD 0.31, 95% CI 0.07-0.56; Figure 8). There was no significant subgroup difference at the immediate posttest (χ^2_1=0.0; *P*=.84), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at the immediate posttest were not statistically different from one another. The iACT had a small effect on reducing psychological distress compared with passive control groups at the immediate posttest (10 studies that involved 770 participants, SMD 0.33, 95% CI 0.18-0.47), but iACT was not significantly different from active control groups (2 studies that involved 120 participants, SMD 0.12, 95% CI −1.92 to 2.16).

A meta-analysis of 4 RCTs with follow-up data (N=314 participants) revealed that iACT had a small effect on reducing psychological distress at follow-up compared with control groups overall (SMD 0.31, 95% CI 0.09-0.54; Figure 9). There was no significant subgroup difference at follow-up (χ^2_1=0.8; *P*=.36), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at follow-up were statistically different from one another. The iACT had a medium effect on psychological distress compared with an active control group at follow-up (1 study that involved 55 participants, SMD 0.54, 95% CI 0.00-1.08), whereas the iACT had a small effect compared with passive control groups at follow-up (3 studies that involved 259 participants, SMD 0.26, 95% CI 0.02-0.51).

Figure 8. Forest plots showing effects of internet-based acceptance and commitment therapy on psychological distress at the immediate posttest. iACT: internet-based acceptance and commitment therapy [15,20,21,25,31,37-39,42,44,45,47].

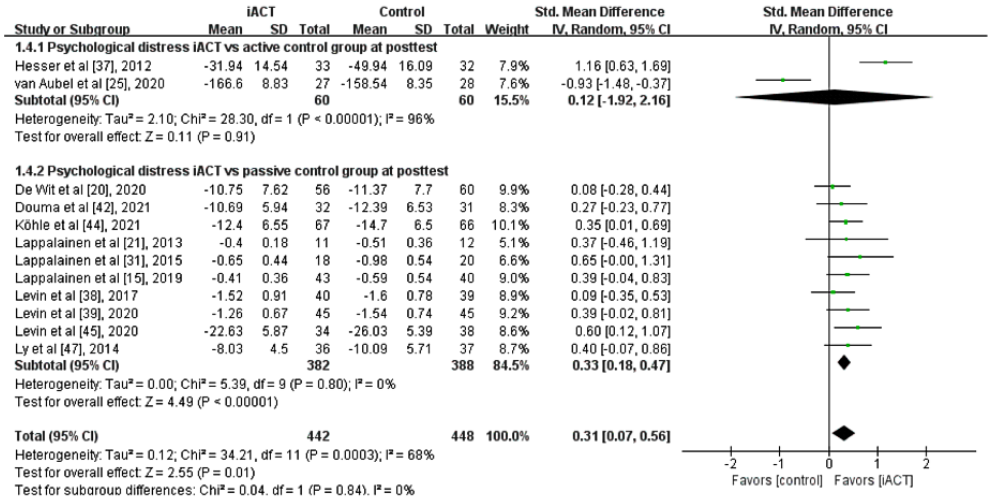
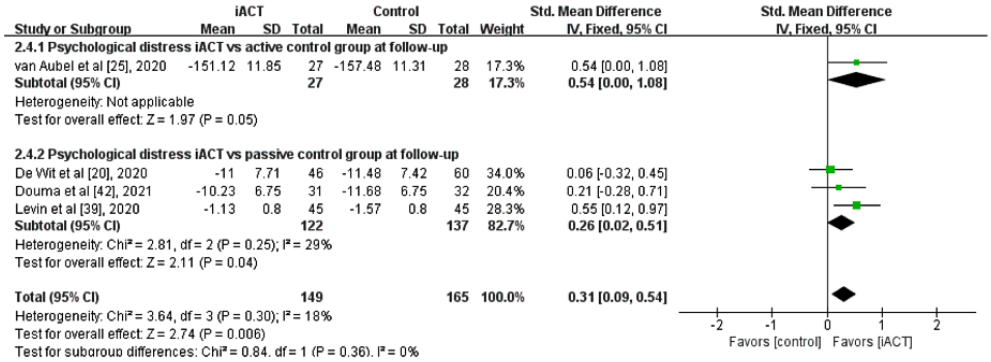


Figure 9. Forest plots showing effects of internet-based acceptance and commitment therapy on psychological distress at follow-up. iACT: internet-based acceptance and commitment therapy [20,25,39,42].



Effects of iACT on Improving QoL

A meta-analysis of 14 RCTs (N=1232 participants) showed that iACT had a small effect on improving QoL at the immediate posttest compared with control groups overall (SMD 0.28, 95% CI 0.14-0.41; Figure 10). There was no statistically significant subgroup difference at the immediate posttest (χ^2_1=0.1; *P*=.79), indicating that the effects of the 2 subgroups (ie, subgroup 1: iACT vs active control groups and subgroup 2: iACT vs passive control groups) at the immediate posttest were not statistically different from one another. The iACT had a small effect on improving QoL compared with passive control groups at the immediate posttest (11 studies that involved 990 participants, SMD 0.29, 95% CI 0.15-0.42), but iACT was not significantly different from active control groups (3 studies that involved 242 participants, SMD 0.23, 95% CI −0.22 to 0.68).

All 3 studies with follow-up data (N=463 participants) compared iACT with passive control groups only, so no subgroup analysis at follow-up was conducted. The iACT had a small effect on improving QoL compared with passive control groups at follow-up (SMD 0.22, 95% CI 0.03-0.40; Figure 11).

Figure 10. Forest plots showing effects of internet-based acceptance and commitment therapy on quality of life at the immediate posttest. iACT: internet-based acceptance and commitment therapy [20,21,26,28,30,34-37,40,41,43,46,50].

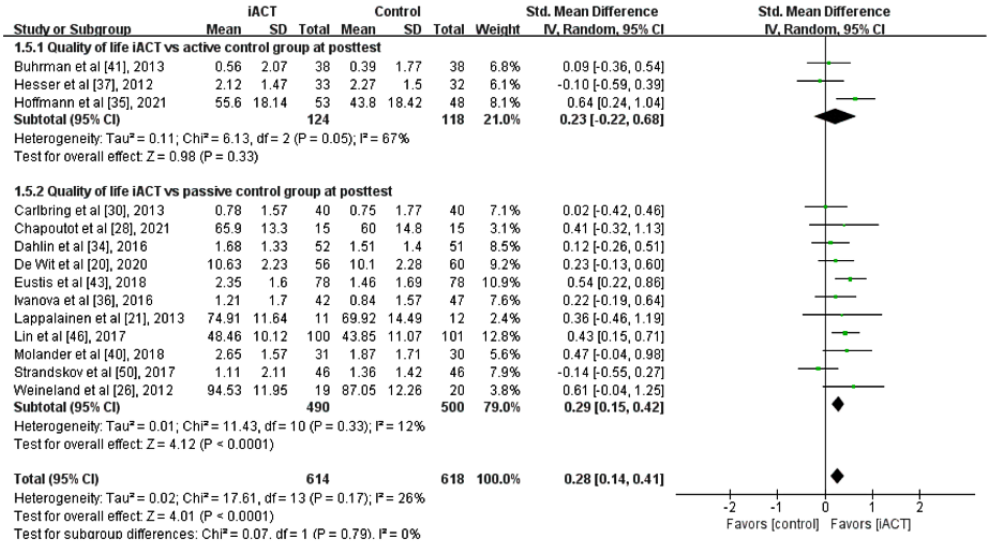
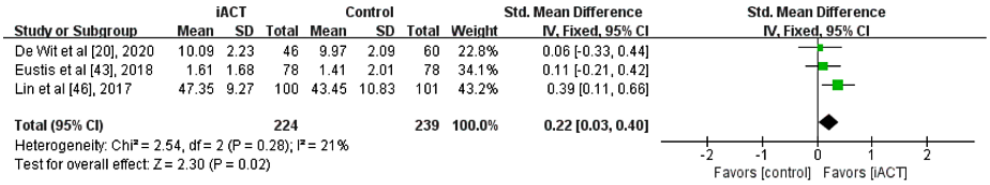


Figure 11. Forest plots showing effects of internet-based acceptance and commitment therapy on quality of life at follow-up. iACT: internet-based acceptance and commitment therapy.



Subgroup Analyses According to the Use of Therapist Guidance

Subgroup analyses showed small effects of iACT with therapist guidance on depressive symptoms at the immediate posttest (26 studies that involved 2466 participants, SMD 0.38, 95% CI 0.29-0.48), anxiety at the immediate posttest (20 studies that involved 2097 participants, SMD 0.30, 95% CI 0.21-0.38) and at follow-up (8 studies that involved 1173 participants, SMD 0.28, 95% CI 0.17-0.40), stress at the immediate posttest (9 studies that involved 788 participants, SMD 0.28, 95% CI 0.14-0.42) and at follow-up (3 studies that involved 336 participants, SMD 0.35, 95% CI 0.01-0.68), and psychological distress at the immediate posttest (11 studies that involved 807 participants, SMD 0.31, 95% CI 0.04-0.57) compared with control groups. A subgroup analysis found a medium effect of iACT with therapist guidance on depressive symptoms at follow-up (11 studies that involved 1335 participants, SMD 0.43, 95% CI 0.27-0.59). However, subgroup analyses revealed small effects of iACT without therapist guidance on depressive symptoms (5 studies that involved 1658 participants, SMD 0.24, 95% CI 0.00-0.48) and anxiety (3 studies that involved 1472 participants, SMD 0.18, 95% CI 0.08-0.28) at the immediate posttest only, with smaller effect sizes than those for iACT with therapist guidance. The iACT without therapist guidance was not significantly different from the control groups in depressive symptoms at follow-up (2 studies that involved 310 participants, SMD 0.14, 95% CI −0.59 to 0.86), anxiety at follow-up (2 studies that involved 310 participants, SMD 0.14, 95% CI −0.36 to 0.64), stress at the immediate posttest (4 studies that involved 1514 participants, SMD 0.08, 95% CI −0.29 to 0.46) and at follow-up (2 studies that involved 310 participants, SMD 0.25, 95% CI −0.48 to 0.99), and psychological distress at the immediate posttest (2 studies that involved 125 participants, SMD 0.09, 95% CI −0.56 to 0.74). There was no statistically significant subgroup difference in all the outcomes, indicating no statistically significant difference among studies according to the use of therapist guidance. Forest plots of these subgroup analyses are illustrated in Figures S1-S7 in Multimedia Appendix 3 [13,15-25,27-51].

Subgroup Analyses According to Delivery Mode

Subgroup analyses showed small effects of web-based ACT on depressive symptoms at the immediate posttest (23 studies that involved 3345 participants, SMD 0.35, 95% CI 0.28-0.42) and at follow-up (10 studies that involved 1213 participants, SMD 0.36, 95% CI 0.16-0.57), anxiety at the immediate posttest (18 studies that involved 3022 participants, SMD 0.29, 95% CI 0.18-0.40) and at follow-up (8 studies that involved 1114 participants, SMD 0.25, 95% CI 0.13-0.37), stress at the immediate posttest (9 studies that involved 2015 participants, SMD 0.21, 95% CI 0.03-0.40), and QoL at the immediate posttest (10 studies that involved 1024 participants, SMD 0.28, 95% CI 0.15-0.40) and at follow-up (2 studies that involved 357 participants, SMD 0.26, 95% CI 0.05-0.47) and medium effects of web-based ACT on psychological distress at the immediate posttest (10 studies that involved 738 participants, SMD 0.40, 95% CI 0.20-0.60) and at follow-up (2 studies that involved 153 participants, SMD 0.41, 95% CI 0.08-0.73) compared with control groups. Subgroup analyses revealed

small effects of iACT accompanied by in-person ACT sessions on depressive symptoms (6 studies that involved 669 participants, SMD 0.28, 95% CI 0.13-0.43) and QoL (3 studies that involved 178 participants, SMD 0.33, 95% CI 0.03-0.63) at the immediate posttest only compared with control groups, findings that may be because of a smaller number of studies included for these subgroup analyses. Few studies compared videoconferencing ACT with control groups (ie, 2 studies for depressive symptoms and anxiety), and these subgroup analyses showed no statistically significant difference of videoconferencing ACT from control groups in these outcomes. There was no statistically significant subgroup difference in all the outcomes (P>.05), indicating no statistically significant difference among the studies according to delivery mode. Forest plots of these subgroup analyses are illustrated in Figures S8-S17 in Multimedia Appendix 4 [13,15-51].

Subgroup Analyses According to the Targeted Participants

Subgroup analyses found medium effects of iACT on depressive symptoms (5 studies that involved 360 participants, SMD 0.62, 95% CI 0.40-0.83]) and anxiety (4 studies that involved 607 participants, SMD 0.48, 95% CI 0.11-0.85) at the immediate posttest compared with control groups when studies directly targeted participants with depressive symptoms and anxiety, respectively. In contrast, subgroup analyses showed small effects of iACT on depressive symptoms (26 studies that involved 3764 participants, SMD 0.31, 95% CI 0.25-0.37) and anxiety (19 studies that involved 2962 participants, SMD 0.24, 95% CI 0.14-0.35) at the immediate posttest compared with control groups when studies did not involve targeted participants for each of the outcomes (eg, studies that involved participants regardless of depressive symptoms at baseline). Subgroup analyses found small effects of iACT on stress at the immediate posttest (11 studies that involved 2199 participants, SMD 0.26, 95% CI 0.17-0.34) and on depressive symptoms (10 studies that involved 1403 participants, SMD 0.33, 95% CI 0.15-0.52) and anxiety (9 studies that involved 1169 participants, SMD 0.26, 95% CI 0.14-0.37) at follow-up compared with control groups when studies did not involve targeted participants for each of the outcomes. However, there was no statistically significant difference of iACT from control groups in stress and psychological distress at the immediate posttest and in depressive symptoms and anxiety at follow-up when studies directly targeted participants for each of the outcomes, findings that may be because of a smaller number of studies included for these subgroup analyses (eg, 2 studies for stress). There was no statistically significant subgroup difference in all the outcomes (P>.05), except for depressive symptoms at the immediate posttest (P=.007). These findings indicate that there was no statistically significant difference among studies according to the use of targeted participants in all the outcomes, except for depressive symptoms at the immediate posttest, for which a statistically significant, larger effect of iACT was found when studies targeted people with depressive symptoms (ie, SMD 0.62 vs SMD 0.31). Forest plots of these subgroup analyses are illustrated in Figures S18-S24 in Multimedia Appendix 5 [13,15-25,27-51].

Risk of Bias of the Included Studies

Of the 39 included studies, 20 (51%) had an unclear risk of bias, 14 (36%) had a low risk of bias, and 5 (13%) had a high risk of bias overall (Multimedia Appendix 6 [13-51]). A domain for blinding of participants and personnel was not considered as the key domain for the overall risk of bias within a study because studies that involved passive control conditions were less able to conceal the group allocation from participants. The overall risk of bias across 39 studies was interpreted as unclear because most information was from studies with an unclear risk of bias [12].

Discussion

Principal Findings

This systematic review with meta-analysis identified 39 studies that assessed the efficacy of iACT for depressive symptoms, anxiety, stress, psychological distress, and QoL. This study found that iACT had small effects on reducing depressive symptoms, anxiety, stress, and psychological distress and improving QoL at the immediate posttest and follow-up. There was no significant effect of iACT on stress at follow-up.

One of the previous meta-analyses found that iACT had a small effect on depressive symptoms and nonsignificant effects on anxiety and QoL; it included 10 studies involving depressive symptoms, 7 studies involving anxiety, and 8 studies involving QoL [9]. As more studies on iACT (ie, 30 studies on iACT) have been conducted since 2016, this study's meta-analyses included 31 studies on depressive symptoms, 22 studies on anxiety, and 14 studies on QoL. This may explain why, contrary to Brown et al [9], this study found small effects of iACT on anxiety and QoL in addition to depressive symptoms with the increased statistical power from a larger total sample size pooled from more studies.

Another prior meta-analysis [10] found small effects of iACT on depressive symptoms (25 studies at posttest) and anxiety (20 studies at posttest), so this study confirms the previous meta-analysis with more studies added (ie, 6 more studies for depressive symptoms and 2 more studies for anxiety). More importantly, this study found small effects of iACT on stress and psychological distress, which were not included in the meta-analyses by Brown et al [9] and Thompson et al [10]. Such positive findings are supported by previous studies that suggested negative relationships of the ACT processes, such as mindfulness and acceptance, with depressive symptoms, anxiety, stress, and psychological distress, and positive relationships with QoL [52,53]. ACT promoting acceptance, mindfulness, and committed action to living in alignment with values may improve QoL and decrease psychological distress by helping those who receive the therapy better manage uncomfortable or negative emotions, thoughts, and experiences; accept them without judgment; and move forward to valued living [54,55]. The findings of this study contribute to the body of evidence regarding the effects of iACT on depressive symptoms, anxiety, stress, psychological distress, and QoL and may be applicable in any population as ACT is a transdiagnostic approach [1].

Subgroup analyses for each outcome were conducted in this study according to the type of control groups but were not conducted in prior meta-analyses [9,10]. Subgroup analyses showed small to medium effects of iACT on all the outcomes at the immediate posttest and follow-up compared with passive control groups. In contrast, subgroup analyses that compared iACT with active control groups (eg, CBT and mental health education) found no differences between groups on stress, psychological distress, and QoL at the immediate posttest or on depressive symptoms, anxiety, and stress at follow-up.

However, it should be noted that relatively fewer studies have compared the effects of iACT with active control conditions. There were 2 to 5 times more studies that compared iACT with passive control conditions than those that compared iACT with active control conditions in all the outcomes, except for anxiety at follow-up. There were fewer studies that compared iACT with active control conditions, especially for stress and psychological distress at the immediate posttest and follow-up. More importantly, active control conditions varied among the included studies, including web-based discussion forums [35,37,41], web-based expressive writing [32,51], web-based mental health education [17], a web-based smoking cessation intervention [14], web-based CBT [13], in-person CBT [27], and in-person documentary discussions [25]. Such differences in the active control conditions might explain the higher statistical heterogeneity when comparing iACT with active control groups than heterogeneity when comparing iACT with passive control groups. Overall, these findings suggest a need for more studies comparing iACT with active control conditions enough to assess whether iACT is comparable or superior to each of the other evidence-based interventions in decreasing depressive symptoms, anxiety, stress, and psychological distress and improving QoL in future meta-analyses.

This study conducted subgroup analyses according to the use of therapist guidance, delivery modes, and the use of targeted participants and found no statistically significant subgroup differences among studies according to these 3 characteristics in all the outcomes, except for the subgroup difference among studies according to the use of targeted participants for depressive symptoms at the immediate posttest (ie, a statistically significant, larger effect of iACT when studies targeted people with depressive symptoms). However, more studies are needed to confirm these findings, especially in subgroup analyses according to delivery modes and targeted participants, as there were only a few studies that involved delivery modes other than web-based ACT modules or that involved directly targeted participants (eg, only 2 studies that involved people with stress).

Limitations

This review had some limitations. The Bonferroni correction method might be needed to adjust the results of multiple comparisons to reduce the false positive error rate (ie, type I errors) in a meta-analysis [56]. As 10 comparisons were made in this meta-analysis, the level of significance (*P* value) after Bonferroni correction was .005 (0.05/10). With the adjusted level of significance, most results remained the same, except for psychological distress at posttest and follow-up and QoL at follow-up. As RevMan does not allow users to choose the levels of significance other than 0.1, 0.05, or 0.01, and the values in the forest plots, including CIs, cannot be changed accordingly, we could not use the adjusted level of significance (ie, *P*=.005) in this meta-analysis. A total of 4 electronic databases were searched; therefore, some relevant articles could have been missed if they were published only in other databases. Only studies written in English were searched and included in this review, which could create a publication bias. One author with extensive experience in comprehensive literature reviews and expertise in ACT searched the literature; therefore, this review did not include 2 independent reviewers in the search process. A recent methodological systematic review found that single screening for study selection in systematic reviews conducted by an experienced reviewer had no impact on the findings of the meta-analysis [57]. The overall risk of bias across the included RCTs was interpreted as unclear, indicating the need for high-quality studies to better determine the efficacy of iACT for psychological distress and QoL. This study did not aim to compare iACT with face-to-face ACT because the focus was on comparing iACT with non-ACT control conditions. Further meta-analyses may consider comparing iACT with face-to-face ACT to determine whether iACT is comparable with face-to-face ACT when more studies are available. Further meta-analyses may consider examining whether the effects of blended ACT, which involves iACT and face-to-face ACT, differ from iACT only or face-to-face ACT only.

Acknowledgments

This research was supported by a grant from the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute and funded by the Ministry of Health and Welfare, Republic of Korea (grant HR21C0885040021).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms used in database searches.
[PDF File (Adobe PDF File), 92 KB - jmir_v24i12e39727_app1.pdf]

Multimedia Appendix 2
Characteristics of the included studies.
[PDF File (Adobe PDF File), 216 KB - jmir_v24i12e39727_app2.pdf]

Multimedia Appendix 3
Forest plots subgroup analyses according to the use of therapist guidance (Figures S1-S7).
[PDF File (Adobe PDF File), 1140 KB - jmir_v24i12e39727_app3.pdf]

Multimedia Appendix 4
Forest plots subgroup analyses according to delivery mode (Figures S8-S17).
[PDF File (Adobe PDF File), 264 KB - jmir_v24i12e39727_app4.pdf]

Multimedia Appendix 5
Forest plots subgroup analyses according to the targeted participants (Figures S18-S24).
[PDF File (Adobe PDF File), 201 KB - jmir_v24i12e39727_app5.pdf]

Multimedia Appendix 6
Forest plots showing effects of internet-based acceptance and commitment therapy on quality of life at follow-up. iACT: internet-based acceptance and commitment therapy [20,43,46].
[PDF File (Adobe PDF File), 122 KB - jmir_v24i12e39727_app6.pdf]

References

1. Dindo L, Van Liew JR, Arch JJ. Acceptance and commitment therapy: a transdiagnostic behavioral intervention for mental health and medical conditions. Neurotherapeutics 2017 Jul;14(3):546-553 [FREE Full text] [doi: 10.1007/s13311-017-0521-3] [Medline: 28271287]
2. Hayes S, Pistorello J, Levin M. Acceptance and commitment therapy as a unified model of behavior change. Counsel Psychol 2012 Sep 26;40(7):976-1002 [FREE Full text] [doi: 10.1177/0011000012460836]
3. Bai Z, Luo S, Zhang L, Wu S, Chi I. Acceptance and Commitment Therapy (ACT) to reduce depression: a systematic review and meta-analysis. J Affect Disord 2020 Jan 01;260:728-737. [doi: 10.1016/j.jad.2019.09.040] [Medline: 31563072]
4. Han A, Yuen HK, Jenkins J. Acceptance and commitment therapy for family caregivers: a systematic review and meta-analysis. J Health Psychol 2021 Jan 10;26(1):82-102. [doi: 10.1177/1359105320941217] [Medline: 32659142]
5. Hughes LS, Clark J, Colclough JA, Dale E, McMillan D. Acceptance and commitment therapy (ACT) for chronic pain: a systematic review and meta-analyses. Clin J Pain 2017 Jun;33(6):552-568. [doi: 10.1097/AJP.0000000000000425] [Medline: 27479642]
6. Jansen JE, Gleeson J, Bendall S, Rice S, Alvarez-Jimenez M. Acceptance- and mindfulness-based interventions for persons with psychosis: a systematic review and meta-analysis. Schizophr Res 2020 Jan;215:25-37. [doi: 10.1016/j.schres.2019.11.016] [Medline: 31780349]
7. Han A. Mindfulness- and acceptance-based interventions for symptom reduction in individuals with multiple sclerosis: a systematic review and meta-analysis. Arch Phys Med Rehabil 2021 Oct;102(10):2022-31.e4. [doi: 10.1016/j.apmr.2021.03.011] [Medline: 33812883]
8. 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]
9. Brown M, Glendenning A, Hoon AE, John A. Effectiveness of web-delivered acceptance and commitment therapy in relation to mental health and well-being: a systematic review and meta-analysis. J Med Internet Res 2016 Aug 24;18(8):e221 [FREE Full text] [doi: 10.2196/jmir.6200] [Medline: 27558740]
10. Thompson EM, Destree L, Albertella L, Fontenelle LF. Internet-based acceptance and commitment therapy: a transdiagnostic systematic review and meta-analysis for mental health outcomes. Behav Ther 2021 Mar;52(2):492-507. [doi: 10.1016/j.beth.2020.07.002] [Medline: 33622516]
11. 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]
12. Higgins J, Thomas J, Chandler J. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Chichester (UK): John Wiley & Sons; Sep 20, 2019.
13. Barrett K, Stewart I. A preliminary comparison of the efficacy of online Acceptance and Commitment Therapy (ACT) and Cognitive Behavioural Therapy (CBT) stress management interventions for social and healthcare workers. Health Soc Care Community 2021 Jan 02;29(1):113-126. [doi: 10.1111/hsc.13074] [Medline: 32613644]

14. Heffner J, Kelly M, Waxmonsky J, Mattocks K, Serfozo E, Bricker JB, et al. Pilot randomized controlled trial of web-delivered acceptance and commitment therapy versus smokefree.gov for smokers with bipolar disorder. Nicotine Tob Res 2020 Aug 24;22(9):1543-1552 [FREE Full text] [doi: 10.1093/ntr/ntz242] [Medline: 31883336]
15. Lappalainen P, Langrial S, Oinas-Kukkonen H, Muotka J, Lappalainen R. ACT for sleep - internet-delivered self-help ACT for sub-clinical and clinical insomnia: a randomized controlled trial. J Contextual Behav Sci 2019 Apr;12:119-127. [doi: 10.1016/j.jcbs.2019.04.001]
16. Levin ME, Pistorello J, Seeley JR, Hayes SC. Feasibility of a prototype web-based acceptance and commitment therapy prevention program for college students. J Am Coll Health 2014;62(1):20-30 [FREE Full text] [doi: 10.1080/07448481.2013.843533] [Medline: 24313693]
17. Levin ME, Hayes SC, Pistorello J, Seeley JR. Web-based self-help for preventing mental health problems in universities: comparing acceptance and commitment training to mental health education. J Clin Psychol 2016 Mar;72(3):207-225. [doi: 10.1002/jclp.22254] [Medline: 26784010]
18. Sagon A, Danitz S, Suvak M, Orsillo S. The Mindful Way through the Semester: evaluating the feasibility of delivering an acceptance-based behavioral program online. J Contextual Behavioral Sci 2018 Jul;9:36-44 [FREE Full text] [doi: 10.1016/j.jcbs.2018.06.004]
19. Viskovich S, Pakenham KI. Randomized controlled trial of a web-based Acceptance and Commitment Therapy (ACT) program to promote mental health in university students. J Clin Psychol 2020 Jun 30;76(6):929-951. [doi: 10.1002/jclp.22848] [Medline: 31468528]
20. De Wit J, Beelen A, Drossaert CH, Kolijn R, Van Den Berg LH, SchrÖder CD, et al. Blended psychosocial support for partners of patients with ALS and PMA: results of a randomized controlled trial. Amyotroph Lateral Scler Frontotemporal Degener 2020 Aug 02;21(5-6):344-354. [doi: 10.1080/21678421.2020.1757114] [Medline: 32362155]
21. Lappalainen P, Kaipainen K, Lappalainen R, Hoffrén H, Myllymäki T, Kinnunen M, et al. Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. JMIR Res Protoc 2013 Jan 09;2(1):e1 [FREE Full text] [doi: 10.2196/resprot.2389] [Medline: 23611946]
22. Lappalainen R, Lappalainen P, Puolakanaho A, Hirvonen R, Eklund K, Ahonen T, et al. The Youth Compass -the effectiveness of an online acceptance and commitment therapy program to promote adolescent mental health: a randomized controlled trial. J Contextual Behav Sci 2021 Apr;20:1-12. [doi: 10.1016/j.jcbs.2021.01.007]
23. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. Behav Res Ther 2016 Mar;78:30-42. [doi: 10.1016/j.brat.2016.01.001] [Medline: 26848517]
24. Scott W, Chilcot J, Guildford B, Daly-Eichenhardt A, McCracken LM. Feasibility randomized-controlled trial of online Acceptance and Commitment Therapy for patients with complex chronic pain in the United Kingdom. Eur J Pain 2018 Apr 28 (forthcoming). [doi: 10.1002/ejp.1236] [Medline: 29704880]
25. van Aubel E, Bakker JM, Batink T, Michielse S, Goossens L, Lange I, et al. Blended care in the treatment of subthreshold symptoms of depression and psychosis in emerging adults: a randomised controlled trial of Acceptance and Commitment Therapy in Daily-Life (ACT-DL). Behav Res Ther 2020 May;128:103592. [doi: 10.1016/j.brat.2020.103592] [Medline: 32146218]
26. Weineland S, Arvidsson D, Kakoulidis TP, Dahl J. Acceptance and commitment therapy for bariatric surgery patients, a pilot RCT. Obes Res Clin Pract 2012 Jan;6(1):e1-90. [doi: 10.1016/j.orcp.2011.04.004] [Medline: 24331170]
27. Witlox M, Garnefski N, Kraaij V, de Waal MW, Smit F, Bohlmeijer E, et al. Blended acceptance and commitment therapy versus face-to-face cognitive behavioral therapy for older adults with anxiety symptoms in primary care: pragmatic single-blind cluster randomized trial. J Med Internet Res 2021 Mar 26;23(3):e24366 [FREE Full text] [doi: 10.2196/24366] [Medline: 33769293]
28. Chapoutot M, Peter-Derex L, Schoendorff B, Faivre T, Bastuji H, Putois B. Telehealth-delivered CBT-I programme enhanced by acceptance and commitment therapy for insomnia and hypnotic dependence: a pilot randomized controlled trial. J Sleep Res 2021 Feb 05;30(1):e13199 [FREE Full text] [doi: 10.1111/jsr.13199] [Medline: 33020985]
29. Muscara F, McCarthy MC, Rayner M, Nicholson JM, Dimovski A, McMillan L, et al. Effect of a videoconference-based online group intervention for traumatic stress in parents of children with life-threatening illness: a randomized clinical trial. JAMA Netw Open 2020 Jul 01;3(7):e208507 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8507] [Medline: 32735335]
30. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki Å, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Disord 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
31. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-based acceptance and commitment therapy for depressive symptoms with minimal support: a randomized controlled trial. Behav Modif 2015 Nov;39(6):805-834. [doi: 10.1177/0145445515598142] [Medline: 26253644]
32. Pots WT, Fledderus M, Meulenbeek PA, ten Klooster PM, Schreurs KM, Bohlmeijer ET. Acceptance and commitment therapy as a web-based intervention for depressive symptoms: randomised controlled trial. Br J Psychiatry 2016 Jan;208(1):69-77. [doi: 10.1192/bjp.bp.114.146068] [Medline: 26250745]

33. Scott W, Guildford BJ, Badenoch J, Driscoll E, Chilcot J, Norton S, et al. Feasibility randomized-controlled trial of online acceptance and commitment therapy for painful peripheral neuropathy in people living with HIV: the OPEN study. Eur J Pain 2021 Aug 26;25(7):1493-1507. [doi: 10.1002/ejp.1762] [Medline: 33711209]
34. Dahlin M, Andersson G, Magnusson K, Johansson T, Sjögren J, Håkansson A, et al. Internet-delivered acceptance-based behaviour therapy for generalized anxiety disorder: a randomized controlled trial. Behav Res Ther 2016 Feb;77:86-95. [doi: 10.1016/j.brat.2015.12.007] [Medline: 26731173]
35. Hoffmann D, Rask CU, Hedman-Lagerlöf E, Jensen JS, Frostholm L. Efficacy of internet-delivered acceptance and commitment therapy for severe health anxiety: results from a randomized, controlled trial. Psychol Med 2020 May 14;51(15):2685-2695. [doi: 10.1017/s0033291720001312]
36. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided acceptance and commitment therapy for social anxiety disorder and/or panic disorder provided via the internet and a smartphone application: a randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
37. Hesser H, Gustafsson T, Lundén C, Henrikson O, Fattahi K, Johnsson E, et al. A randomized controlled trial of Internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. J Consult Clin Psychol 2012 Aug;80(4):649-661. [doi: 10.1037/a0027021] [Medline: 22250855]
38. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017 Jan 01;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
39. Levin ME, Krafft J, Hicks ET, Pierce B, Twohig MP. A randomized dismantling trial of the open and engaged components of acceptance and commitment therapy in an online intervention for distressed college students. Behav Res Ther 2020 Mar;126:103557. [doi: 10.1016/j.brat.2020.103557] [Medline: 32014692]
40. Molander P, Hesser H, Weineland S, Bergwall K, Buck S, Jäder Malmlöf J, et al. Internet-based acceptance and commitment therapy for psychological distress experienced by people with hearing problems: a pilot randomized controlled trial. Cogn Behav Ther 2018 Mar 12;47(2):169-184. [doi: 10.1080/16506073.2017.1365929] [Medline: 28895787]
41. Buhrman M, Skoglund A, Husell J, Bergström K, Gordh T, Hursti T, et al. Guided internet-delivered acceptance and commitment therapy for chronic pain patients: a randomized controlled trial. Behav Res Ther 2013 Jun;51(6):307-315. [doi: 10.1016/j.brat.2013.02.010] [Medline: 23548250]
42. Douma M, Maurice-Stam H, Gorter B, Krol Y, Verkleij M, Wiltink L, et al. Online psychosocial group intervention for parents: positive effects on anxiety and depression. J Pediatr Psychol 2021 Feb 19;46(2):123-134 [FREE Full text] [Medline: 33230541]
43. Eustis EH, Hayes-Skelton SA, Orsillo SM, Roemer L. Surviving and thriving during stress: a randomized clinical trial comparing a brief web-based therapist-assisted acceptance-based behavioral intervention versus waitlist control for college students. Behav Ther 2018 Nov;49(6):889-903. [doi: 10.1016/j.beth.2018.05.009] [Medline: 30316488]
44. Köhle N, Drossaert CH, Ten Klooster PM, Schreurs KM, Hagedoorn M, Van Uden-Kraan CF, et al. Web-based self-help intervention for partners of cancer patients based on acceptance and commitment therapy and self-compassion training: a randomized controlled trial with automated versus personal feedback. Support Care Cancer 2021 Sep 19;29(9):5115-5125 [FREE Full text] [doi: 10.1007/s00520-021-06051-w] [Medline: 33608762]
45. Levin M, Petersen J, Durward C, Bingeman B, Davis E, Nelson C, et al. A randomized controlled trial of online acceptance and commitment therapy to improve diet and physical activity among adults who are overweight/obese. Transl Behav Med 2021 Jun 17;11(6):1216-1225. [doi: 10.1093/tbm/ibaa123] [Medline: 33289785]
46. 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]
47. Ly K, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. Internet Interv 2014 Jul;1(3):95-101 [FREE Full text] [doi: 10.1016/j.invent.2014.06.003]
48. Sairanen E, Lappalainen R, Lappalainen P, Kaipainen K, Carlstedt F, Anclair M, et al. Effectiveness of a web-based acceptance and commitment therapy intervention for wellbeing of parents whose children have chronic conditions: a randomized controlled trial. J Contextual Behav Sci 2019 Jul;13:94-102. [doi: 10.1016/j.jcbs.2019.07.004]
49. Simister HD, Tkachuk GA, Shay BL, Vincent N, Pear JJ, Skrabek RQ. Randomized controlled trial of online acceptance and commitment therapy for fibromyalgia. J Pain 2018 Jul;19(7):741-753. [doi: 10.1016/j.jpain.2018.02.004] [Medline: 29481976]
50. Strandskov SW, Ghaderi A, Andersson H, Parmskog N, Hjort E, Wärn AS, et al. Effects of tailored and ACT-influenced internet-based CBT for eating disorders and the relation between knowledge acquisition and outcome: a randomized controlled trial. Behav Ther 2017 Sep;48(5):624-637. [doi: 10.1016/j.beth.2017.02.002] [Medline: 28711113]
51. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KM. Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]

52. Aaby A, Ravn SL, Kasch H, Andersen TE. The associations of acceptance with quality of life and mental health following spinal cord injury: a systematic review. Spinal Cord 2020 Feb 12;58(2):130-148. [doi: 10.1038/s41393-019-0379-9] [Medline: 31719667]
53. Pagnini F, Cavalera C, Rovaris M, Mendozzi L, Molinari E, Phillips D, et al. Longitudinal associations between mindfulness and well-being in people with multiple sclerosis. Int J Clin Health Psychol 2019 Jan;19(1):22-30 [FREE Full text] [doi: 10.1016/j.ijchp.2018.11.003] [Medline: 30619494]
54. Bendelin N, Björkdahl P, Risell M, Nelson KZ, Gerdle B, Andersson G, et al. Patients' experiences of internet-based Acceptance and commitment therapy for chronic pain: a qualitative study. BMC Musculoskelet Disord 2020 Apr 06;21(1):212 [FREE Full text] [doi: 10.1186/s12891-020-03198-1] [Medline: 32252707]
55. Large R, Samuel V, Morris R. A changed reality: experience of an acceptance and commitment therapy (ACT) group after stroke. Neuropsychol Rehabil 2020 Sep 29;30(8):1477-1496. [doi: 10.1080/09602011.2019.1589531] [Medline: 30924741]
56. Polanin JR, Pigott TD. The use of meta-analytic statistical significance testing. Res Synth Methods 2015 Mar;6(1):63-73. [doi: 10.1002/jrsm.1124] [Medline: 26035470]
57. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol 2019 Jun 28;19(1):132 [FREE Full text] [doi: 10.1186/s12874-019-0782-0] [Medline: 31253092]

Abbreviations

ACT: acceptance and commitment therapy
CBT: cognitive behavioral therapy
iACT: internet-based acceptance and commitment therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by R Kukafka, G Eysenbach; submitted 20.05.22; peer-reviewed by E Feofanova, S Sarejloo; comments to author 13.07.22; revised version received 24.07.22; accepted 08.10.22; published 09.12.22.

Please cite as:
Han A, Kim TH
Efficacy of Internet-Based Acceptance and Commitment Therapy for Depressive Symptoms, Anxiety, Stress, Psychological Distress, and Quality of Life: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(12):e39727
URL: https://www.jmir.org/2022/12/e39727
doi:10.2196/39727
PMID:36485030

Review

Dropout Rate in Digital Health Interventions for the Prevention of Skin Cancer: Systematic Review, Meta-analysis, and Metaregression

Juan-Carlos Hernández-Rodríguez[1], MSc, MD; Cristina García-Muñoz[2], PT, PhD; Juan Ortiz-Álvarez[1], MD; Francesc Saigí-Rubió[3], BSc, PhD; Julián Conejo-Mir[1,4], MD, PhD; Jose-Juan Pereyra-Rodriguez[1,4], MD, PhD

[1]Department of Dermatology, Virgen del Rocío University Hospital, Seville, Spain

[2]Department of Nursing and Physical Therapy, University of Cadiz, Cadiz, Spain

[3]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

[4]Department of Medicine, University of Seville, Seville, Spain

Corresponding Author:
Cristina García-Muñoz, PT, PhD
Department of Nursing and Physical Therapy
University of Cadiz
Avda. Ana de Viya 52
Cadiz, 11009
Spain
Phone: 34 956 01 90 00
Email: ccriss.g@gmail.com

Abstract

Background: Digital strategies are innovative approaches to the prevention of skin cancer, but the attrition following this kind of intervention needs to be analyzed.

Objective: The aim of this paper is to assess the dropouts from studies focused on digital strategies for the prevention of skin cancer.

Methods: We conducted this systematic review with meta-analyses and metaregression according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statements. Search terms for skin cancer, digital strategies, and prevention were combined to search PubMed, Scopus, Web of Science, CINAHL, and Cochrane Library from inception until July 2022. Randomized clinical trials that reported dropouts of participants and compared digital strategies with other interventions to prevent skin cancer in healthy or disease-free participants were included. Two independent reviewers extracted data for analysis. The Revised Cochrane Collaboration Bias tool was employed. We calculated the pooled dropout rate of participants through a meta-analysis of proportions and examined whether dropout was more or less frequent in digital interventions against comparators via an odds ratio (OR) meta-analysis. Data were pooled using a random-effects model. Subgroup meta-analyses were conducted in a meta-analysis of proportions and OR meta-analysis to assess the dropout events when data were sorted by digital interventions or control comparator. A univariate metaregression based on a random-effects model assessed possible moderators of dropout. Participants' dropout rates as pooled proportions were calculated for all groups combined, and the digital and comparator groups separately. OR>1 indicated higher dropouts for digital-based interventions. Metaregressions were performed for age, sex, length of intervention, and sample size.

Results: A total of 17 studies were included. The overall pooled dropout rate was 9.5% (95% CI 5.0-17.5). The subgroup meta-analysis of proportions revealed a dropout rate of 11.6% for digital strategies (95% CI 6.8-19.0) and 10.0% for comparators (95% CI 5.5-17.7). A trend of higher dropout rates for digital strategies was observed in the overall (OR 1.16, 95% CI 0.98-1.36) and subgroup OR meta-analysis, but no significant differences were found between the groups. None of the covariates moderated the effect size in the univariate metaregression.

Conclusions: Digital strategies had a higher dropout rate compared to other prevention interventions, but the difference was not significant. Standardization is needed regarding reporting the number of and reasons for dropouts.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42022329669; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=329669

(J Med Internet Res 2022;24(12):e42397) doi:10.2196/42397

KEYWORDS

skin cancer; digital health; dropout; prevention; systematic review; meta-analysis; meta analyses; review methodology; cancer; skin; dermatology; attrition; digital intervention; digital treatment; eHealth; randomized controlled trial; RCT

Introduction

Digital strategies have experienced a boom in use in prevention programs for skin cancer in recent years. Primary and secondary prevention programs are the mainstay to reduce the incidence rate of skin cancer [1]. In fact, recent publications have stated a stabilization in melanoma incidence in young cohorts, due to governmental efforts to promote prevention programs [2]. Nonetheless, cases of melanoma will continue to rise in the coming years, primarily in older adults [3]. The continuous rise in the incidence of skin cancer in recent decades suggests a current global public threat [4,5].

Digital strategies seem to be more effective in the prevention of skin cancer than other conventional strategies [6]. The former can be defined as interventions provided through a digital environment such as web-based interventions, smartphone apps, SMS text messaging, web-based videos, or wearable devices [7]. Digital approaches to the prevention of skin cancer present additional advantages such as feedback, interactivity, accessibility, and gamification, which make them suitable and attractive for stakeholders [8,9]. Conversely, possible drawbacks of digital strategies in dermatology could be their availability, financial aspects, reliability, security, confidentially, and lack of education and training of the user [10]. Given all these issues, the feasibility of randomized clinical trials (RCTs) in digital health research continues to be discussed [11,12]. However, digital strategies such as telemedicine in different areas of health care are expected to continue growing in the coming years [13].

The engagement of the patients with the prevention and digital strategies determines their effectiveness. Despite the increasing interest of researchers in implementing RCTs that analyze digital strategies, there is still no consensus in the literature on whether they positively or negatively influence the dropout and adherence of participants [14,15]. However, some authors have reported that the dropout rate was higher in digital strategies than analogue interventions [16,17]. Some of the reasons for the higher loss of participants could be the participant's reluctance to join remote research studies and mistrust in sharing data [18].

Dropout or attrition is a constant challenge for researchers in RCTs and other longitudinal studies [19,20]. In addition, characteristics of the target population could influence attrition, because maintaining prevention behaviors in healthy participants could be challenging [21]. The absence of perceiving disease, geographical location, or accessibility are some of the factors that could lead to the failure of long-term prevention strategies [22,23]. Disentangling the factors and trend in dropouts in RCTs would help researchers develop future digital interventions for the prevention of skin cancer.

No previous studies have analyzed dropout in digital strategies for skin cancer prevention; therefore, our aim was to systematically assess and meta-analyze the existing RCTs to calculate the overall pooled dropout rate and to examine possible factors that could influence the dropout of users.

Methods

Protocol and Registration

We conducted this systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline, 2020 [24]. Before the start of the study, the review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO; CRD42022329669).

Data Sources and Search Strategy

Two researchers (J-CH-R and CG-M) performed an independent electronic search in PubMed, Scopus, Web of Science, Cochrane Library, and CINAHL. The search included all records from the inception of the databases up to July 10, 2022. Search terms for digital strategies (*"virtual," "online," "web-based," "internet-based," "digital," "e-Health," "m-Health," "App,"* and *"mApp"*), skin cancer (*"melanoma," "cutaneous melanoma," "malignant melanoma,"* and *"skin cancer"*), prevention (*"prevention"* and *"sun protection"*), and risk factors (*"tanned," "sunburn,"* and *"UV exposure"*) were employed. These were combined using the Boolean operators "AND" and "OR." Details of the search strategy can be found in Appendix S1 in Multimedia Appendix 1.

Eligibility Criteria and Outcomes of Interest

We developed the eligibility criteria following the PICOS model (ie, patient, intervention, comparison, outcome, and study design) shown in Table 1.

Table 1. Eligibility criteria based on the PICOS[a] model.

PICOS model	Inclusion criteria	Exclusion criteria
Population	Participants free of skin cancer during the study period	Participants with skin diseases during the study period
Intervention	Digital prevention strategies	Preventions approaches not focused on digital strategies
Comparator	Any type of comparator	Digital prevention strategies as comparator
Outcomes	Number of participants who dropout during the study period	Studies in which the dropout number was not reported, or indirect calculation was not allowed
Study design	Randomized controlled trials written in English	Any other type of study design

[a]PICOS: patient, intervention, comparison, outcome, and study design.

Data Management and Selection Process

To manage data, Mendeley Desktop (version 1.19.8; Elsevier) was used to detect duplicates and carry out the screening process. Two independent researchers (J-CH-R and CG-M) screened records by title and abstract, and later performed a complete read of the studies to select those that met the mentioned criteria. Any disagreement was deliberated with a third researcher, J-JP-R.

Assessment of Methodological Quality

We assessed methodological quality and risk of bias using The Cochrane Risk of Bias tool version 2 (ROB-2) [25]. This tool is composed of the following five domains: bias from randomization process, intended intervention, missing outcome data, measurement of outcomes, and selection of the reported results. The overall judgment is classified as "low," "some concerns," or "high" risk of bias. We also conducted subgroup analysis to determine how dropout events could be affected by the level of methodological quality and methodological threats such as blinding.

Data Extraction and Qualitative Synthesis

The following data were extracted from the RCTs included in the systematic review: authors or year and country, study population, recruited sample, analyzed sample, sex, experimental and control intervention, dropout rate, reasons for dropouts, and length of intervention. When the number or rate of dropouts was not directly provided in the manuscripts, both were calculated.

Quantitative Assessment of Data

A dropout was considered when a participant did not complete the intervention or follow-up period, after the randomization process. For studies that included more than 2 groups of intervention, we separately analyzed the comparison groups two by two. Dropout data were extracted from the text of the randomized controlled trials provided in either a flowchart, in the description of participants, in the results sections, or in the discussion.

To analyze data, we used the free software R Studio version 4.1.1. (R foundation for Statistical Computing) metafor (version 3.0-2) [26], *meta* (version 5.1-1) [27], and *dmetar* (version 0.0.9000) [28] packages. The analysis consisted of overall and subgroups proportion and odds ratio (OR)–based meta-analyses and metaregression.

A random-effects model was employed in all meta-analyses considering possible heterogeneity between our selected RCTs. Furthermore, heterogeneity was assessed with I^2, with values exceeding 50% indicating large heterogeneity. The subgroup meta-analysis and metaregression was run when at least 3 arms of study were available.

The meta-analysis of proportions allowed us to calculate the overall pooled dropout rate with its 95% CI of all arms of the studies included in our review [29,30]. Additionally, a subgroup analysis was performed to calculate the pooled dropout rate for digital or comparator interventions and to determine which type of intervention resulted in the highest dropout rate. This analysis was complemented by an OR subgroup analysis ordered by digital or intervention comparator to determine whether the probability of losing the participants was greater in one group or another.

The OR meta-analysis evaluated whether the event (dropout) was more or less frequent in the digital or comparator intervention. When the OR was less than 1, dropouts were less likely in digital strategies. To assess the measure of effect on binary outcomes, the OR with a 95% CI was calculated, and the inverse variance method was used to adjust the pooled estimations to sparse data. The restricted maximum-likelihood estimator for τ^2 estimated the variance among RCTs [31]. When studies reported zero events in one or all groups of intervention, we added a 0.5 continuity correction to the meta-analyses so that these studies could contribute to the overall sample size of the review [32]. The OR meta-analyses were conducted and subsequently described in terms of absolute values. The results of the meta-analyses were displayed in forest plots.

A sensitivity analysis was carried out to detect how studies influenced the effect size. When a study was identified as an outlier based on the dropout variable, it was removed from the analysis. Furthermore, to confirm previous results, we performed an exploratory analysis using the L'Abbé, Baujat plot, Leave-One-Out meta-analysis, and influence plot.

A univariate metaregression analysis based on a random-effects model assessed the continuous variables of age, female percentage, male percentage, length of intervention in months, and sample size as covariates of the occurrence of dropouts. These predictors were selected to determine how the characteristics of the participants and interventions could influence dropouts [33]. Bubble plots were used to illustrate

how a covariate modified the effect size in the metaregression analysis.

Publication Bias Assessment

We examined the effects of small studies and publication bias based on the symmetry of the contour-enhanced funnel plot. The Harbord and Egger bias test were used to confirm the absence of asymmetry in the funnel plot (P>.05).

Results

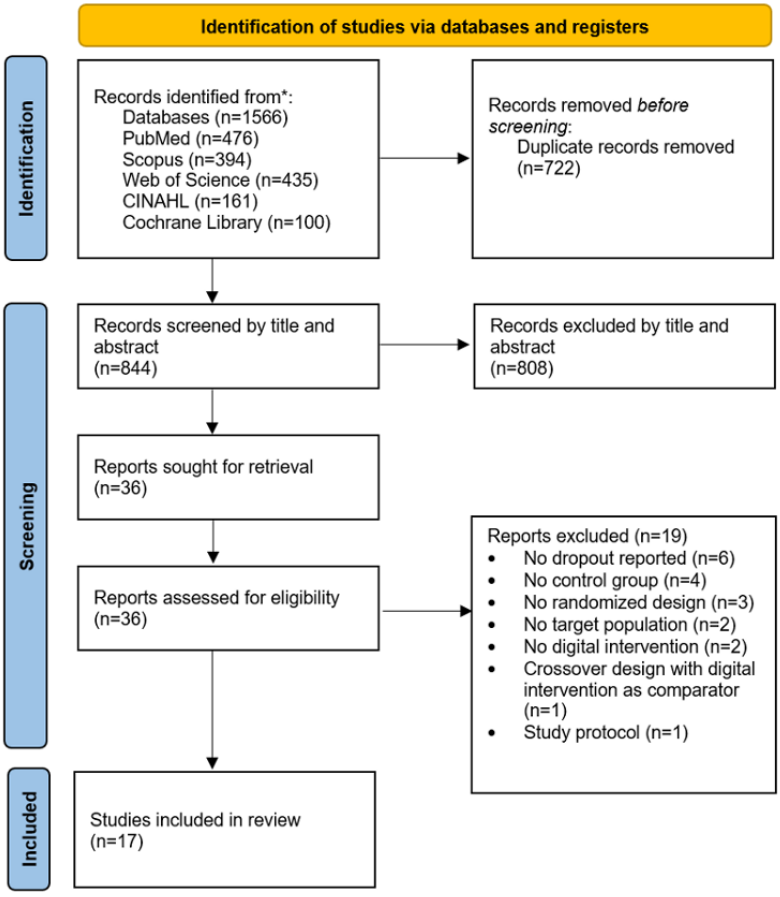
Study Selection and Methodological Quality Assessment

A total of 1566 studies were identified in the database search. After removing duplicates, the screening process, and complete reading of the records that met the eligibility criteria, 17 RCTs were finally included in the review [34-50]. The complete process is shown in Figure 1. Details of the excluded records are presented in Table S1 in Multimedia Appendix 1.

Regarding methodological quality, 14 (82%) of 17 RCTs showed "some concerns" based on the summary score of ROB-2. Moreover, 2 (12%) RCTs [44,47] showed a "low" risk of bias, and only 1 (6%) had a "high" risk of bias [49] (Figure S1 in Multimedia Appendix 1). The latter showed a "high" risk of bias because baseline differences between groups were observed.

Regarding the subgroup analyses, an analysis sorted by participants' blinding condition could not be performed because most of the studies were not blind or the blinding was not clearly specified. The subgroup meta-analysis sorted by the ROB-2 scores (Figure S2 in Multimedia Appendix 1) showed that a "low" overall score could indicate lower attrition in nondigital prevention strategies. However, due to the limited number of "low"-risk studies, the results should be interpreted with caution.

Figure 1. Flow diagram of trials selection based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines.



Study Design and Population Characteristics

A sample of 6593 healthy participants and people free of disease during the study period was analyzed. The age of the participants ranged from 12.6 to 54.3 years. The digital strategies used in the included RCTs were web-based interventions in 8 studies [35,40,42,43,46,48-50], 6 involved apps [36-38,44,45,47], 3 involved SMS [35,36,39], 2 involved video [34,41], and 1 involved a wearable device [47]. Conversely, the comparator groups involved no intervention in 11 studies [35,37,43-48,50] and active controls in 6 studies [34,36,38,39,41,49].

The total number of dropouts for all arms of the included studies was 1120, with 681 (60.80%) in experimental interventions and 439 (39.20%) in controls. The reason for the dropout of participants was reported as loss during follow-up in 9 of the 17 RCTs [34,37,38,40-43,47,49] and not answering the final questionnaire in 4 studies [39,44,45,50]; 2 studies did not report the reason for dropout [35,48]. The main characteristics of the studies are displayed in Table 2.

Table 2. Summary of the included studies in the systematic review.

Source	Population	Recruited or analyzed (n)	Percentage of sex, age (years), or mean (SD)	Experimental intervention	Comparator intervention	Dropout rate (%)	Reason for dropouts (EG/CG[a])	Length of intervention (months)
Armstrong et al, 2011 [34]; United States	English speakers aged >18 years	EG: 47/43; CG: 47/40; n=94	Female: 50%; male: 50%; 37.2 years	Online video addressing how sunscreen works to protect skin	Active (brochure)	EG: 8.5% (4/47); CG: 14.9% (7/47)	Lost to follow-up	3
Böttcher et al, 2019 [35]; Germany	Young organ transplant recipients	EG1: 44/39; EG2: 49/40; CG: 44/33; n=137	Female: 44.5%; male: 55.5%; 12.6 years	EG1: SMS text message providing sun protection advice; EG2: WBI[b] with sun protection training	No intervention (waitlist)	EG1: 11.4% (5/44); EG2: 18.4% (9/49); CG: 25.0% (11/44)	N/R[c]	12
Bowen et al, 2019 [36]; United States	First-degree relatives of melanoma cases	EG: 157/141; CG: 156/137; n=313	Female: 63.6%; male: 36.4%; 51.3 years	WBI with weekly messages of melanoma prevention behaviors	No intervention (waitlist)	EG: 10.2% (16/157); CG: 12.2% (19/156)	Lost to follow-up	12
Brinker et al, 2020 [37]; Brazil	Secondary school pupils	EG: 734/734; CG: 839/839; n=1573	Female: 51.6%; male: 48.4%; 15.9 (SD 1.3) years	App that modifies a selfie according to different levels of UV exposure for future 5 to 25 years based on individual skin type	No intervention	EG: 17.3% (127/734); CG: 6.20% (52/839)	Lost to follow-up	6
Buller et al, 2015 [38]; United States	Adults aged >18 years owning a smartphone	EG: 96/89; CG: 106/104; n=202	Female: 73.5%; male: 26.5%; 33.3 (SD 9.8) years	App giving feedback on sun protection and alerted users to apply or to reapply sunscreen and to get out of the sun	No intervention	EG: 7.3% (7/96); CG: 1.9% (2/106)	Lost to follow-up and survey not completed	3
Craciun et al, 2011 [39]; United Kingdom, Germany, Portugal, and Romania	Female volunteers	EG1: 74/74; EG2: 70/70; CG: 61/61; n=205	Male: 0%; female: 100%; 25.1 (SD 8.7) years	EG1: WBI volitional theory–based; EG2: WBI motivational theory–based	No intervention	0%	Not applied	1
Hacker et al, 2018 [40]; Australia	Young adults aged 18-35 years	EG1: 41/35; EG2: 42/36; CG: 41/36; n=124	Female: 65.8%; male: 31.5%; 25.8 years	EG1: app that displays the daily UV index and gives sun protection advice; EG2: wearable with UV dosimeter	No intervention	EG1: 14.6% (6/41); EG2: 14.3% (6/42); CG: 12.2% (5/41)	Lost to follow-up	3
Heckman et al, 2016 [41]; United States	Adults aged 18-25 years	EG1: 287/195; EG2: 338/205; CG: 340/229; n=965	Female: 66.1%; male: 33.9%; 21.8 (SD 2.2) years	EG1: WBI with a tailored intervention based on the Integrative Model of Behavioral Prediction; EG2: WBI with the Skin Cancer Foundation website	No intervention	EG1: 32.1% (92/287); EG2: 39.4% (133/338); CG: 32.7% (111/340)	N/R	3
Hillhouse at al, 2017 [42]; United States	Female adolescents	EG: 214/182; CG: 229/206; n=443	Female: 100%; male: 0%; 15.2 (SD 2.0) years	WBI to reduce IT[d] motivations	Active (placebo)	EG: 15.9% (32/214); CG: 10.1% (23/229)	Lost to follow-up	6
Manne et al, 2021 [43]; United States	Participants at increased risk for melanoma aged 18-89 years	EG: 56/43; CG: 60/56; n=116	Female: 69.8%; male: 30.2%; 51.1 (SD 15.2) years	WBI to improve SSE[e] and sun protection	No intervention	EG: 76.8% (13/56); CG: 93.3% (4/60)	Survey not completed	3

Source	Population	Recruited or analyzed (n)	Percentage of sex, age (years), or mean (SD)	Experimental intervention	Comparator intervention	Dropout rate (%)	Reason for dropouts (EG/CG[a])	Length of intervention (months)
Marek et al, 2018 [44]; United States	Adults aged ≥18 years	EG1: 18/18; EG2: 17/17; EG3: 17/17; CG: 17/17; n=69	Female: 61.1%; male: 38.9%; 54.3 (SD 13.9) years	EG1: app allowing total body photography; EG2: SMS to remind SSE; EG3: SMS+ accountability partner	Active (accountability partner)	0%	Not applied	6
Reilly et al, 2021 [45]; Scotland	Adults aged >18 years who survived stage 0-2C primary cutaneous melanoma	EG: 121/82; CG: 119/86; n=240	N/A[f]	App to encourage and improve SSE	No intervention	EG: 32.2% (39/121); CG: 27.7% (33/119)	Lost to follow-up	12
Robinson et al, 2016 [46]; United States	Kidney transplant recipients	EG: 84/78; CG: 86/83; n=170	Female 40.6%; male: 59.4%; 50.0 years	App with educational sun protection content	Active (usual education)	EG: 7.1% (6/84); CG: 3.5% (3/86)	Lost to follow-up	1.5
Robinson et al, 2021 [47]; United States	Female adults	EG: 494/390; CG: 495/414; n=989	Female: 100%; male: 0%; 47.0 years	SMS to remind SSE	Active (brochure)	EG: 21.1% (104/494); CG: 16.4% (81/495)	Survey not completed and discontinued intervention (EG)	3
Stapleton et al, 2015 [48]; United States	Female adults aged 18-25 years with IT in the past 12 months	EG: 94/74; CG: 93/85; n=186	Female: 100%; male: 0%; 19.8 (SD1.4) years	WBI with psychoeducational content to reduce IT	No intervention	EG: 8.5% (8/94); CG: 8.6% (8/93)	No response	1.5
Tsai et al, 2017 [49]; United States	Adults aged ≥18 years	EG: 71/42; CG: 72/34; n=143	Female: 74.1%; male: 25.9%; 42.3 years	Online melanoma video tutorial + brochure	Active (brochure)	EG: 40.8% (29/71); CG: 52.8% (38/72)	Lost to follow-up	1
Vuong et al, 2018 [50]; Australia	General practice patients	EG: 134/89; CG: 138/96; n=272	Female: 71.7%; male: 28.3%; 45.5 years	WBI with tailored melanoma risk assessment and prevention + usual education	Active (usual education)	EG: 33.9% (45/134); CG: 30.4% (42/138)	Lost to follow-up	1.5

[a]CG: comparator group; EG: experimental group.

[b]WBI: web-based intervention.

[c]N/R: not reported.

[d]IT: indoor tanning.

[e]SSE: skin self-examination.

[f]N/A: not applicable.

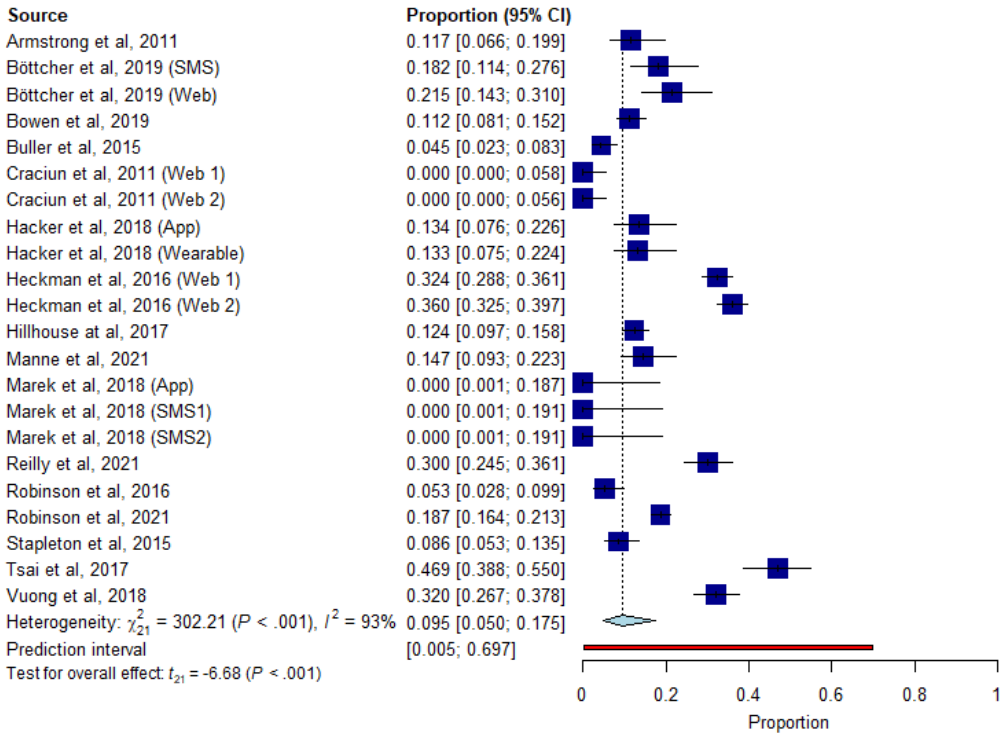
Sensitivity Analysis

The initial sensitivity analysis included a total of 23 arms from the randomized controlled trials of the review. After the sensitivity analysis, the study conducted by Brinker et al [44] was removed because it was identified as an outlier that influenced the effect size. The details of the sensitivity analysis are shown in Figures S3-S6 in Multimedia Appendix 1. Figure S7 in Multimedia Appendix 1 shows a funnel plot with absence of asymmetry, as confirmed by the Harbord test (P=.66) and Egger bias test (P=.69).

Meta-analysis of Proportions

The meta-analysis of proportions included 22 arms (k) of study and 2610 subjects among whom there were 419 dropouts. An overall pooled dropout rate of 9.5% (95% CI, 5.0-17.5) was calculated (Figure 2; [34-36,38-50]). In the subgroup meta-analysis, digital strategies showed a higher dropout rate of 11.6% (95% CI 6.8-19.0) compared to 10.0% (95% CI 5.5-17.7) in the comparators. These results are displayed in forest plots, respectively, in Figures S8 and S9 in Multimedia Appendix 1.

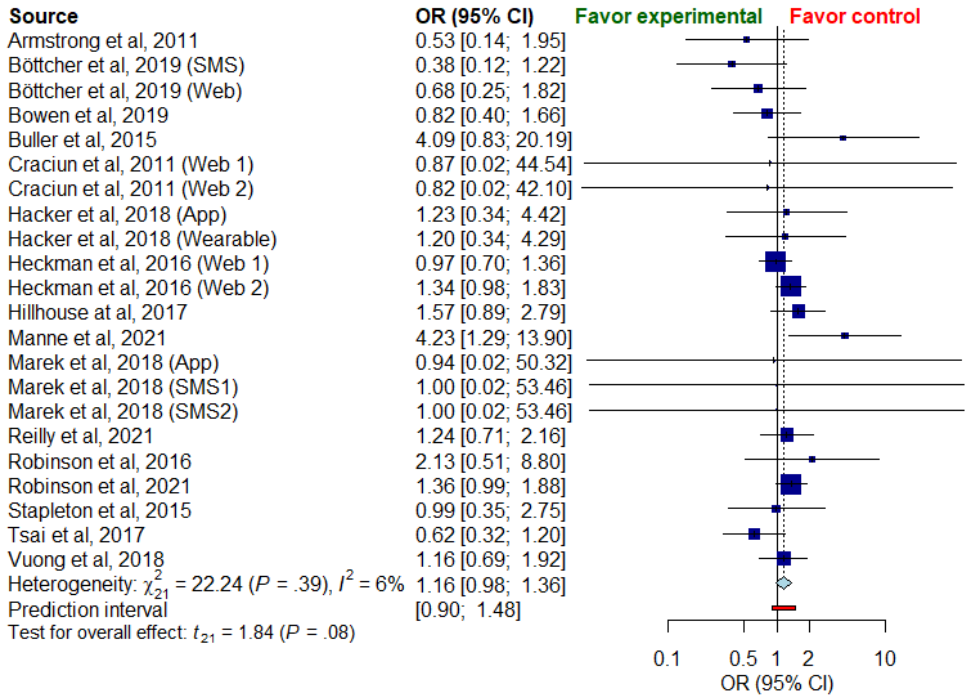
Figure 2. Forest plot of overall meta-analysis of proportions for all groups of studies.



Odds Ratio Meta-analysis

A slight trend for a higher number of dropouts was observed in digital strategies with an OR of 1.16 (95% CI 0.98-1.36), but there were no significant differences between the experimental and control approaches (*P*=.39). The I^2 was 6% (95% CI 0-38) indicating a lack of heterogeneity between the studies analyzed for the overall and subgroup meta-analysis (Figure 3; [34-36,38-50]).

Figure 3. Forest plot of overall odds ratio meta-analysis for all groups of studies.



We performed a meta-analysis of subgroups divided by the type of digital strategy and the comparison groups. Only the strategies that were analyzed in more than two RCTs were included in the OR meta-analysis. As Figure S10 in Multimedia Appendix 1 shows, none of the digital interventions assessed differed significantly in the number of dropouts compared with the comparator strategies. The OR score was 0.88 for SMS (95% CI, 0.30-2.53), 1.17 for web-based interventions (95% CI

0.94-1.47), and 1.44 for Apps (95% CI 0.88-2.35). Our findings in the comparator subgroup analysis showed no significant differences, even when comparing digital strategies with active control (OR 1.13; 95% CI 0.82-1.54) or no-intervention groups (OR 1.14; 95% CI 0.90-1.44; Figure S11 in Multimedia Appendix 1).

Metaregression

Univariate metaregression analysis (Table 3) for age, female percentage, male percentage, and length of intervention in months and sample size did not show any significant association with the effect size of the study. Metaregression bubble plots for these analyses are presented in Figures S12-S16 in Multimedia Appendix 1.

Table 3. Univariate metaregression analysis.

Covariate	Coefficient (95% CI)[a]	SE	*t* value	*P* value
Age	0.05 (–0.01 to 0.02)	0.24	–0.08	.53
Percentage of female	0.008 (–0.001 to 0.018)	0.004	1.78	.09
Percentage of male	–0.008 (–0.02 to 0.001)	0.005	–1.79	.09
Length of intervention (months)	–0.023 (–0.07 to 0.03)	0.023	–0.98	.34
Sample size	0.0004 (–0.0002 to 0.0009)	0.0002	1.45	.16

[a]According to the random-effects model.

Discussion

Principal Findings

This systematic review synthesizes information on the attrition of RCTs based on eHealth interventions for the prevention of skin cancer. Quantitative analysis evaluated the pooled dropout rate and dropout OR, in addition to moderators that could influence the dropout of subjects in the meta-analyzed RCTs. Although the digital strategies employed within studies used different platforms or devices, all of them were focused on skin cancer prevention and were supervised by expert dermatologists.

The meta-analysis of proportions showed a pooled dropout rate of 9.5%, with a dropout rate of 11.6% and 10.0% for the eHealth interventions and comparators, respectively. These results are in line with the findings by Walters et al [51], who reviewed the retention in RCTs of health technology programs in the United Kingdom. This review established that there was a dropout rate of up to 11% in a significant proportion of RCTs. Dropout rates of 5% are likely to introduce bias, while if 20% is exceeded, this could affect the validity of the study due to insufficiency during data analysis [52,53]. No background research was found performing similar analyses in the dermatology literature, so the comparison of rates was not viable.

Eysenbach et al [20] hypothesized that the nature of digital strategies tends to a higher loss of participants, a phenomenon called the “Law of attrition.” Although a slightly higher dropout rate was observed in digital strategies compared with comparator groups in our proportion and OR meta-analysis, the difference was not significant. Our findings refute the “Law of attrition” in those studies that aim to prevent skin cancer through these innovative interventions.

Previous systematic reviews, such as Bevens et al [54], focused on the analysis of attrition of digital strategies in people with multiple sclerosis and found no significant differences between dropout rates in participants allocated to digital or control interventions. Although our findings are in line with these previous findings, the target population and research conditions differed from ours, so comparison of findings are difficult.

As in the overall OR meta-analysis, the subgroup meta-analysis sorted by type of digital strategy and comparators found no significant differences in dropout rate. Only SMS text messaging presented a lower odd of dropout compared with other digital interventions, but without statistical significance. Reminder-based interventions such as SMS seem to promote adherence in chronic conditions, but further research is still needed [55]. It is noteworthy that the dropout rates in participants allocated to no intervention showed losses similar to digital strategies, reflecting the prior expectation that they could be affected by nonexperimental factors and the loss of perspective of therapeutics goals [56].

Our metaregression found that none of the covariates moderated the interventions’ effect size. Nonetheless, Torous et al [17] obtained higher dropout rates in studies with larger sample sizes that used apps for depressive symptoms, possibly related to a lower rate of individual follow-up and feedback from subjects.

In addition to the moderator analysis, assessment of the reasons for dropping out could be a way to identify barriers to reduce attrition in future RCTs. However, the lack of transparency and homogeneity in reporting reasons for participants’ dropout in the studies included in this review made the aforementioned task challenging. The main reported cause of attrition in our RCTs was loss to follow-up, but this aspect did not show the real reason for the loss of participants.

Research Implications

As previously mentioned, dropout could threaten internal or external validity in studies. We recommend that researchers use our overall pooled dropout rate to calculate the sample size of future trials, avoiding possible threats. The overrecruitment of 10.1% in the sample size of RCTs may be a suitable way to overcome external validity risks [57,58].

Although our OR meta-analysis showed no differences in attrition between digital strategies and comparator interventions, in order to obtain conclusive results that can be turned into daily

clinical practice, we point out the need for further research with head-to-head comparison between digital and conventional interventions (eg, education programs or brochures) for the prevention of skin cancer [59]. Dropout rates have previously been directly related to the acceptability and feasibility of the intervention [60,61].

Given the scarce information and lack of transparency provided by studies when reporting the number of and reasons for dropouts, a deep change in the research framework is needed. To overcome this obstacle, relevant guidelines such as Consolidated Standards of Reporting Trials report the need to detail the reasons and the number of participants lost during the study period [62,63]. Accurately following these guidelines would pave the way for researchers to find suitable dropout prevention plans. Previous literature, based on user experience with digital strategies, indicates that reliability, lack of technological education, lack of satisfaction with intervention, and sparse human feedback seem to be the main barriers to their use [63-65]. We encourage future researchers who aim to develop a digital strategy or perform RCT protocols to implement solutions to the mentioned barriers such as gamification, tailored and customizable e-interventions, personalized feedback, or programmed reminders (eg, mail and SMS). The gamification principles of meaningful purpose, meaningful choice, supporting player archetypes, feedback, and visibility proposed by Floryan et al [66] could enhance the user experience and engagement within digital health interventions. Gamification could increase motivation, reinforce learning objectives, and increase enjoyment and positive experiences in dermatological education and prevention approaches [9]. Likewise, programmed reminders are an effective way to promote prevention habits, highlighted by the use of text messages in dermatology [64]. Reminders associated with professional supervision have shown even greater results in prevention programs [67].

Given that RCTs are the first step required to translate research results into clinical settings, success in decreasing the number of participants dropping out within the research context could improve long-term engagement in digital programs for the prevention of skin cancer.

Strengths and Limitations

This review has several strengths. Our study provided an initial analysis of the dropout from RCTs to prevent skin cancer through digital strategies. Our computed rates could help calculate sample sizes in future studies. We performed a sensitivity analysis that helped us detect outliers and confirm the absence of publication bias. Moreover, the subgroups and metaregression analyses allowed us to understand how loss of participants could be modified by different predictors.

The main limitation of our review is that potential literature from other databases with non-English records could have been missing. Furthermore, our outcomes may have been conditioned by the heterogeneity of the experimental interventions in the included studies. Some of the studies compared digital strategies with no intervention, so we cannot assert that dropouts from these groups could be related to external factors. Evidence from the subgroup meta-analysis sorted by an active comparator group should be interpreted with caution because of the low number of analyzed studies; further research is needed to obtain strong evidence. We were unable to propose tailored advice to improve retention for this kind of RCT owing to the sparse information on reasons for dropout provided by the authors.

Conclusions

This systematic review and meta-analysis calculated an overall pooled participant dropout rate of 9.5% (95% CI 5.0-17.5), which should be considered in the calculation of sample size in RCTs aimed at preventing skin cancer using digital health interventions. Although a slightly higher pooled dropout rate was recorded for digital strategies, the OR-based meta-analysis did not show significant differences against the comparator groups. Our meta-analyses of subgroups sorted by digital and comparator interventions did not present significant statistical differences. Age, sex, length of the intervention, and sample size did not modify the effect size, so they were not moderators of dropout. We highlight the need to follow the guidelines and standardize reporting of the number of and reasons for participants' dropout because this will be the only effective way to design a successful plan to reduce the loss of participants in studies that analyze digital approaches to prevent skin cancer.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental material.
[DOCX File , 9184 KB - jmir_v24i12e42397_app1.docx]

References

1. Giesey R, Mostow E, Lloyd J. Digital strategies for dermatology patient education. Cutis 2018 Mar;101(3):209-210. [Medline: 29718025]
2. Arnold M, Singh D, Laversanne M, Vignat J, Vaccarella S, Meheus F, et al. Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040. JAMA Dermatol 2022 May 01;158(5):495-503. [doi: 10.1001/jamadermatol.2022.0160] [Medline: 35353115]

3. Whiteman DC, Green AC, Olsen CM. The Growing Burden of Invasive Melanoma: Projections of Incidence Rates and Numbers of New Cases in Six Susceptible Populations through 2031. J Invest Dermatol 2016 Jun;136(6):1161-1171 [FREE Full text] [doi: 10.1016/j.jid.2016.01.035] [Medline: 26902923]
4. Urban K, Mehrmal S, Uppal P, Giesey RL, Delost GR. The global burden of skin cancer: A longitudinal analysis from the Global Burden of Disease Study, 1990-2017. JAAD Int 2021 Mar;2:98-108 [FREE Full text] [doi: 10.1016/j.jdin.2020.10.013] [Medline: 34409358]
5. Zhang W, Zeng W, Jiang A, He Z, Shen X, Dong X, et al. Global, regional and national incidence, mortality and disability-adjusted life-years of skin cancers and trend analysis from 1990 to 2019: An analysis of the Global Burden of Disease Study 2019. Cancer Med 2021 Jul 09;10(14):4905-4922 [FREE Full text] [doi: 10.1002/cam4.4046] [Medline: 34105887]
6. Niu Z, Bhurosy T, Heckman CJ. Digital interventions for promoting sun protection and skin self-examination behaviors: A systematic review. Prev Med Rep 2022 Apr;26:101709 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101709] [Medline: 35529530]
7. Classification of self-care interventions for health: a shared language to describe the uses of self-care interventions. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/350480/9789240039469-eng.pdf?sequence=1&isAllowed=y [accessed 2022-12-01]
8. Khan N, Marvel FA, Wang J, Martin SS. Digital Health Technologies to Promote Lifestyle Change and Adherence. Curr Treat Options Cardiovasc Med 2017 Aug;19(8):60. [doi: 10.1007/s11936-017-0560-4] [Medline: 28647844]
9. Szeto MD, Strock D, Anderson J, Sivesind TE, Vorwald VM, Rietcheck HR, et al. Gamification and Game-Based Strategies for Dermatology Education: Narrative Review. JMIR Dermatol 2021 Aug 30;4(2):e30325. [doi: 10.2196/30325]
10. Ariens LF, Schussler-Raymakers FM, Frima C, Flinterman A, Hamminga E, Arents BW, et al. Barriers and Facilitators to eHealth Use in Daily Practice: Perspectives of Patients and Professionals in Dermatology. J Med Internet Res 2017 Sep 05;19(9):e300 [FREE Full text] [doi: 10.2196/jmir.7512] [Medline: 28874336]
11. Law LM, Wason JMS. Design of telehealth trials--introducing adaptive approaches. Int J Med Inform 2014 Dec;83(12):870-880 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.09.002] [Medline: 25293533]
12. Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: a systematic review of reviews. Int J Med Inform 2012 Jan;81(1):1-11 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.10.009] [Medline: 22104370]
13. Waseh S, Dicker AP. Telemedicine Training in Undergraduate Medical Education: Mixed-Methods Review. JMIR Med Educ 2019 Apr 08;5(1):e12515 [FREE Full text] [doi: 10.2196/12515] [Medline: 30958269]
14. 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]
15. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The Effectiveness of Prompts to Promote Engagement With Digital Interventions: A Systematic Review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
16. Olivier CB, Middleton SK, Purington N, Shashidhar S, Hereford J, Mahaffey KW, et al. Why digital health trials can fail: Lessons learned from a randomized trial of health coaching and virtual cardiac rehabilitation. Cardiovasc Digit Health J 2021 Apr;2(2):101-108 [FREE Full text] [doi: 10.1016/j.cvdhj.2021.01.003] [Medline: 35265897]
17. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
18. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
19. Peterson JC, Pirraglia PA, Wells MT, Charlson ME. Attrition in longitudinal randomized controlled trials: home visits make a difference. BMC Med Res Methodol 2012 Nov 23;12(1):178 [FREE Full text] [doi: 10.1186/1471-2288-12-178] [Medline: 23176384]
20. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
21. Skarin F, Wästlund E, Gustafsson H. Maintaining or Losing Intervention-Induced Health-Related Behavior Change. A Mixed Methods Field Study. Front Psychol 2021;12:688192 [FREE Full text] [doi: 10.3389/fpsyg.2021.688192] [Medline: 34220651]
22. Morenz AM, Wescott S, Mostaghimi A, Sequist TD, Tobey M. Evaluation of Barriers to Telehealth Programs and Dermatological Care for American Indian Individuals in Rural Communities. JAMA Dermatol 2019 Aug 01;155(8):899-905 [FREE Full text] [doi: 10.1001/jamadermatol.2019.0872] [Medline: 31215975]
23. Krist AH, Tong ST, Aycock RA, Longo DR. Engaging Patients in Decision-Making and Behavior Change to Promote Prevention. Stud Health Technol Inform 2017;240:284-302 [FREE Full text] [Medline: 28972524]

24. 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]
25. 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 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
26. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J. Stat. Soft 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
27. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019 Nov;22(4):153-160. [doi: 10.1136/ebmental-2019-300117] [Medline: 31563865]
28. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
29. Schwarzer G, Chemaitelly H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. Res Synth Methods 2019 Sep 23;10(3):476-483 [FREE Full text] [doi: 10.1002/jrsm.1348] [Medline: 30945438]
30. Barker TH, Migliavaca CB, Stein C, Colpani V, Falavigna M, Aromataris E, et al. Conducting proportional meta-analysis in different types of systematic reviews: a guide for synthesisers of evidence. BMC Med Res Methodol 2021 Sep 20;21(1):189 [FREE Full text] [doi: 10.1186/s12874-021-01381-z] [Medline: 34544368]
31. Viechtbauer W. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. Journal of Educational and Behavioral Statistics 2016 Nov 23;30(3):261-293. [doi: 10.3102/10769986030003261]
32. Gart JJ, Zweifel JR. On the Bias of Various Estimators of the Logit and Its Variance with Application to Quantal Bioassay. Biometrika 1967 Jun;54(1/2):181. [doi: 10.2307/2333861]
33. Carayol M, Delpierre C, Bernard P, Ninot G. Population-, intervention- and methodology-related characteristics of clinical trials impact exercise efficacy during adjuvant therapy for breast cancer: a meta-regression analysis. Psychooncology 2015 Jul;24(7):737-747. [doi: 10.1002/pon.3727] [Medline: 25483860]
34. Armstrong AW, Idriss NZ, Kim RH. Effects of video-based, online education on behavioral and knowledge outcomes in sunscreen use: a randomized controlled trial. Patient Educ Couns 2011 May;83(2):273-277. [doi: 10.1016/j.pec.2010.04.033] [Medline: 20570081]
35. Böttcher S, Buck C, Zeeb H, Laschewski G, Hauer C, Wagner G, et al. Randomised controlled trial to evaluate the influence of mHealth and eHealth skin cancer prevention education among young organ transplant recipients: the HIPPOlino intervention study. BMJ Open 2019 Dec 16;9(12):e028842 [FREE Full text] [doi: 10.1136/bmjopen-2018-028842] [Medline: 31848158]
36. Bowen DJ, Hay J, Meischke H, Mayer JA, Harris-Wai J, Burke W. Randomized trial of a web-based survivor intervention on melanoma prevention behaviors of first-degree relatives. Cancer Causes Control 2019 Mar 27;30(3):225-233 [FREE Full text] [doi: 10.1007/s10552-018-1096-y] [Medline: 30483971]
37. Brinker TJ, Faria BL, de Faria OM, Klode J, Schadendorf D, Utikal JS, et al. Effect of a Face-Aging Mobile App-Based Intervention on Skin Cancer Protection Behavior in Secondary Schools in Brazil: A Cluster-Randomized Clinical Trial. JAMA Dermatol 2020 Jul 01;156(7):737-745 [FREE Full text] [doi: 10.1001/jamadermatol.2020.0511] [Medline: 32374352]
38. Buller DB, Berwick M, Lantz K, Buller MK, Shane J, Kane I, et al. Evaluation of immediate and 12-week effects of a smartphone sun-safety mobile application: a randomized clinical trial. JAMA Dermatol 2015 May;151(5):505-512 [FREE Full text] [doi: 10.1001/jamadermatol.2014.3894] [Medline: 25629819]
39. Craciun C, Schüz N, Lippke S, Schwarzer R. Facilitating sunscreen use in women by a theory-based online intervention: a randomized controlled trial. J Health Psychol 2012 Mar;17(2):207-216. [doi: 10.1177/1359105311414955] [Medline: 21752862]
40. Hacker E, Horsham C, Vagenas D, Jones L, Lowe J, Janda M. A Mobile Technology Intervention With Ultraviolet Radiation Dosimeters and Smartphone Apps for Skin Cancer Prevention in Young Adults: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Nov 28;6(11):e199 [FREE Full text] [doi: 10.2196/mhealth.9854] [Medline: 30487115]
41. Heckman CJ, Darlow SD, Ritterband LM, Handorf EA, Manne SL. Efficacy of an Intervention to Alter Skin Cancer Risk Behaviors in Young Adults. Am J Prev Med 2016 Jul;51(1):1-11 [FREE Full text] [doi: 10.1016/j.amepre.2015.11.008] [Medline: 26810358]
42. Hillhouse J, Turrisi R, Scaglione NM, Cleveland MJ, Baker K, Florence LC. A Web-Based Intervention to Reduce Indoor Tanning Motivations in Adolescents: a Randomized Controlled Trial. Prev Sci 2017 Feb;18(2):131-140 [FREE Full text] [doi: 10.1007/s11121-016-0698-4] [Medline: 27549602]
43. Manne SL, Marchetti MA, Kashy DA, Heckman CJ, Ritterband LM, Thorndike FP, et al. mySmartCheck, a Digital Intervention to Promote Skin Self-examination Among Individuals Diagnosed With or at Risk for Melanoma: A Randomized Clinical Trial. Ann Behav Med 2022 Aug 02;56(8):791-803 [FREE Full text] [doi: 10.1093/abm/kaab090] [Medline: 34637495]

44. Marek AJ, Chu EY, Ming ME, Khan ZA, Kovarik CL. Piloting the Use of Smartphones, Reminders, and Accountability Partners to Promote Skin Self-Examinations in Patients with Total Body Photography: A Randomized Controlled Trial. Am J Clin Dermatol 2018 Oct;19(5):779-785 [FREE Full text] [doi: 10.1007/s40257-018-0372-7] [Medline: 30062632]
45. Reilly F, Contstable L, Brant W, Rahman K, Durrani A, Burrows N, et al. Achieving integrated self-directed Cancer aftercare (ASICA) for melanoma: how a digital intervention to support total skin self-examination was used by people treated for cutaneous melanoma. BMC Cancer 2021 Nov 13;21(1):1217 [FREE Full text] [doi: 10.1186/s12885-021-08959-2] [Medline: 34774015]
46. Robinson JK, Friedewald JJ, Desai A, Gordon EJ. A Randomized Controlled Trial of a Mobile Medical App for Kidney Transplant Recipients. Transplantation Direct 2016;2(1):e51. [doi: 10.1097/txd.0000000000000561]
47. Robinson JK, Wahood S, Ly S, Kirk J, Yoon J, Sterritt J, et al. Melanoma detection by skin self-examination targeting at-risk women: A randomized controlled trial with telemedicine support for concerning moles. Prev Med Rep 2021 Dec;24:101532 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101532] [Medline: 34976609]
48. Stapleton JL, Manne SL, Darabos K, Greene K, Ray AE, Turner AL, et al. Randomized controlled trial of a web-based indoor tanning intervention: Acceptability and preliminary outcomes. Health Psychol 2015 Dec;34S:1278-1285 [FREE Full text] [doi: 10.1037/hea0000254] [Medline: 26651469]
49. Tsai S, Frank SH, Bordeaux JS. Improving Sun-Protective Behaviors and Self-Skin Examinations Among African Americans: A Randomized Controlled Trial. Dermatol Surg 2018;44(4):512-518. [doi: 10.1097/dss.0000000000001366]
50. Vuong K, Armstrong BK, McGeechan K, Cust AE. Personalized melanoma risk assessments and tailored prevention advice: a pragmatic randomized controlled trial in Australian general practice. Fam Pract 2019 Mar 20;36(2):237-246. [doi: 10.1093/fampra/cmy040] [Medline: 29800131]
51. Walters SJ, Bonacho Dos Anjos Henriques-Cadby I, Bortolami O, Flight L, Hind D, Jacques RM, et al. Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom Health Technology Assessment Programme. BMJ Open 2017 Mar 20;7(3):e015276 [FREE Full text] [doi: 10.1136/bmjopen-2016-015276] [Medline: 28320800]
52. Fewtrell MS, Kennedy K, Singhal A, Martin RM, Ness A, Hadders-Algra M, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch Dis Child 2008 Jun;93(6):458-461 [FREE Full text] [doi: 10.1136/adc.2007.127316] [Medline: 18495909]
53. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. The Lancet 2002 Mar;359(9308):781-785. [doi: 10.1016/s0140-6736(02)07882-0]
54. Bevens W, Weiland T, Gray K, Jelinek G, Neate S, Simpson-Yap S. Attrition Within Digital Health Interventions for People With Multiple Sclerosis: Systematic Review and Meta-analysis. J Med Internet Res 2022 Feb 09;24(2):e27735 [FREE Full text] [doi: 10.2196/27735] [Medline: 35138262]
55. DeKoekkoek T, Given B, Given CW, Ridenour K, Schueller M, Spoelstra SL. mHealth SMS text messaging interventions and to promote medication adherence: an integrative review. J Clin Nurs 2015 Oct;24(19-20):2722-2735. [doi: 10.1111/jocn.12918] [Medline: 26216256]
56. Piolanti A, Foran HM. Efficacy of Interventions to Prevent Physical and Sexual Dating Violence Among Adolescents: A Systematic Review and Meta-analysis. JAMA Pediatr 2022 Feb 01;176(2):142-149 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.4829] [Medline: 34842911]
57. Cooper CL, Whitehead A, Pottrill E, Julious SA, Walters SJ. Are pilot trials useful for predicting randomisation and attrition rates in definitive studies: A review of publicly funded trials. Clin Trials 2018 Apr 23;15(2):189-196 [FREE Full text] [doi: 10.1177/1740774517752113] [Medline: 29361833]
58. Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. BMJ Evid Based Med 2018 Feb 24;23(1):21-22. [doi: 10.1136/ebmed-2017-110883] [Medline: 29367321]
59. Sim WMB, Zeng MX, Rojas-Garcia A. The effectiveness of educational programmes in promoting sun protection among children under the age of 18: a systematic review and meta-analysis. J Eur Acad Dermatol Venereol 2021 Nov;35(11):2154-2165. [doi: 10.1111/jdv.17599] [Medline: 34379846]
60. Cuijpers P, Quero S, Noma H, Ciharova M, Miguel C, Karyotaki E, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry 2021 Jun;20(2):283-293 [FREE Full text] [doi: 10.1002/wps.20860] [Medline: 34002502]
61. Sreedaran P, Beniwal RP, Chari U, Smitha TS, Vidhya Shree SV, Gupta V, et al. A Randomized Controlled Trial to Assess Feasibility and Acceptability of Telephone-Based Psychosocial Interventions in Individuals who Attempted Suicide. Indian J Psychol Med 2021 Mar 20;43(2):144-149 [FREE Full text] [doi: 10.1177/0253717620939272] [Medline: 34376890]
62. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
63. Leslie M, Beatty L, Hulbert-Williams L, Pendrous R, Cartwright T, Jackson R, Finding My Way UK Trial Steering Group, et al. Web-Based Psychological Interventions for People Living With and Beyond Cancer: Meta-Review of What Works and What Does Not for Maximizing Recruitment, Engagement, and Efficacy. JMIR Cancer 2022 Jul 08;8(3):e36255 [FREE Full text] [doi: 10.2196/36255] [Medline: 35802418]

64. Choi J, Cho Y, Woo H. mHealth Approaches in Managing Skin Cancer: Systematic Review of Evidence-Based Research Using Integrative Mapping. JMIR Mhealth Uhealth 2018 Aug 02;6(8):e164 [FREE Full text] [doi: 10.2196/mhealth.8554] [Medline: 30072362]
65. Wilks CR, Yin Q, Zuromski KL. User Experience Affects Dropout from Internet-Delivered Dialectical Behavior Therapy. Telemed J E Health 2020 Jun 01;26(6):794-797 [FREE Full text] [doi: 10.1089/tmj.2019.0124] [Medline: 31502945]
66. Floryan M, Chow PI, Schueller SM, Ritterband LM. The Model of Gamification Principles for Digital Health Interventions: Evaluation of Validity and Potential Utility. J Med Internet Res 2020 Jun 10;22(6):e16506 [FREE Full text] [doi: 10.2196/16506] [Medline: 32519965]
67. Timpel P, Oswald S, Schwarz PEH, Harst L. Mapping the Evidence on the Effectiveness of Telemedicine Interventions in Diabetes, Dyslipidemia, and Hypertension: An Umbrella Review of Systematic Reviews and Meta-Analyses. J Med Internet Res 2020 Mar 18;22(3):e16791 [FREE Full text] [doi: 10.2196/16791] [Medline: 32186516]

Abbreviations

OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RB-2: The Cochrane Risk of Bias tool version 2
RCT: randomized clinical trial

Edited by G Eysenbach; submitted 03.09.22; peer-reviewed by M Gacto-Sánchez, L Harst; comments to author 28.10.22; revised version received 10.11.22; accepted 28.11.22; published 09.12.22.

Please cite as:
Hernández-Rodríguez JC, García-Muñoz C, Ortiz-Álvarez J, Saigí-Rubió F, Conejo-Mir J, Pereyra-Rodriguez JJ
Dropout Rate in Digital Health Interventions for the Prevention of Skin Cancer: Systematic Review, Meta-analysis, and Metaregression
J Med Internet Res 2022;24(12):e42397
URL: https://www.jmir.org/2022/12/e42397
doi:10.2196/42397
PMID:36485027

Review

The Effectiveness of Technology-Based Cardiopulmonary Resuscitation Training on the Skills and Knowledge of Adolescents: Systematic Review and Meta-analysis

Xiu Ming Amanda Lim[1], BSc; Wenxin Ariel Liao[1], BSc; Wenru Wang[2], BSc, MSc, PhD; Betsy Seah[2], BSc, PhD

[1]Woodlands Health, Singapore, Singapore

[2]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Betsy Seah, BSc, PhD
Alice Lee Centre for Nursing Studies
Yong Loo Lin School of Medicine
National University of Singapore
Clinical Research Centre
Blk MD11 #02-02 ,10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66013213
Email: nurseah@nus.edu.sg

Abstract

Background: Cardiopulmonary resuscitation (CPR) training for adolescents is a prominent strategy to increase the number of community first responders who can recognize cardiac arrest and initiate CPR. More schools are adopting technology-based CPR training modalities to reduce class time and reliance on instructor availability and increase their capacity for wider training dissemination. However, it remains unclear whether these technology-based modalities are comparable with standard training.

Objective: This study aimed to systematically review and perform meta-analyses to evaluate the effectiveness of technology-based CPR training on adolescents' CPR skills and knowledge.

Methods: Searches were conducted in PubMed, Embase, Cochrane Library, Ovid MEDLINE, CINAHL, PsycINFO, Education Resources Information Center, ProQuest Dissertations and Theses Global, and Scopus from inception to June 25, 2021. Eligible randomized controlled trials (RCTs) compared technology-based training with standard training for adolescents aged 12 to 18 years. Studies were appraised using the Cochrane risk-of-bias tool. Random-effects meta-analyses were performed using Review Manager (The Cochrane Collaboration). Subgroup analyses were conducted to explore sources of heterogeneity. Overall certainty of evidence was appraised using the Grading of Recommendations Assessment, Development, and Evaluation approach.

Results: Seventeen RCTs involving 5578 adolescents were included. Most of the studies had unclear risks of selection bias (9/17, 53%) and high risks of performance bias (16/17, 94%). Interventions that included instructor guidance increased the likelihood of adolescents checking the responsiveness of the person experiencing cardiac arrest (risk ratio 1.39, 95% CI 1.19-1.63) and calling the emergency medical services (risk ratio 1.11, 95% CI 1.00-1.24). Self-directed technology-based CPR training without instructor guidance was associated with poorer overall skill performance (Cohen *d*=–0.74, 95% CI –1.02 to –0.45). Training without hands-on practice increased mean compression rates (mean difference 9.38, 95% CI 5.75-13.01), whereas real-time feedback potentially yielded slower compression rates. Instructor-guided training with hands-on practice (Cohen *d*=0.45, 95% CI 0.13-0.78) and the use of computer programs or mobile apps (Cohen *d*=0.62, 95% CI 0.37-0.86) improved knowledge scores. However, certainty of evidence was very low.

Conclusions: Instructor-guided technology-based CPR training that includes hands-on practice and real-time feedback is noninferior to standard training in CPR skills and knowledge among adolescents. Our findings supported the use of technology-based components such as videos, computer programs, or mobile apps for self-directed theoretical instruction. However, instructor guidance, hands-on practice, and real-time feedback are still necessary components of training to achieve better learning outcomes for adolescents. Such a blended learning approach may reduce class time and reliance on instructor availability. Because of the high heterogeneity of the studies reviewed, the findings from this study should be interpreted with caution. More high-quality

RCTs with large sample sizes and follow-up data are needed. Finally, technology-based training can be considered a routine refresher training modality in schools for future research.

(J Med Internet Res 2022;24(12):e36423) doi:10.2196/36423

KEYWORDS

cardiac arrest; education; methods; first responders; resuscitation training; effectiveness; adolescents; schoolchildren

Introduction

Background

Out-of-hospital cardiac arrests (OHCAs) are associated with poor survival and neurological outcomes [1]. Prognoses are improved when bystanders promptly recognize cardiac arrest and initiate cardiopulmonary resuscitation (CPR) [2]. Of note, many laypeople lack the ability to identify OHCA or act appropriately. The American Heart Association and European Resuscitation Council advocate compulsory annual CPR training for individuals aged ≥12 years [3]. Generally, students receiving formal education (ie, middle school and high school students) are aged between 12 and 18 years. Introducing CPR training in schools can equip a large proportion of the country's population with fundamental lifesaving knowledge and skills. The early introduction of these lifesaving skills during one's developmental years not only promotes altruism but also increases one's willingness to help people experiencing OHCA [4]. Such school-based education sessions can prepare both students and teachers in responding to cardiac arrest incidents within schools and the community at large. Moreover, regular refresher training can be arranged easily in schools because most children attend formal education [5]. However, the lack of stringent guidelines gives schools full autonomy to conduct diverse training modalities, some of which are yet to be supported by empirical evidence [6].

Standard CPR training involves didactic face-to-face lessons, skill demonstrations by qualified instructors, and hands-on practice on manikins in small groups [7]. Although this modality has been regarded as the gold standard, there are often not enough qualified instructors for large-scale implementation in schools. Such training requires numerous manikins and equipment, which are costly [8]. Furthermore, standard training consumes substantial class time and impedes adoption by schools [9].

Technology-Based CPR Training

International resuscitation guidelines suggest the incorporation of technology into CPR training as alternatives to standard training [10,11]. Particularly in the age of the COVID-19 pandemic, technology-based CPR training has become increasingly prominent. These interventions are facilitated by digital technology, including instructional videos, web-based learning, computer programs, mobile apps, or advanced manikin software [12]. Many of them use self-directed learning to decrease reliance on the availability of qualified instructors [13]. Technology-based CPR training may also be cost-effective because fewer resources are required [14]. Finally, training duration is kept minimal, which complements hectic school curricula [15]. Hence, there is a tremendous potential for the proliferation of technology-based CPR training among adolescents.

Two systematic reviews were conducted on CPR training modalities for adolescents. Plant and Taylor [16] concluded that all modalities, including technology-based training, improved knowledge and skills. Reveruzzi et al [17] added that qualified instructors, videos, and hands-on practice improved training outcomes. Both reviews had broad eligibility criteria and no restrictions on study design. This contributed to heterogenous results that prevented pooling of training effects using meta-analyses. Other systematic reviews involving technology-based training focused on health care students, health care professionals, and adult laypeople [18-20]. The conclusions from these reviews cannot be generalized to adolescents because different teaching approaches might be necessary to cater to learners of different age groups [10]. This review used meta-analysis to synthesize evidence on the effectiveness of technology-based CPR training compared with standard training in improving the skills and knowledge of adolescents aged 12 to 18 years.

Methods

This study adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [21].

Search Strategy

Searches were conducted in PubMed, Embase, Cochrane Library, Ovid MEDLINE, CINAHL, PsycINFO, Education Resources Information Center, ProQuest Dissertations and Theses Global, and Scopus from inception to June 25, 2021. The search terms included *adolescent**, *schoolchild**, *student**, *cardiopulmonary resuscitation*, *basic life support*, *train**, and *teach**. Synonyms were combined with the Boolean operator *OR*. Population and intervention concepts were then combined, such as *adolescents* AND CPR AND *training*. Refer to Multimedia Appendix 1 for the database search strategies. ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform were searched for ongoing and unpublished trials. Hand searching of the *Resuscitation* journal was performed for articles published between January 2000 and June 2021. The reference lists of relevant trials and systematic reviews were screened to identify additional studies.

Inclusion Criteria

Randomized controlled trials (RCTs) were included if they met the following criteria: (1) participants were adolescents aged 12 to 18 years; (2) participants received CPR training that included technology-based components such as videos, web-based learning, computer programs, mobile apps, or

manikin software with real-time feedback; (3) technology-based CPR training was compared with standard CPR training (without the technology-based intervention component); and (4) the RCTs reported CPR skills or knowledge. CPR skills are defined as the ability to perform CPR techniques objectively measured via manikin software or as evaluated by instructors. Theoretical knowledge scores are measured by self-reported instruments, including multi-item questionnaires or multiple-choice–question tests (Table 1).

Table 1. Inclusion and exclusion criteria.

Variable	Inclusion criteria	Exclusion criteria
Study characteristics		
Study design	RCTs[a] and cluster RCTs	Nonrandomized studies, observational studies, qualitative studies, and reviews
Publication type	Full-text journal publications, conference proceedings, and unpublished dissertations or theses	Editorials and letters
Publication year	No limit	N/A[b]
Language	English only	Languages other than English
PICO[c] framework		
Population	Schoolchildren aged 12 to 18 years	Schoolchildren with physical disabilities that may affect their ability to perform CPR[d] (eg, those who are blind, deaf, or have a speech disability)
Intervention	CPR training with technology-based components, including videos, computer programs, mobile apps, and real-time audiovisual feedback	CPR training with popular songs only
Comparison	Standard resuscitation training without technology-based component	N/A
Outcomes		
Skill performance	Overall performance (cumulative score from skills checklist); components of cardiopulmonary resuscitation, including checking responsiveness, checking the airway and breathing, calling the EMS[e], compression depth, compression rate, correct hand position, correct compression:ventilation ratio, total compressions, correct ventilation, AED[f] pad placement, and use of AED	N/A
Knowledge	Theoretical knowledge scores	N/A

[a]RCT: randomized controlled trial.

[b]N/A: not applicable.

[c]PICO: Population, Intervention, Comparison, and Outcomes.

[d]CPR: cardiopulmonary resuscitation.

[e]EMS: emergency medical services.

[f]AED: automated external defibrillator.

Study Selection

All retrieved records were imported into EndNote X9 (Clarivate) for deduplication. Titles and abstracts of records were screened by 2 independent reviewers (AL and WX) for relevance. After removing irrelevant records, full texts of potential studies were independently assessed for eligibility. Discrepancies were resolved through discussion with a third reviewer (BS).

Data Extraction

AL and WX collected data independently using a standardized data extraction form. Extracted data included publication year, country, study design, setting, participants, sample size, interventions, comparators, outcome measures, and instruments. Posttraining and retention data were extracted, with retention defined as at least 4 weeks after training. Indicators of trial quality were also extracted; for example, attrition rate, intention to treat, and trial registration. Results of studies reported in >1 publication were extracted as 1 study. Authors were contacted when data were incomplete or unclear. Discrepancies in extracted data were resolved through discussion with BS.

Quality Appraisal

AL and WX performed quality appraisal independently for all included studies. Discrepancies were resolved through discussion with BS. The Cochrane risk-of-bias tool was used to appraise studies for risks of bias [22]. The Grading of Recommendations Assessment, Development, and Evaluation approach was used to appraise certainty of evidence for the

main outcomes [23]. Ratings were categorized as high, moderate, low, or very low.

Data Analysis

Meta-analyses were performed with Review Manager (The Cochrane Collaboration) and presented as forest plots where appropriate. The Mantel-Haenszel approach and risk ratio (RR) were selected for dichotomous outcomes, whereas the inverse-variance approach pooled mean differences (MDs) for continuous outcomes. Continuous outcomes measured using different scales were presented as standardized MDs or Cohen *d*. When mean and SD were not reported, values were estimated using median and IQR. Overall intervention effects were interpreted using the *Z* statistic, with level of significance set at $P \leq .05$.

Heterogeneity was evaluated using the Cochran *Q* test and I^2 statistic, with level of significance set at $P \leq .10$. A random-effects model was used because of variation in training characteristics [24]; for instance, the studies used different modes of technology-based instruction as well as different types of CPR instructors such as health care professionals, schoolteachers, or medical students, which may affect effect sizes across the studies. For meta-analyses with significant heterogeneity and with at least 6 comparisons, sensitivity or subgroup analyses were performed [25]. The subgroup analyses explored potential effect modifiers, including instructor guidance, hands-on practice, and training modalities. Sensitivity analyses were conducted when meta-analyses yielded significant heterogeneity that was attributable to an outlier study. Funnel plots were not performed because of the limited number of trials included in each meta-analysis. Where quantitative analysis could not be determined from the meta-analysis, findings were presented narratively.

Ethics Approval

The preparation of this paper did not involve primary research or data collection involving human participants; therefore, no institutional review board examination or approval was required.

Results

Study Characteristics

The search process is illustrated in Figure 1. Seventeen RCTs were included in this review. Table 2 summarizes the characteristics of the included studies. Of the 28 intervention arms included in this review, 4 (14%) [26-29] were excluded on account of irrelevance. Studies were conducted across 11 countries: Belgium, Canada, Iran, Italy, South Korea, the Netherlands, Spain, Sweden, Turkey, the United Kingdom, and the United States. A total of 5578 secondary school or high school students were recruited (sample sizes ranged from 62 to 1426). Of the 17 studies, 6 (35%) excluded students with prior or recent CPR training [26,27,30-33].

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

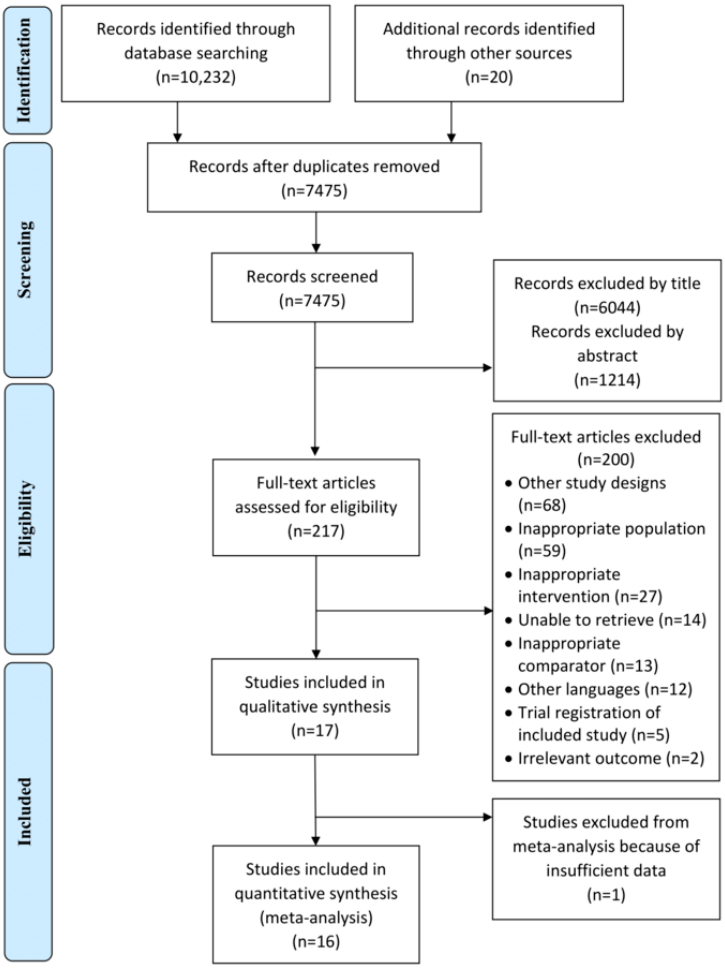


Table 2. Characteristics of the included studies.

Study authors, year; country	Study design	Sample size	Technology-based training	Standard training	Outcomes	Attrition (%)	ITT[a]; missing data management	Protocol; trial registration; funding
Beskind et al [28], 2016; United States	Three-arm cluster RCT[b,c]	I[d]: 58, C[e]: 54	Brief video	Standard instructor-led training	Skills	11.2	No; no	No; no; yes
Chamdawala et al [34], 2021; United States	Two-arm RCT	I: 110, C: 110	QCPR[f] real-time visual feedback	Standard instructor-led training	Skills and knowledge	13.6	No; no	No; no; yes
Cortegiani et al [35], 2017; Italy	Two-arm RCT	I: 60, C: 65	QCPR real-time visual feedback	Standard instructor-led training	Skills	13.2	No; no	Yes; yes; no
Cuijpers et al [36], 2011; Netherlands	Three-arm RCT	I1: 33, I2: 34, C: 37	I1: 1 hour e-learning+1 hour instructor-led training, I2: 1 hour e-learning	Standard instructor-led training	Skills	NR[g]	NR; NR	NR; NR; NR
Doucet et al [37], 2019; Belgium	Two-arm RCT	I: 83, C: 82	StartnHart app	Standard instructor-led training	Skills	0	N/A[h]	No; no; no
Han et al [38], 2021; Korea	Two-arm RCT	I: 31, C: 31	e-Learning+videoconferencing	Standard instructor-led training	Skills	0	N/A	No; no; yes
Iserbyt et al [31], 2014; Belgium	Two-arm RCT	I: 59, C: 69	Video instruction	Static picture instruction	Skills	7.2	No; no	No; no; no
Marchiori et al [39], 2012; Spain	Two-arm cluster RCT	I: 187, C: 144	Video game	Standard instructor-led training	Knowledge	3.8	No; no	No; no; yes
Morrison et al [40], 2012; Canada	Two-arm RCT	I: 140, C: 124	CPR[i] Anytime video self-instruction+instructor-led training for AED[j]	Standard instructor-led training	Skills	21	NR; NR	NR; NR; NR
Nord et al [41], 2017; Sweden	Two-arm cluster RCT	I: 645 or 208, C: 587 or 224	Web course+classroom-based instructor-facilitated training with app (static pictures) or video instruction	Classroom-based instructor-facilitated training with app (static pictures) or video instruction	Skills and knowledge	13.6	No; no	No; yes; yes
Norman [26], 1984; United States	Three-arm RCT[c]	I: 39, C: 39	Video instruction	Standard instructor-led training	Skills and knowledge	17.1	No; no	NR; NR; NR
Onan et al [42], 2019; Turkey	Three-arm cluster RCT	I1: 25, I2: 25, C: 25	I1: video instruction, I2: video instruction with real-time feedback	Standard instructor-led training (theory only)	Skills and knowledge	7.2	No; no	No; no; no
Otero-Agra et al [32], 2019; Spain	Four-arm cluster RCT	I1: 151, I2: 140, I3: 109, C: 89	I1: mandatory and graded team-based training with real-time feedback for competition, I2: mandatory and graded individual training with real-time feedback, I3: individual training with real-time feedback	Standard instructor-led training	Skills	0	N/A	No; no; no
Reder et al [29], 2006; United States	Four-arm cluster RCT[c]	I1: 213, I2: 170, C: 206	I1: interactive computer session, I2: interactive computer session with practice	Standard instructor-led training	Skills and knowledge	22.8	No; no	No; no; NR
Rezaei et al [33], 2013; Iran	Two-arm cluster RCT	I: 42, C: 42	Prerecorded video demonstration	Standard instructor-led training	Skills and knowledge	0	N/A	No; no; yes

Study authors, year; country	Study design	Sample size	Technology-based training	Standard training	Outcomes	Attrition (%)	ITT[a]; missing data management	Protocol; trial registration; funding
Van Raemdonck et al [27], 2014; Belgium	Four-arm RCT[c]	I1: 44, I2: 42, C: 43	I1: video instruction, I2: video instruction with low-cost manikin	Standard instructor-led training	Skills	66.3	No; no	No; no; yes
Yeung et al [30], 2017; United Kingdom	Three-arm cluster RCT	I1: 21, I2: 24, C: 19	I1: Lifesaver app, I2: Lifesaver app+standard instructor-led training	Standard instructor-led training	Skills	21	No; no	No; yes; yes

[a]ITT: intention to treat.

[b]RCT: randomized controlled trial.

[c]One comparison arm was not included in this review because of an irrelevant comparator.

[d]I: intervention.

[e]C: control.

[f]QCPR: quality cardiopulmonary resuscitation.

[g]NR: not reported.

[h]N/A: not applicable.

[i]CPR: cardiopulmonary resuscitation.

[j]AED: automated external defibrillator.

Descriptions of Interventions and Comparators

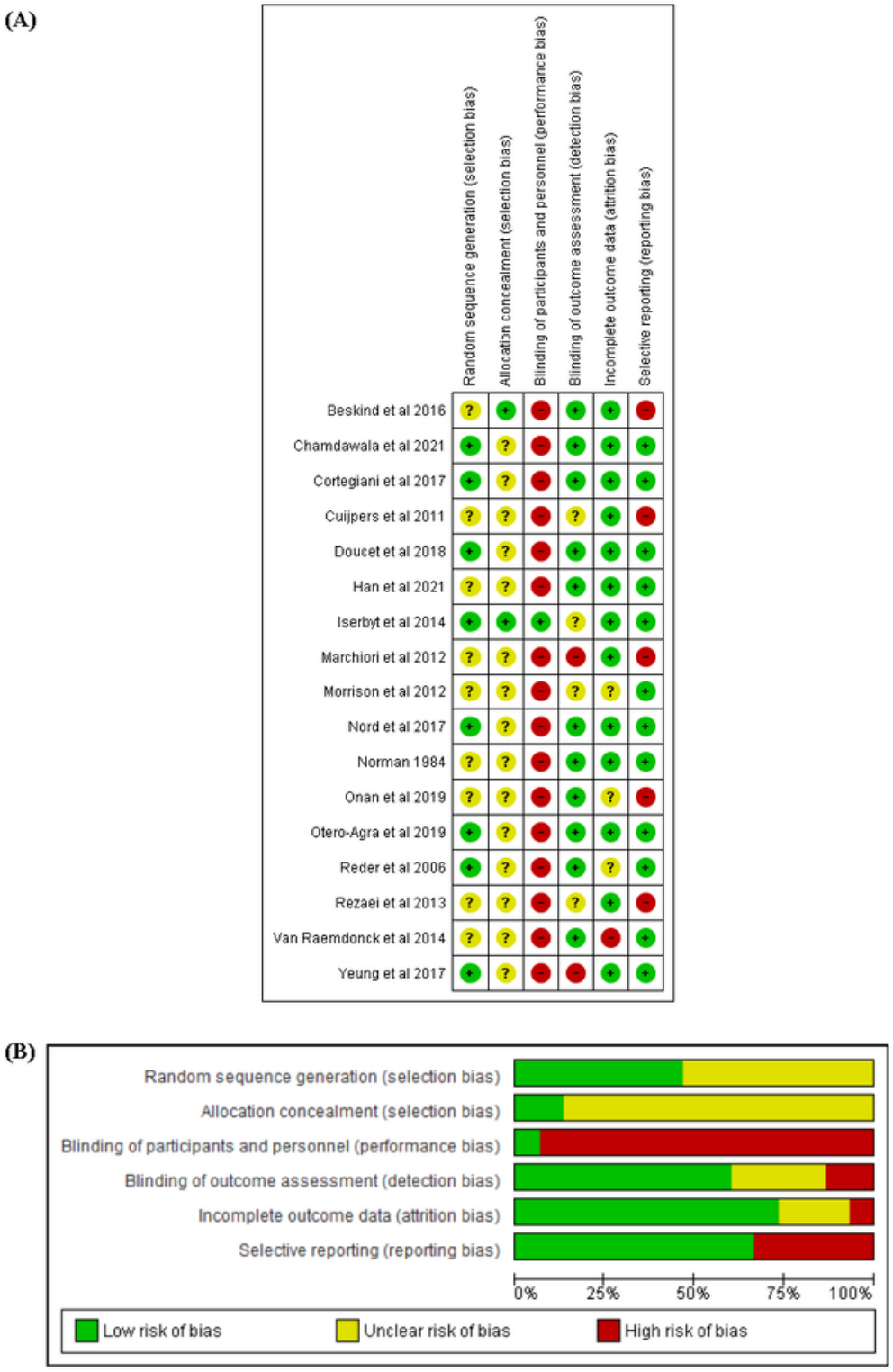
All studies adhered to national or international resuscitation guidelines, except for the study by Rezaei et al [33], which did not mention them. Only the study by Iserbyt et al [31] used a multimedia learning theory to guide the intervention. Training was either self-directed [27-31,33,37,39] or instructor guided [26,29,30,33-36,38,40-42]. Trained schoolteachers or medical students served as instructors or facilitators in 29% (5/17) of the studies [29,31,33,41,42]. The interventions comprised video instruction [26-28,31,33,40,42], computer programs or mobile apps [29,30,36-39,41,42], or real-time feedback [32,34,35]. Of the 24 intervention arms, 4 (17%) omitted hands-on practice on manikins [28-30,33]. All studies reported up to 2 training sessions over a span of 3 weeks, with each session lasting from 1.5 minutes to 4 hours. The length of follow-up ranged from 2 months [28,29] to 1 year [34].

The standard training used included face-to-face qualified-instructor–led demonstration, with hands-on practice on manikins. Other comparators included static pictures [31], classroom-based video instruction [31,34,38,41], or didactic teaching without hands-on practice [42]. Refer to Multimedia Appendix 2 for details.

Quality Assessment

Most (12/17, 71%) of the studies had overall moderate risk of bias (Figure 2). There was low risk of selection bias in 47% (8/17) of the RCTs because of adequate random sequence allocation and in 12% (2/17) because of allocation concealment. Because of the nature of CPR training, blinding of participants and personnel was not possible in all of the trials. Of the 17 RCTs, 7 (41%) had low risk of detection bias because of the blinding of outcome assessors and another 4 (24%) trials had all outcomes objectively measured through manikins, minimizing bias attributable to the lack of blinding. Of the 17 studies, 13 (76%) had low risk for attrition bias because of similar reasons for attrition across the groups or no missing data. Although only 12% (2/17) of the studies published protocols, 59% (10/17) reported all outcomes completely and were thus rated low risk for reporting bias. Certainty of evidence was appraised as *very low* for skills and knowledge using the Grading of Recommendations Assessment, Development, and Evaluation approach (Multimedia Appendix 3). Domains were downgraded because of high risks of bias, statistical and methodological heterogeneity, small sample sizes, and wide CIs.

Figure 2. Risk-of-bias (A) summary and (B) graph.



Effectiveness of Technology-Based Training on Overall Performance

Overview

Overall performance is the cumulative score from a skills checklist, with components presented in Table 3. The sole use of self-directed learning yielded poorer overall performance after the intervention. At 6 months, technology-based training potentially improved overall performance.

Table 3. Meta-analyses: cardiopulmonary resuscitation skill components.

Outcome and time point	Trials (N=16), n (%)	Arms (N=23), n (%)	Sample size (N)	Statistical approach	Effect estimate (95% CI)	Overall effect		I^2 (%)
						Z statistic	*P* value	
First link in chain of survival: early recognition and calling for help								
Checking responsiveness								
After training	5 (31)	6 (26)	884	M-H[a], random effects	1.16[b] (0.90 to 1.50)	1.14	.25	88
Checking airway								
After training	5 (31)	6 (26)	1370	M-H, random effects	0.93[b] (0.78 to 1.10)	0.85	.39	60
Retention	2 (13)	3 (13)	892	M-H, random effects	0.90[b] (0.72 to 1.13)	0.89	.37	25
Checking breathing								
After training	4 (25)	5 (22)	719	M-H, random effects	1.18[b] (0.92 to 1.50)	1.31	.19	68
Calling EMS[c]								
After training	6 (38)	7 (30)	996	M-H, random effects	1.10[b] (0.92 to 1.31)	1.07	.28	81
Retention	2 (13)	2 (9)	511	M-H, random effects	1.01[b] (0.92 to 1.10)	0.14	.89	0
Second link in chain of survival: early CPR[d]								
Overall compression quality (%)								
After training	3 (19)	4 (17)	824	IV[e], random effects	23.96[f] (19.84 to 28.09)	11.40	*<.001*[g]	0
Mean compression depth (mm)								
After training	10 (63)	13 (57)	1619	IV, random effects	1.16[f] (–2.49 to 4.82)	0.62	.53	95
Retention	6 (38)	8 (35)	1179	IV, random effects	0.73[f] (–3.07 to 4.52)	0.38	.71	94
Correct compression depth								
After training	6 (38)	8 (35)	1447	M-H, random effects	1.04[b] (0.90 to 1.21)	0.54	.59	43
Retention	2 (13)	3 (13)	528	M-H, random effects	0.76[b] (0.59 to 0.97)	2.17	*.03*	0
Mean compression rate (number of compressions per minute)								
After training	11 (69)	15 (65)	2107	IV, random effects	–3.25[f] (–7.57 to 1.07)	1.47	.14	88
Retention	6 (38)	8 (35)	1179	IV, random effects	–2.47[f] (–7.48 to 2.54)	0.97	.33	85
Correct compression rate								
After training	5 (31)	7 (30)	601	M-H, random effects	0.89[b] (0.75 to 1.07)	1.22	.22	38
Correct hand position								
After training	7 (44)	10 (43)	1617	M-H, random effects	0.93[b] (0.84 to 1.03)	1.38	.17	44
Retention	3 (19)	5 (22)	1021	M-H, random effects	0.86[b] (0.65 to 1.14)	1.06	.29	56
Correct ventilation								
After training	8 (50)	11 (48)	1680	M-H, random effects	0.86[b] (0.67 to 1.10)	1.23	.22	69
Retention	3 (19)	5 (22)	1056	M-H, random effects	0.64[b] (0.41 to 0.99)	2.00	*.05*	78
Correct compression:ventilation ratio								
After training	2 (13)	2 (9)	597	M-H, random effects	0.99[b] (0.87 to 1.13)	0.15	.88	34
Total compressions in 2 minutes								
After training	2 (13)	3 (13)	614	IV, random effects	–22.84[f] (–30.35 to –15.33)	5.96	*<.001*	0
Third link in chain of survival: early defibrillation								

Outcome and time point	Trials (N=16), n (%)	Arms (N=23), n (%)	Sample size (N)	Statistical approach	Effect estimate (95% CI)	Overall effect		I^2 (%)
						Z statistic	*P* value	
Correct AED [h] pad placement								
After training	2 (13)	3 (13)	729	M-H, random effects	0.94[b] (0.86 to 1.02)	1.58	.11	54
Correct use of AED								
After training	2 (13)	3 (13)	729	M-H, random effects	0.98[b] (0.94 to 1.01)	1.15	.25	68

[a]M-H: Mantel-Haenszel.

[b]RR: risk ratio.

[c]EMS: emergency medical services.

[d]CPR: cardiopulmonary resuscitation.

[e]IV: inverse variance.

[f]MD: mean difference.

[g]Results of significance are presented in italics.

[h]AED: automated external defibrillator.

Posttraining Performance

Of the 16 RCTs included in the meta-analyses for posttraining performance, 6 (38%; arms: 8/23, 35%) involving 1121 students reported overall performance scores from skills checklists [30,36,37,40-42]. Meta-analysis revealed high heterogeneity (I^2=89%; *P*<.001). Subgroup analyses revealed that self-directed learning yielded significantly poorer overall performance (Cohen *d*=–0.74, 95% CI –1.02 to –0.45; *P*<.001; Table 4).

Table 4. Subgroup analyses based on instructor guidance: overall performance scores.

Subgroup analyses	Comparisons (n)	Effect estimate (95% CI)	Subgroup effect		I^2 (%)	Subgroup differences	
			Z statistic	*P* value		I^2 (%)	*P* value
Self-directed learning	2	–0.74 (–1.02 to –0.45)	5.02	*<.001* [a]	0	92.8	*<.001*
Instructor-guided learning	6	0.28 (–0.17 to 0.73)	1.22	.22	88	N/A[b]	N/A

[a]Results of significance are presented in italics.

[b]N/A: not applicable.

Retention

Of the 16 RCTs, 3 (19%; arms: 4/23, 17%) involving 727 students reported overall performance at 6 months [30,40,41]. Only instructor-guided training involving a participative mobile app significantly improved performance retention [30]. Interventions that used video instruction, a supplementary web-based course, or self-directed learning with a participative mobile app yielded performance similar to that of standard training [30,40,41].

Effectiveness of Technology-Based Training on CPR Skill Components

Meta-analyses performed for CPR skill components are summarized in Table 3.

Posttraining Performance

Of the 16 RCTs, 3 (19%; arms: 4/23, 17%) involving 824 students reported overall compression quality calculated by manikin software (Figure 3) [32,34,35]. Technology-based training significantly improved compression quality (MD 23.96, 95% CI 19.84-28.09; *P*<.001; Table 3). With significant heterogeneity reported for the other CPR components, sensitivity or subgroup analyses were performed for instructor guidance, training components, and training modalities (Table 5 and Multimedia Appendix 4). Sensitivity analyses of outlier studies did not yield statistically significant data.

Figure 3. Meta-analysis: overall compression quality.

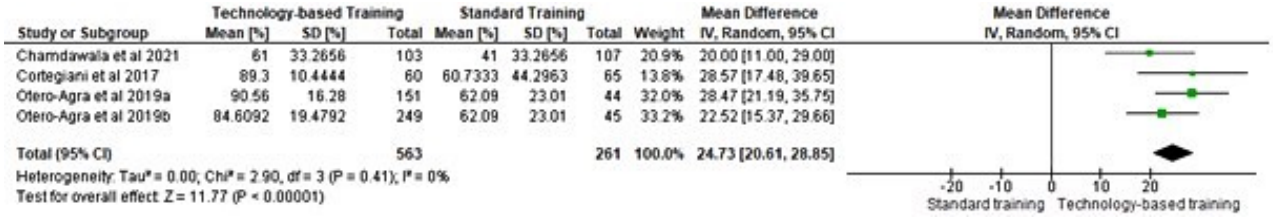


Table 5. Subgroup analyses: cardiopulmonary resuscitation skill components after training and at retention.

Subgroup analyses	Comparisons (n)	Effect estimate (95% CI)	Subgroup effect		I^2 (%)	Subgroup differences	
			Z statistic	*P* value		I^2 (%)	*P* value
Subgroup analyses based on instructor guidance							
Checking responsiveness (after training): 5 trials (6 arms)							
Self-directed learning	3	1.07 (0.83 to 1.38)	0.50	.61	86	67	*.08*[a]
Instructor-guided learning	3	1.39 (1.19 to 1.63)	4.10	*<.001*	0	—[b]	—
Checking airway (after training): 5 trials (6 arms)							
Self-directed learning	4	0.84 (0.75 to 0.94)	3.05	*.002*	0	4.6	.31
Instructor-guided learning	2	1.02 (0.71 to 1.48)	0.13	.90	63	—	—
Calling EMS[c] or help (after training): 6 trials (7 arms)							
Self-directed learning	4	1.10 (0.85 to 1.43)	0.72	.47	85	0	.93
Instructor-guided learning	3	1.11 (1.00 to 1.24)	2.01	*.04*	0	—	—
Mean compression depth (after training): 10 trials (13 arms)							
Self-directed learning	6	–3.16 (–8.17 to 1.85)	1.23	.22	93	83.9	*.01*
Instructor-guided learning	7	3.94 (1.50 to 6.37)	3.17	*.002*	78	—	—
Correct hand position (after training): 7 trials (10 arms)							
Self-directed learning	5	0.84 (0.71 to 0.99)	2.10	*.04*	48	64.1	*.09*
Instructor-guided learning	5	1.11 (0.83 to 1.47)	0.71	.48	74	—	—
Subgroup analyses based on hands-on practice							
Mean compression depth (after training): 10 trials (13 arms)							
Hands-on practice	11	2.20 (0.08 to 4.32)	2.03	*.04*	79	70.6	*.07*
Without practical training	2	–6.52 (–15.53 to 2.50)	1.42	.16	94	—	—
Mean compression rate (after training): 11 trials (15 arms)							
Hands-on practice	13	–5.47 (–9.26 to –1.68)	2.83	*.005*	81	96.7	*<.001*
Without practical training	2	9.38 (5.75 to 13.01)	5.07	*<.001*	0	—	—
Mean compression rate (retention): 6 trials (8 arms)							
Hands-on practice	6	–3.88 (–9.79 to 2.03)	1.29	.20	86	62.1	*.10*
Without practical training	2	1.80 (–1.67 to 5.27)	1.02	.31	0	—	—
Subgroup analyses based on training modalities							
Correct hand position (after training): 7 trials (10 arms)							
Video instruction	4	0.78 (0.61 to 1.00)	1.92	*.05*	0	9.3	.33
Computer program or mobile app	5	0.99 (0.82 to 1.18)	0.16	.87	66	—	—
Real-time feedback only	1	0.93 (0.84 to 1.02)	1.53	.13	N/A[d]	—	—

[a]Results of significance are presented in italics.

[b]Not available.

[c]EMS: emergency medical services.

[d]N/A: not applicable.

Instructor Guidance

Instructor-guided training significantly increased the likelihood of checking the responsiveness of people experiencing cardiac arrest (RR 1.39, 95% CI 1.19-1.63; *P*<.001) and calling the emergency medical services (RR 1.11, 95% CI 1.00-1.24; *P*=.04). Although heterogeneity was high, instructor-guided training potentially increased mean compression depth. All of the instructor-guided intervention arms reported MDs favoring technology-based training (statistical significance in 4 out of 7 [57%] arms [30,34,38,42] and insignificance in 3 out of 7 [43%] arms [35,41,42]). The sole use of self-directed learning

significantly decreased the likelihood of checking the airway (RR 0.84, 95% CI 0.75-0.94; *P*=.002) and correct hand position (RR 0.84, 95% CI 0.71-0.99; *P*=.04). Overall, instructor-guided training improved skills, whereas the sole use of self-directed learning yielded poorer skills.

Hands-on Practice

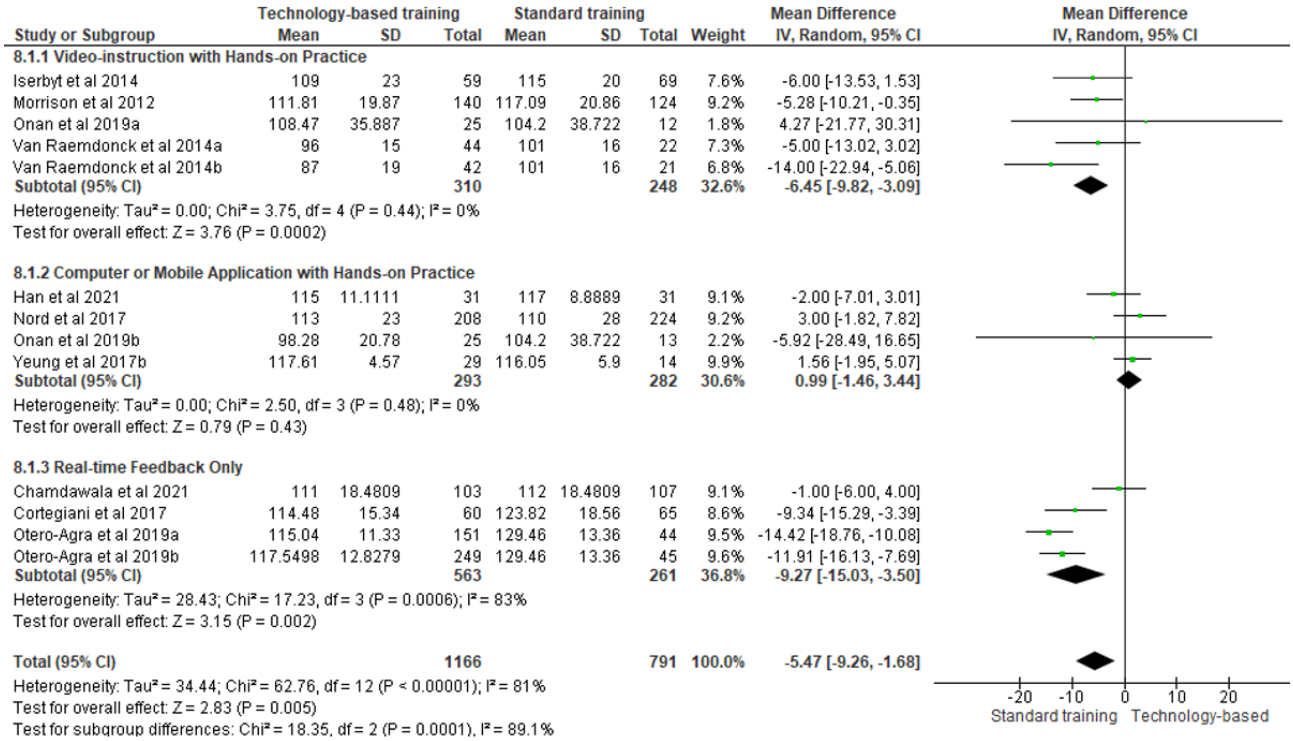
Despite high heterogeneity levels, hands-on practice potentially increased mean compression depth. Technology-based training interventions yielded significantly deeper chest compressions in 36% (4/11) of the intervention arms [30,34,38,42] and similar compression depths compared with standard training in 64% (7/11) of the intervention arms [27,31,35,40-42]. Similarly, although heterogeneity was high, hands-on practice potentially yielded slower compression rates than standard training; these reported compression rates were all within the guidelines of 100 to 120 compressions per minute. Training without hands-on practice significantly increased mean compression rate (MD 9.38, 95% CI 5.75-13.01; *P*<.001; I^2=0%).

Training Modalities

Subgroup analyses of studies involving hands-on practice using different training modalities (Figure 4) revealed that video instruction with hands-on practice yielded significantly slower compression rates (MD –6.45, 95% CI –9.82 to –3.09; *P*<.001; I^2=0%). Real-time feedback also potentially yielded slower compression rates, although heterogeneity was significantly high. Of the 4 arms involving real-time feedback that reported slower compression rates in the intervention groups, statistical significance was reached in 3 (75%) arms [32,34,35], whereas insignificance was reported in 1 (25%) arm [34].

Video instruction significantly decreased the likelihood of correct hand position (RR 0.78, 95% CI 0.61-1.00; *P*=.05; Table 5).

Figure 4. Subgroup analysis: mean compression rate after training.



Retention

Technology-based training decreased the likelihood of correct compression depth (RR 0.76, 95% CI 0.59-0.97; *P*=.03; I^2=0%; Table 3). Further analyses revealed that training with hands-on practice potentially significantly decreased mean compression rate compared with training without hands-on practice (I^2=62.1%; *P*=.10; Table 5). Technology-based training also potentially decreased the likelihood of correct ventilation at 2 to 6 months (RR 0.64, 95% CI 0.41-0.99; *P*=.05; I^2=78%; Table 3). Events of correct ventilation were significantly fewer in 60% (3/5) of the intervention arms [27,29] and similar to standard training in 40% (2/5) of the intervention arms [29,41].

Effectiveness of Technology-Based Training on Knowledge

Posttraining Performance

Of the 17 RCTs, 6 (35%; arms: 8/24, 33%) involving 2253 students reported knowledge scores using questionnaires [26,29,33,34,41,42]. Owing to high heterogeneity (I^2=89%; *P*<.001), subgroup analyses were performed (Multimedia Appendix 5). Instructor-guided training with hands-on practice potentially improved knowledge scores (Cohen *d*=0.45, 95% CI 0.13-0.78; *P*=.006; I^2=84%). Only Chamdawala et al [34] reported insignificantly poorer knowledge scores after training with real-time feedback, and this contributed to the considerable heterogeneity. Of the 4 studies that reported MDs favoring technology-based training, statistical significance was achieved

in 2 (50%) [29,41], whereas insignificance was reported in 1 (25%) [26]. *P* value was not reported by Onan et al [42].

Computer programs or mobile apps potentially improved knowledge scores (Cohen *d*=0.62, 95% CI 0.37-0.86; *P*<.001; I^2=74%). The studies reported MDs favoring technology-based training. However, statistical significance was reported only in 33% (1/3) of the studies [41]. *P* values were not reported by Onan et al [42] and Reder et al [29].

Marchiori et al [39] lacked sufficient data to be included in the meta-analysis but reported that video game–based training significantly improved knowledge scores.

Overall, the effect of technology-based training on knowledge after training remains inconclusive. However, instructor guidance, hands-on practice, and computer programs or mobile apps potentially improved knowledge.

Retention

Meta-analysis on knowledge scores at 2 to 6 months pooled from 18% (3/17) of the trials (arms: 4/24, 17%), which involved 1862 students, revealed high heterogeneity (I^2=89%; *P*<.001) [29,34,41]. Of these 3 studies, 1 (33%) [34] reported an insignificant difference in knowledge scores between training with real-time feedback and standard training, whereas 2 (67%) [29,41] reported significant improvements in knowledge scores. Overall, technology-based training potentially improved knowledge up to 6 months.

Discussion

Our review showed that technology-based CPR training involving instructor guidance, hands-on practice, and real-time feedback yielded favorable outcomes for secondary school and high school students after the intervention. Technology-based training also potentially improved overall skills performance and knowledge at retention.

CPR Skills

Posttraining Performance

Consistent with a recent meta-analysis conducted among laypeople and health care professionals [20], our study demonstrated that technology-based and standard CPR training produced comparable skills. The findings showed that instructor-guided training increased the likelihood of checking the responsiveness of people experiencing cardiac arrest and calling the emergency medical services and potentially increased compression depth. A meta-analysis [43] also reported better learning outcomes among health professionals who received instructor-supervised training compared with self-regulated learning. Instructors play important roles in increasing student motivation and providing personalized feedback on psychomotor skills [37,44]. These attributes, which are absent in self-directed learning, contribute to skill acquisition [45]. Consequently, the sole use of self-directed learning yielded poorer overall performance and reduced the likelihood of checking the airway and correct hand position. Similarly, a narrative review found that self-directed training potentially reduced overall CPR pass rates compared with instructor-led training [18]. Our findings suggest that instructor guidance remains essential for improved CPR performance in adolescents.

The findings revealed that technology-based training with hands-on practice potentially increased compression depth. In all of the included studies, the mean compression depth ranged from 30 mm to 53 mm, less than the maximum acceptable compression depth of 60 mm [46]. Adolescents often struggle with achieving adequate compression depths because of physical factors; for example, body weight [47]. Thus, practice is essential to acquire and reinforce proper compression techniques through trial and error. The incorporation of these participative and practical components boosted training success in adolescents [17]. In addition, video instruction with hands-on practice reduced mean compression rates, which were within the 2015 recommended guidelines [46] of 100 to 120 compressions per minute [31,40,42]. The study by Van Raemdonck et al [27] reported mean rates of <100 compressions per minute, considering that it applied the European Resuscitation Council 2005 guidelines, which accept 80 to 120 compressions per minute. Contrarily, training without hands-on practice increased compression rates. Without hands-on practice, instructions to *push hard, push fast* at *100 compressions per minute* may result in an overestimation of compression rates. Prior studies on technology-based training without practice also reported increased mean compression rates [48]. Overall, our findings suggest the importance of hands-on practice for improved CPR performance in adolescents.

In our review, real-time visual feedback improved overall compression quality, which comprises compression depth and rate, chest recoil, and hand position. Prior studies also reported that feedback devices contribute to improved chest compressions among health care professionals and adult laypeople [49]. Visual feedback allows trainees to contrast their performance against target parameters and correct themselves according to real-time performance data, improving skill acquisition. Real-time feedback also potentially yielded slower compression rates than standard training, and these mean rates were within 100 to 120 compressions per minute. The control groups in 67% (2/3) of the trials exceeded 120 compressions per minute [32,35]. Our findings suggest that real-time feedback improves chest compressions and possibly enhances adherence of compression rates to resuscitation guidelines.

However, video instruction reduced the likelihood of correct hand position. Similarly, an RCT [50] reported that video instruction training for health care staff yielded suboptimal hand position. As 67% (2/3) of the studies in our meta-analysis lacked clear descriptions of their instructional videos [27,42], it is challenging to examine explanations for this finding. One possible reason might be the inadequate emphasis on essential information; for example, anatomical landmarks for correct hand position in the instructional videos [31].

Retention

Technology-based training potentially improved overall performance at 6 months. Similarly, prior studies demonstrated improved skill retention in adolescents for up to 8 months after technology-based training [14]. However, technology-based training also reduced the likelihood of correct compression

depth and potentially reduced the likelihood of correct ventilation at 2 to 6 months. Without refresher training, regression of skill performance from the second month is expected [16]. Skill regression in compression depth and ventilation may be more evident because these skills are considered challenging for adolescents [47]. Our findings suggest that regular refresher training is necessary to prevent skill decay.

Knowledge

Our findings were consistent with those of a past meta-analysis [20] that reported improved knowledge after digital resuscitation training among laypeople and health care professionals. In particular, instructor guidance, hands-on practice, and computer programs or mobile apps potentially yielded higher knowledge scores. Instructors improve students' theoretical understanding by simplifying complex concepts, identifying individual areas of weaknesses, and promptly clarifying doubts [51]. Hands-on practice allows students to put theory into practice and enhance knowledge acquisition and retention [17]. The participative features in computer programs or mobile apps increase students' interest and help them to grasp concepts quickly [52]. Students can access training materials via electronic devices easily to reinforce knowledge and improve knowledge retention [14].

However, knowledge questionnaires were not standardized across the studies. Recently, a questionnaire assessing adolescents' CPR knowledge was developed and validated [53]. Adoption of standardized assessment by future studies will be beneficial because intervention effects on knowledge can be easily compared across studies. In addition, learning theories improved CPR knowledge [20]. However, only Iserbyt et al [31] in this review used learning theory to guide their intervention.

Strengths and Limitations

The strengths of this review include an extensive search in multiple bibliographic databases and comprehensive synthesis of results. However, the review was limited by the low quality of the included studies. Most (16/17, 94%) of the studies inadequately reported or took measures to reduce selection and performance biases. In addition, variations in intervention designs across the studies increased heterogeneity; for instance, videos and computer programs or mobile apps may emphasize theoretical knowledge, whereas interventions involving real-time feedback focused on CPR skills. Furthermore, several (11/17, 65%) of the technology-based interventions involved active interaction and engagement with students, whereas others (7/17, 41%) involved passive learning through videos. These variations made it challenging to draw conclusions on training elements required for optimal effectiveness. Finally, this review only included trials published in English.

Implications for Research

More high-quality RCTs with clear descriptions of study procedures—for example, allocation concealment and blinding of participants and personnel—are needed. These efforts will improve the credibility of evidence and contribute to stronger conclusions on the effectiveness of technology-based training for adolescents. Future studies should consider incorporating learning theories to guide their interventions [20]. Technology-based training can be considered a routine refresher training modality in schools for future research.

Implications for Practice

Overall, technology-based training demonstrated equivalence or improvements in skills and knowledge after training and at retention when compared with standard training among adolescents. From an educational perspective, the noninferiority of technology-based training offers a desirable alternative to standard training. Schools can consider using videos, computer programs, or mobile apps for self-directed theoretical instruction. However, instructor guidance and hands-on practice are still necessary components of training. Real-time feedback devices may also be used to increase students' compliance to resuscitation guidelines. Such a blended learning approach, comprising technology-based resources and face-to-face teaching, may reduce class time and reliance on instructor availability and increase schools' capacity for wider training dissemination. Furthermore, refresher training should focus on challenging skills; for example, compression depth and ventilation.

Conclusions

This review explored the use of technology-based training as an alternative to standard CPR training among secondary school and high school students. Our findings supported the use of technology-based components such as videos, computer programs, or mobile apps for self-directed theoretical instruction; these components potentially improve skills and knowledge retention. However, instructor guidance, hands-on practice, and real-time feedback are still necessary components of training to achieve better learning outcomes for adolescents. Such a blended learning approach may reduce class time and reliance on instructor availability. Regular refresher training is necessary for challenging skills such as compression depth and ventilation. Caution must be exercised when interpreting the results of this review because of the high heterogeneity of intervention characteristics. The overall low quality of evidence indicated the need for high-quality RCTs with large sample sizes and follow-up data.

Acknowledgments

The authors would like to thank Ms Wong Suei Nee for reviewing the search strategies and Associate Professor Wilson Tam for his advice regarding data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search terms and keywords.
[PDF File (Adobe PDF File), 99 KB - jmir_v24i12e36423_app1.pdf]

Multimedia Appendix 2
Descriptions of interventions.
[PDF File (Adobe PDF File), 104 KB - jmir_v24i12e36423_app2.pdf]

Multimedia Appendix 3
Grading of Recommendations Assessment, Development, and Evaluation.
[PDF File (Adobe PDF File), 143 KB - jmir_v24i12e36423_app3.pdf]

Multimedia Appendix 4
Results of subgroup analyses for skills and knowledge.
[PDF File (Adobe PDF File), 153 KB - jmir_v24i12e36423_app4.pdf]

Multimedia Appendix 5
Subgroup analyses of knowledge after training.
[PDF File (Adobe PDF File), 101 KB - jmir_v24i12e36423_app5.pdf]

References

1. Green CR, Botha JA, Tiruvoipati R. Cognitive function, quality of life and mental health in survivors of out-of-hospital cardiac arrest: a review. Anaesth Intensive Care 2015 Sep 01;43(5):568-576. [doi: 10.1177/0310057x1504300504]
2. Doan TN, Schultz BV, Rashford S, Bosley E. Surviving out-of-hospital cardiac arrest: the important role of bystander interventions. Australas Emerg Care 2020 Mar;23(1):47-54. [doi: 10.1016/j.auec.2019.12.003] [Medline: 31926962]
3. Böttiger BW, Bossaert L, Castrén M, Cimpoesu D, Georgiou M, Greif R, Board of European Resuscitation Council (ERC). Kids Save Lives - ERC position statement on school children education in CPR.: "Hands that help - Training children is training for life". Resuscitation 2016 Aug;105:A1-A3. [doi: 10.1016/j.resuscitation.2016.06.005] [Medline: 27339096]
4. Anantharaman V. Developing resuscitation programmes in the community: the tasks ahead for the National Resuscitation Council. Singapore Med J 2011 Aug;52(8):634-641 [FREE Full text] [Medline: 21879226]
5. Bohn A, Van Aken H, Möllhoff T, Wienzek H, Kimmeyer P, Wild E, et al. Teaching resuscitation in schools: annual tuition by trained teachers is effective starting at age 10. A four-year prospective cohort study. Resuscitation 2012 May;83(5):619-625. [doi: 10.1016/j.resuscitation.2012.01.020] [Medline: 22286049]
6. Nord A. Bystander CPR New aspects of CPR training among students and the importance of bystander education level on survival. Linköping Sweden: Linköping University Electronic Press; 2017.
7. Cheng A, Magid DJ, Auerbach M, Bhanji F, Bigham BL, Blewer AL, et al. Part 6: resuscitation education science: 2020 American heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020 Oct 20;142(16_suppl_2):S551-S579 [FREE Full text] [doi: 10.1161/CIR.0000000000000903] [Medline: 33081527]
8. Lockey AS, Barton K, Yoxall H. Opportunities and barriers to cardiopulmonary resuscitation training in English secondary schools. Eur J Emerg Med 2016;23(5):381-385. [doi: 10.1097/mej.0000000000000307]
9. Salvatierra GG, Palazzo SJ, Emery A. High school CPR/AED training in Washington state. Public Health Nurs 2017 May 05;34(3):238-244. [doi: 10.1111/phn.12293] [Medline: 27704621]
10. Wyllie J, Bruinenberg J, Roehr CC, Rüdiger M, Trevisanuto D, Urlesberger B. European resuscitation council guidelines for resuscitation 2015: section 7. Resuscitation and support of transition of babies at birth. Resuscitation 2015 Oct;95:249-263. [doi: 10.1016/j.resuscitation.2015.07.029] [Medline: 26477415]
11. Merchant RM, Topjian AA, Panchal AR, Cheng A, Aziz K, Berg KM, Adult BasicAdvanced Life Support, Pediatric BasicAdvanced Life Support, Neonatal Life Support, Resuscitation Education Science,Systems of Care Writing Groups. Part 1: executive summary: 2020 American heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020 Oct 20;142(16_suppl_2):S337-S357 [FREE Full text] [doi: 10.1161/CIR.0000000000000918] [Medline: 33081530]
12. Bull S. Technology-Based Health Promotion. Thousand Oaks, California, United States: SAGE Publications; 2010.
13. Saks K, Leijen Ä. Distinguishing self-directed and self-regulated learning and measuring them in the e-learning context. Procedia Soc Behav Sci 2014 Feb;112:190-198. [doi: 10.1016/j.sbspro.2014.01.1155]

14. Semeraro F, Frisoli A, Loconsole C, Mastronicola N, Stroppa F, Ristagno G, et al. Kids (learn how to) save lives in the school with the serious game Relive. Resuscitation 2017 Jul;116:27-32. [doi: 10.1016/j.resuscitation.2017.04.038] [Medline: 28476478]
15. Cheskes S. High school CPR training: it's only an APP away!!. Resuscitation 2017 Nov;120:A9-10. [doi: 10.1016/j.resuscitation.2017.08.241] [Medline: 28974343]
16. Plant N, Taylor K. How best to teach CPR to schoolchildren: a systematic review. Resuscitation 2013 Apr;84(4):415-421. [doi: 10.1016/j.resuscitation.2012.12.008] [Medline: 23246989]
17. Reveruzzi B, Buckley L, Sheehan M. School-based first aid training programs: a systematic review. J Sch Health 2016 Apr;86(4):266-272. [doi: 10.1111/josh.12373] [Medline: 26930238]
18. Hsieh M, Bhanji F, Chiang W, Yang C, Chien K, Ma MH. Comparing the effect of self-instruction with that of traditional instruction in basic life support courses-a systematic review. Resuscitation 2016 Nov;108:8-19. [doi: 10.1016/j.resuscitation.2016.08.021] [Medline: 27581252]
19. Riggs M, Franklin R, Saylany L. Associations between cardiopulmonary resuscitation (CPR) knowledge, self-efficacy, training history and willingness to perform CPR and CPR psychomotor skills: a systematic review. Resuscitation 2019 May;138:259-272. [doi: 10.1016/j.resuscitation.2019.03.019] [Medline: 30928504]
20. Lau Y, Nyoe RS, Wong SN, Ab Hamid ZB, Leong BS, Lau ST. Effectiveness of digital resuscitation training in improving knowledge and skills: a systematic review and meta-analysis of randomised controlled trials. Resuscitation 2018 Oct;131:14-23. [doi: 10.1016/j.resuscitation.2018.07.033] [Medline: 30071263]
21. 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]
22. 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. BMJ 2011 Oct 18;343(oct18 2):d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
23. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
24. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 2010 Apr 21;1(2):97-111. [doi: 10.1002/jrsm.12] [Medline: 26061376]
25. Fu R, Gartlehner G, Grant M, Shamliyan T, Sedrakyan A, Wilt TJ, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the effective health care program. J Clin Epidemiol 2011 Nov;64(11):1187-1197. [doi: 10.1016/j.jclinepi.2010.08.010] [Medline: 21477993]
26. Norman RW. A comparison of the effectiveness and efficiency of three demonstration conditions in the acquisition of a motor skill (time, cardiopulmonary resuscitation). The Florida State University ProQuest Dissertations Publishing. 1984. URL: https://tinyurl.com/msf2kmwr [accessed 2021-06-30]
27. Van Raemdonck V, Monsieurs KG, Aerenhouts D, De Martelaer K. Teaching basic life support: a prospective randomized study on low-cost training strategies in secondary schools. Eur J Emerg Med 2014 Aug;21(4):284-290. [doi: 10.1097/MEJ.0000000000000071] [Medline: 23978956]
28. Beskind DL, Stolz U, Thiede R, Hoyer R, Burns W, Brown J, et al. Viewing a brief chest-compression-only CPR video improves bystander CPR performance and responsiveness in high school students: a cluster randomized trial. Resuscitation 2016 Jul;104:28-33. [doi: 10.1016/j.resuscitation.2016.03.022] [Medline: 27112909]
29. Reder S, Cummings P, Quan L. Comparison of three instructional methods for teaching cardiopulmonary resuscitation and use of an automatic external defibrillator to high school students. Resuscitation 2006 Jun;69(3):443-453. [doi: 10.1016/j.resuscitation.2005.08.020] [Medline: 16678958]
30. Yeung J, Kovic I, Vidacic M, Skilton E, Higgins D, Melody T, et al. The school Lifesavers study-a randomised controlled trial comparing the impact of Lifesaver only, face-to-face training only, and Lifesaver with face-to-face training on CPR knowledge, skills and attitudes in UK school children. Resuscitation 2017 Nov;120:138-145. [doi: 10.1016/j.resuscitation.2017.08.010] [Medline: 28974344]
31. Iserbyt P, Charlier N, Mols L. Learning basic life support (BLS) with tablet PCs in reciprocal learning at school: are videos superior to pictures? A randomized controlled trial. Resuscitation 2014 Jun;85(6):809-813. [doi: 10.1016/j.resuscitation.2014.01.018] [Medline: 24503451]
32. Otero-Agra M, Barcala-Furelos R, Besada-Saavedra I, Peixoto-Pino L, Martínez-Isasi S, Rodríguez-Núñez A. Let the kids play: gamification as a CPR training methodology in secondary school students. A quasi-experimental manikin simulation study. Emerg Med J 2019 Nov 13;36(11):653-659. [doi: 10.1136/emermed-2018-208108] [Medline: 31409636]
33. Rezaei A, Safari moradabadi A, Montazerghaem H, Miri HR, Paknahad A, Alavi A. Comparing two training methods, traditional CPR skill and distance training at the Red-crescent organization in Hormozgan province. Life Sci J 2013;10(9s):140.

34. Chamdawala H, Meltzer JA, Shankar V, Elachi D, Jarzynka SM, Nixon AF. Cardiopulmonary resuscitation skill training and retention in teens (CPR START): a randomized control trial in high school students. Resusc Plus 2021 Mar;5:100079 [FREE Full text] [doi: 10.1016/j.resplu.2021.100079] [Medline: 34223345]
35. Cortegiani A, Russotto V, Montalto F, Iozzo P, Meschis R, Pugliesi M, et al. Use of a real-time training software (Laerdal QCPR®) compared to instructor-based feedback for high-quality chest compressions acquisition in secondary school students: a randomized trial. PLoS One 2017 Jan 5;12(1):e0169591 [FREE Full text] [doi: 10.1371/journal.pone.0169591] [Medline: 28056076]
36. Cuijpers P, Schakelaar E, van Merrienboer J, Gorgels A. AP052 three instructional methods for CPR training in secondary schools: practical training remains necessary. Resuscitation 2011 Oct;82:S21. [doi: 10.1016/s0300-9572(11)70085-5]
37. Doucet L, Lammens R, Hendrickx S, Dewolf P. App-based learning as an alternative for instructors in teaching basic life support to school children: a randomized control trial. Acta Clin Belg 2019 Oct 14;74(5):317-325. [doi: 10.1080/17843286.2018.1500766] [Medline: 30105946]
38. Han S, Park HJ, Nah S, Lee EH, Lee HJ, Park JO, et al. Instructor-led distance learning for training students in cardiopulmonary resuscitation: a randomized controlled study. PLoS One 2021 May 6;16(5):e0251277 [FREE Full text] [doi: 10.1371/journal.pone.0251277] [Medline: 33956873]
39. Marchiori E, Ferrer G, Fernández-Manjón B, Povar-Marco J, Suberviola JF, Giménez-Valverde A. Video-game instruction in basic life support maneuvers. Emergencias 2012 Dec;24(6):433-437.
40. Morrison LJ, Yu M, Hurst B, Dainty KN, Buick JE, Cote J, et al. CPR anytime any school: a randomized trial of training strategies to teach CPR and use of an AED to high school students. In: Proceedings of the Canadian Association of Emergency Physicians Annual Conference. 2012 Presented at: Canadian Association of Emergency Physicians Annual Conference; Jun, 2012; Niagara Falls ON, Canada.
41. Nord A, Svensson L, Claesson A, Herlitz J, Hult H, Kreitz-Sandberg S, et al. The effect of a national web course "Help-Brain-Heart" as a supplemental learning tool before CPR training: a cluster randomised trial. Scand J Trauma Resusc Emerg Med 2017 Sep 12;25(1):93 [FREE Full text] [doi: 10.1186/s13049-017-0439-0] [Medline: 28899418]
42. Onan A, Turan S, Elcin M, Erbil B, Bulut Ş. The effectiveness of traditional Basic Life Support training and alternative technology-enhanced methods in high schools. Hong Kong J Emerg Med 2018 Jun 14;26(1):44-52.
43. Brydges R, Manzone J, Shanks D, Hatala R, Hamstra SJ, Zendejas B, et al. Self-regulated learning in simulation-based training: a systematic review and meta-analysis. Med Educ 2015 Apr 20;49(4):368-378. [doi: 10.1111/medu.12649] [Medline: 25800297]
44. Saraç L, Ok A. The effects of different instructional methods on students' acquisition and retention of cardiopulmonary resuscitation skills. Resuscitation 2010 May;81(5):555-561. [doi: 10.1016/j.resuscitation.2009.08.030] [Medline: 20149946]
45. Süss-Havemann C, Kosan J, Seibold T, Dibbern NM, Daubmann A, Kubitz JC, et al. Implementation of Basic Life Support training in schools: a randomised controlled trial evaluating self-regulated learning as alternative training concept. BMC Public Health 2020 Jan 13;20(1):50 [FREE Full text] [doi: 10.1186/s12889-020-8161-7] [Medline: 31931770]
46. Perkins GD, Handley AJ, Koster RW, Castrén M, Smyth MA, Olasveengen T, Adult basic life supportautomated external defibrillation section Collaborators. European resuscitation council guidelines for resuscitation 2015: section 2. Adult basic life support and automated external defibrillation. Resuscitation 2015 Oct;95:81-99. [doi: 10.1016/j.resuscitation.2015.07.015] [Medline: 26477420]
47. Abelairas-Gómez C, Rodríguez-Núñez A, Casillas-Cabana M, Romo-Pérez V, Barcala-Furelos R. Schoolchildren as life savers: at what age do they become strong enough? Resuscitation 2014 Jun;85(6):814-819. [doi: 10.1016/j.resuscitation.2014.03.001] [Medline: 24614187]
48. Panchal AR, Meziab O, Stolz U, Anderson W, Bartlett M, Spaite DW, et al. The impact of ultra-brief chest compression-only CPR video training on responsiveness, compression rate, and hands-off time interval among bystanders in a shopping mall. Resuscitation 2014 Sep;85(9):1287-1290. [doi: 10.1016/j.resuscitation.2014.06.013] [Medline: 24983200]
49. González-Salvado V, Rodríguez-Ruiz E, Abelairas-Gómez C, Ruano-Raviña A, Peña-Gil C, González-Juanatey JR, et al. Training adult laypeople in basic life support. A systematic review. Rev Esp Cardiol (Engl Ed) 2020 Jan;73(1):53-68 [FREE Full text] [doi: 10.1016/j.rec.2018.11.013] [Medline: 30808611]
50. Akhtar N, Field RA, Davies RP, Perkins GD. Video-based CPR training--the importance of quality assurance. Resuscitation 2012 Jul;83(7):e158-e159. [doi: 10.1016/j.resuscitation.2012.03.006] [Medline: 22429974]
51. Montgomery C, Kardong-Edgren SE, Oermann MH, Odom-Maryon T. Student satisfaction and self report of CPR competency: heartCode BLS courses, instructor-led CPR courses, and monthly voice advisory manikin practice for CPR skill maintenance. Int J Nurs Educ Scholarsh 2012 Jun 03;9:/j/ijnes.2012.9.issue-1/1548-923X.2361/1548-923X.2361.xml. [doi: 10.1515/1548-923X.2361] [Medline: 22673961]
52. Raja R, Nagasubramani PC. Impact of modern technology in education. J Applied Advance Res 2018 May 10:S33-S35. [doi: 10.21839/JAAR.2018.v3is1.165]
53. Borovnik Lesjak V, Šorgo A, Strnad M. Development, validation and assessment of the test on knowledge about basic life support and use of automated external defibrillator among schoolchildren. Scand J Trauma Resusc Emerg Med 2019 Dec 23;27(1):114 [FREE Full text] [doi: 10.1186/s13049-019-0683-6] [Medline: 31870420]

Abbreviations

CPR: cardiopulmonary resuscitation
MD: mean difference
OCHA: out-of-hospital cardiac arrest
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: risk ratio

Edited by G Eysenbach; submitted 14.01.22; peer-reviewed by AKC Wong, L Suppan, SLS Drop; comments to author 19.09.22; revised version received 15.10.22; accepted 20.10.22; published 15.12.22.

Please cite as:
Lim XMA, Liao WA, Wang W, Seah B
The Effectiveness of Technology-Based Cardiopulmonary Resuscitation Training on the Skills and Knowledge of Adolescents: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(12):e36423
URL: https://www.jmir.org/2022/12/e36423
doi:10.2196/36423
PMID:36520524

Review

Effect of Mobile Phone App–Based Interventions on Quality of Life and Psychological Symptoms Among Adult Cancer Survivors: Systematic Review and Meta-analysis of Randomized Controlled Trials

Minghui Qin[1,2*], MD; Bo Chen[2*], MD; Shaohua Sun[3], MD; Xiaodong Liu[4], PhD

[1]Department of Traditional Chinese Medicine, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Art and Science, Xiangyang, China

[2]Center for Clinical Evidence-Based and Translational Medicine, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang, China

[3]Department of Oncology, Suizhou Central Hospital, Hubei University of Medicine, Suizhou, China

[4]Information Center, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Art and Science, Xiangyang, China

[*]these authors contributed equally

Corresponding Author:
Xiaodong Liu, PhD
Information Center
Xiangyang Central Hospital
Affiliated Hospital of Hubei University of Art and Science
No 136, Jingzhou Street, Xiangcheng District
Xiangyang, 441021
China
Phone: 86 18602770213
Email: lxd20000@163.com

Abstract

Background: Most patients with cancer experience psychological or physical distress, which can adversely affect their quality of life (QOL). Smartphone app interventions are increasingly being used to improve QOL and psychological outcomes in patients with cancer. However, there is insufficient evidence regarding the effect of this type of intervention, with conflicting results in the literature.

Objective: In this systematic review and meta-analysis, we investigated the effectiveness of mobile phone app interventions on QOL and psychological outcomes in adult patients with cancer, with a special focus on intervention duration, type of cancer, intervention theory, treatment strategy, and intervention delivery format.

Methods: We conducted a literature search of PubMed, Web of Science, the Cochrane Library, Embase, Scopus, China National Knowledge Infrastructure, and WanFang to identify studies involving apps that focused on cancer survivors and QOL or psychological symptoms published from inception to October 30, 2022. We selected only randomized controlled trials that met the inclusion criteria and performed systematic review and meta-analysis. The standardized mean difference (SMD) with a 95% CI was pooled when needed. Sensitivity and subgroup analyses were also conducted.

Results: In total, 30 randomized controlled trials with a total of 5353 participants were included in this meta-analysis. Compared with routine care, app interventions might improve QOL (SMD=0.39, 95% CI 0.27-0.51; *P*<.001); enhance self-efficacy (SMD=0.15, 95% CI 0.02-0.29; *P*=.03); and alleviate anxiety (SMD=−0.64, 95% CI −0.73 to −0.56; *P*<.001), depression (SMD=−0.33, 95% CI −0.58 to −0.08; *P*=.009), and distress (SMD=−0.34, 95% CI −0.61 to −0.08; *P*=.01). Short-term (duration of ≤3 months), physician-patient interaction (2-way communication using a smartphone app), and cognitive behavioral therapy interventions might be the most effective for improving QOL and alleviating adverse psychological effects.

Conclusions: Our study showed that interventions using mobile health apps might improve QOL and self-efficacy as well as alleviate anxiety, depression, and distress in adult cancer survivors. However, these results should be interpreted with caution because of the heterogeneity of the interventions and the study design. More rigorous trials are warranted to confirm the suitable duration and validate the different intervention theories as well as address methodological flaws in previous studies.

Trial Registration: PROSPERO CRD42022370599; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=370599

(J Med Internet Res 2022;24(12):e39799) doi:10.2196/39799

KEYWORDS

mobile health app; mHealth app; quality of life; psychological symptoms; cancer survivors; systematic review and meta-analysis; mobile phone

Introduction

Background

Worldwide, the number of new cancer cases diagnosed each year is rapidly increasing, from 14.1 million in 2012 to an estimated 21.6 million in 2030 [1]. With advancements in early detection and clinical treatment techniques, these patients now have a better prognosis and longer life expectancy [2]. However, approximately 30% to 40% of patients with cancer have at least one psychological or physical symptom, such as anxiety, depression, or distress [3-5], and up to 50% of women diagnosed with breast cancer experience psychological issues at some point in their illness [6], which may negatively affect their quality of life (QOL) and make them more stressed [7,8].

Although psychological problems are common in patients with cancer, they are not inevitable, and appropriate interventions can reduce the impact of these problems. Following the emergence and worldwide spread of COVID-19, the growing popularity of smartphone health apps may represent an opportunity to improve cancer care and management. These apps can be used to collect objective data about patients' behavior and behavior monitoring, which could help patients change their behavior, promote self-monitoring of symptoms, and enhance patients' sense of empowerment and willingness to care for themselves [9] while allowing them to communicate with their health care team from a distance [10,11].

Various randomized controlled trials (RCTs) have found that mobile health (mHealth) interventions may be effective for adult cancer survivors. For example, mHealth interventions have increased the number of women screened for breast cancer [12]. Similarly, among patients with pancreatic ductal adenocarcinoma receiving chemotherapy, a mobile app intervention provided adequate nutritional and psychological support [13]. In addition, a web-based exercise intervention successfully increased the number of patients with cancer who engaged in physical activity [14]. Okunade et al [15] also predicted that telemedicine would be integrated into the care of patients in oncology following the COVID-19 pandemic; however, sufficient evidence to guide such integration has not been established. Owing to the issue of patients' access, or lack thereof, to app interventions, it is difficult to design and implement unbiased, blinded RCTs to determine their true effects. The evidence for the efficacy of mHealth app interventions in cancer treatment might be unreliable. Some studies have demonstrated that smartphone app interventions benefit mental health [16,17]. By contrast, other studies have found no association between smartphone app interventions and psychological outcomes [18,19]. Further studies have suggested that apps increase patient anxiety and depression by enriching cancer information, which reminds them of what they are experiencing [20]. Thus, given this contradictory evidence, clarifying the psychological effects of app interventions remains difficult.

Although several systematic reviews have addressed the psychological impact of teleinterventions on cancer survivors [21-24], contradictory results remain. A meta-analysis that included 20 telehealth interventions found that the interventions improved patients' QOL and self-efficacy and reduced depression, distress, and perceived stress. However, the interventions did not have any significant effect on anxiety [21]. Similarly, another meta-analysis of 14 phone-based interventions found that these interventions reduced anxiety and improved QOL but did not have any significant effect on depression [24]. No meta-analysis has comprehensively and specifically assessed the impact of smartphone apps on QOL and psychological symptoms in cancer survivors. Smartphone apps have natural advantages over websites and SMS text messaging, such as personalized design, rich mobile device features based on smartphones (cameras, phones, GPS, and contact lists), and timely push features. Therefore, smartphone app interventions may have higher adherence.

Objectives

We conducted a systematic review and meta-analysis of RCTs to determine the effects of app interventions on QOL and psychological outcomes in adult cancer survivors. We also performed various subgroup analyses according to intervention duration, type of cancer, intervention theory, treatment strategy, and intervention delivery format to investigate the effects of app interventions.

Methods

The meta-analysis adhered to the Cochrane Handbook guidelines for conducting systematic reviews and meta-analyses and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and was registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42022370599).

Ethical Considerations

This review did not require informed consent or ethics approval as the data were obtained from previously published studies.

Article Selection and Search Strategy

We searched the following databases from inception to October 30, 2022: PubMed, Web of Science, Embase, the Cochrane Library, Scopus, China National Knowledge Infrastructure, and WanFang. For the literature search, we combined Medical

Subject Headings and non–Medical Subject Heading terms, including ("cancer" OR "tumor" OR "neoplasms" OR "neoplasia") AND ("mHealth applications" OR "mHealth" OR "portable software application" OR "app" OR "apps" OR "app-based" OR "electronic") AND ("randomized controlled trial" OR "controlled clinical trial" OR "randomized" OR "placebo" OR "clinical trials as topic" OR "randomly" OR "trial") NOT ("animals") NOT ("humans" AND "animals"). There were no language restrictions. Additional relevant studies were identified by manually searching the references of the screened articles and reviews (Multimedia Appendix 1).

Inclusion and Exclusion Criteria

The following criteria were used to determine whether to include each study: (1) adults with cancer (of any type or stage); (2) telehealth or telemedicine interventions delivered via an mHealth app; (3) a control group involving routine care, including usual care, waitlist control, conventional care, or health education delivered without the use of an mHealth app; (4) the outcome being QOL and psychological outcomes (including depression, anxiety, distress, and self-efficacy) with no restrictions on the measurement tools used; and (5) RCT study design. We excluded studies that used websites, SMS text messaging, email, or other technological interventions that did not include mHealth apps and studies that used mHealth apps without involving patients with cancer (eg, health care professionals who used mHealth apps). In addition, we excluded study protocols, reviews, and studies lacking complete data. The publication date was not restricted in any way.

Data Extraction and Risk-of-Bias Assessment

The data management software EndNote X9 (Clarivate Analytics) was used. In total, 2 researchers (QMH and CB) independently extracted the data based on the qualifying criteria. Disagreements were resolved through discussion between the evaluators. If the data were duplicated or shared between studies, the most recently published or more comprehensive study was used in the analysis. We extracted the following data from each included study: first author, publication date, country, intervention theory, sample size, participant characteristics (mean age, type of cancer, and stage of cancer), intervention duration, treatment strategy, format of intervention delivery, and outcome measurements. The Cochrane risk-of-bias assessment tool was used to determine the risk of bias (Multimedia Appendix 2).

Statistical Analysis

Following data extraction from the publications, heterogeneity tests and statistical analyses were conducted using RevMan (version 5.3; The Cochrane Collaboration) and Stata (version 16.0; StataCorp) software. As these included studies used various measuring tools, the standardized mean difference (SMD) with a 95% CI was used to estimate intervention effects on QOL, depression, anxiety, distress, and self-efficacy. If SDs were not provided, they were calculated using the available data. A 2-sided $P<.05$ was used to indicate a statistically significant difference in the overall effect. To determine the statistical heterogeneity of the included studies, the I^2 statistic and P value were used. A fixed-effects model was used to pool the results if $I^2 \leq 50\%$ and $P>.10$; if heterogeneity was significant ($P<.10$ and $I^2>50\%$), a random-effects model was used to pool the results. If necessary and feasible, subgroup and sensitivity analyses were conducted to identify possible sources of between-study heterogeneity. Subgroup analyses were conducted based on intervention duration, type of cancer, intervention theory, treatment strategy, and intervention delivery format. Sensitivity analyses were carried out by omitting 3% (1/30) of the studies and modifying the pooling model (random-effects or fixed-effects models). To assess publication bias, the Begg and Egger regression tests were used.

Results

Characteristics of the Included Studies and Risk of Bias

The PRISMA flowchart depicts the extensive search process (Figure 1). Initially, 1491 articles were identified, with 38 (2.55%) records being further evaluated as potentially eligible. Finally, the meta-analysis included 2.01% (30/1491) of RCTs (with 5353 participants). Table 1 summarizes the characteristics of these studies. Each included study had a sample size ranging from 38 to 829. Participants' age ranged from 41.9 (SD 11.30) to 67.1 (SD 10.4) years on average. The interventions lasted from 1 week to 12 months, with a median follow-up time of 2.8 months. Of the 30 studies, 13 (43%) included only patients with breast cancer; 7 (23%) used cognitive behavioral therapy interventions; and 9 (30%) and 7 (23%) included only patients treated with surgery and chemotherapy, respectively. In addition, different scales were used to assess the outcomes.

The assessment of the risk of bias is shown in in Figure 2 [16-20,25-49] and Multimedia Appendix 2 [16-20,25-49]. The process of random sequence generation was explicitly described in 90% (27/30) of the studies. In 43% (13/30) of the studies, allocation concealment was adequately reported. A total of 63% (19/30) of the studies had a high risk of bias because of patients' access or lack thereof to the mHealth app interventions, which made it difficult to blind participants and researchers. Regarding attrition bias, 33% (10/30) of the studies were rated as having an unclear risk of bias because of insufficient information on attrition. In comparison, 7% (2/30) of the studies were rated as having a high risk of bias because of high attrition rates. In total, 47% (14/30) of the studies published study protocols and reported all prespecified outcomes and were rated as having a low risk of reporting bias.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection. CNKI: China National Knowledge Infrastructure.

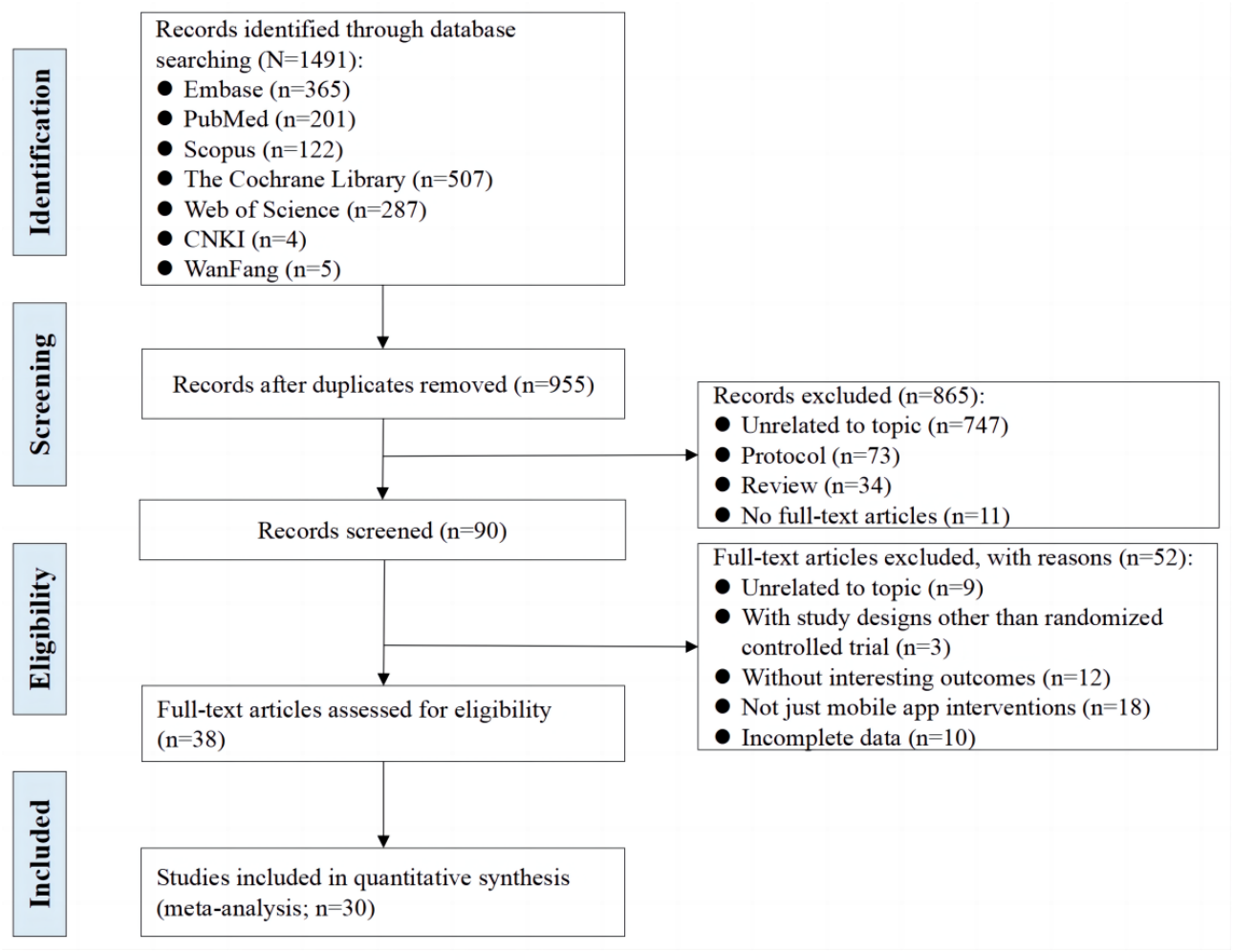


Table 1. Characteristics of the randomized controlled trial studies (N=30).

Author, year, and country	Intervention theory	Sample size, N		Patient characteristics					Intervention	Format of intervention delivery	Intervention duration	Outcomes and outcome measures
		Intervention	Control	Age (years), intervention	Age (years), control	Type of cancer	Stage of cancer	Treatment category				
Børøsund et al [25], 2021, Northern Europe	Cognitive behavioral theory	84	88	Mean 51.7 (SD 10.5)	Mean 52.3 (SD 12.0)	Breast cancer, brain cancer, prostate cancer, and others	Not restricted	Surgery, chemotherapy, radiation, and immune therapy	Intervention group: StressProffen app; control group: usual care	Interactive format (smartphone–based 2-way communication)	12 months	Anxiety (HADS-A[a]), depression (HADS-D[b]), and HRQOL[c] (SF-36[d])
Çınar et al [26], 2021, Turkey	Evidence-based symptom care theory	31	33	Mean 45.9 (SD 8.3)	Mean 45.5 (SD 9.8)	Breast cancer	Stage I to III	Surgery	Intervention group: mHealth[e] app–based patient education; control group: routine care	Interactive format (smartphone–based 2-way communication)	12 months	QOL[f] (FACT-ES QLS[g]) and distress (NCCN-DT[h])
Fjell et al [27], 2020, Sweden	Unclear	74	75	Mean 48.0 (SD 10.6)	Mean 50.0 (SD 11.6)	Breast cancer	Not restricted	Chemotherapy	Intervention group: Interaktor app; control group: standard care	Interactive format (smartphone–based 2-way communication)	18 weeks	Distress (MSAS-GDI[i]) and QOL (EORTC QLQ-C30[j])
Foley et al [20], 2016, Ireland	Unclear	13	26	Median 54 (IQR 49.5-61.5)	Median 52 (IQR 44-64)	Breast cancer	Not restricted	Surgery	Intervention group: Apple iPad; control group: standard care information	Didactic format (smartphone–based 1-way communication)	1 week	Anxiety (HADS-A) and depression (HADS-D)
Ghanbari et al [17], 2021, Iran	Cognitive behavioral theory	41	41	Mean 46.9 (SD 9.83)	Mean 46.0 (SD 8.80)	Nonmetastatic breast cancer	Not restricted	Not restricted	Intervention group: BC-Szone app; control group: waitlist control	Interactive format (smartphone–based 2-way communication)	5 weeks	Anxiety (STAI[k])
Greer et al [28], 2019, United States	Cognitive behavioral theory	72	73	Mean 55.86 (SD 10.08)	Mean 57.03 (SD 12.42)	Gastrointestinal cancer, gynecological cancer, lung cancer, breast cancer, and others	Stage IV or metastatic disease	Surgery, chemotherapy, radiation, and immune therapy	Intervention group: CBT[l] mHealth app; control group: health education control	Interactive format (smartphone–based 2-way communication)	3 months	Anxiety (HADS-A), depression (HADS-D), and QOL (PHQ-9[m])

Author, year, and country	Intervention theory	Sample size, N		Patient characteristics					Intervention	Format of intervention delivery	Intervention duration	Outcomes and outcome measures
		Intervention	Control	Age (years), intervention	Age (years), control	Type of cancer	Stage of cancer	Treatment category				
Greer et al [29], 2020, United States	Unclear	91	90	Mean 52.85 (SD 13.74)	Mean 53.76 (SD 12.08)	Hematologic cancer, non–small cell lung cancer, breast cancer, high-grade glioma, sarcoma, and others	Not restricted	Chemotherapy	Intervention group: mHealth app; control group: standard care	Didactic format (smartphone–based 1-way communication)	3 months	QOL (FACT-G[n])
Ham et al [30], 2019, Korea	CBT	21	21	Mean 41.90 (SD 11.30)	Mean 47.10 (SD 11.19)	Breast cancer, gynecological cancer, thyroid cancer, sarcoma, and others	Stage 0 to III	Surgery, radiotherapy, chemotherapy, and other treatments	Intervention group: HARUToday app; control group: waitlist control	Didactic format (smartphone–based 1-way communication)	10 weeks	Depression (BDI-II[o]), QOL (SF-36), and anxiety (STAI)
Handa et al [18], 2020, Japan	Unclear	47	48	Mean 49.9 (SD 0.2)	Mean 49.9 (SD 9.2)	Breast cancer	Not restricted	Chemotherapy	Intervention group: BPSS app; control group: ordinary instructions	Didactic format (smartphone–based 1-way communication)	12 weeks	Anxiety (HADS-A) and depression (HADS-D)
Karaaslan-Eser and Ayaz-Alkaya [31], 2021, Turkey	Unclear	42	42	Mean 60.33 (SD 9.31)	Mean 62.14 (SD 9.97)	Colorectal cancer, gastrointestinal stromal tumor, lung cancer, renal cell carcinoma, hepatocellular carcinoma, cholangiocarcinoma, and breast cancer	Stage III to IV	Oral anticancer agents	Intervention group: OKT-ED app; control group: standard care	Interactive format (smartphone–based 2-way communication)	6 months	Distress (MSAS-GDI)
Kim et al [19], 2018, Korea	Unclear	36	40	Median 49.8	Median 52.1	Breast cancer	Stage IV	Chemotherapy	Intervention group: mHealth game app; control group: conventional education	Interactive format (smartphone–based 2-way communication)	3 weeks	QOL (WHO-QOL-BREF[p] questionnaire), anxiety (STAI), and depression (BDI[q])

Author, year, and country	Intervention theory	Sample size, N		Patient characteristics					Intervention	Format of intervention delivery	Intervention duration	Outcomes and outcome measures
		Intervention	Control	Age (years), intervention	Age (years), control	Type of cancer	Stage of cancer	Treatment category				
Kubo et al [16], 2020, United States	Mindfulness-based therapy	31	46	Mean 65.8 (SD 8.8)	Mean 67.1 (SD 10.4)	Breast, hematological, gastrointestinal, lung, urological, and gynecological cancer	Not restricted	Not restricted	Intervention group: Headspace app; control group: waitlist control	Interactive format (smartphone–based 2-way communication)	3 months	Anxiety (HADS-A), depression (HADS-D), QOL (FACIT-Pal[r]), and distress (NCCN-DT)
Park et al [32], 2021, South Korea	Unclear	31	30	Mean 52.07 (SD 9.34)	Mean 54.74 (SD 7.87)	Breast cancer	Stage 0 to III	Surgery	Intervention group: Pillsy mHealth app; control group: usual care	Didactic format (smartphone–based 1-way communication)	4 weeks	Depression (Center for Epidemiologic Studies Depression Scale) and self-efficacy (General Self-Efficacy Scale)
Peng et al [33], 2020, China	Unclear	152	150	Mean 55.6 (SD 6.8)	Mean 56.3 (SD 7.0)	Not restricted	Not restricted	Not restricted	Intervention group: WeChat app; control group: usual care	Interactive format (smartphone–based 2-way communication)	3 days	QOL (cancer-related quality of life), anxiety (GAD-7[s]), and depression (PHQ-9)
Spahrkäs et al [34]; 2020, Australia, Canada, United Kingdom, and United States	CBT	519	280	Mean 56.7 (SD 9.99)	Mean 56.2 (SD 9.42)	Breast cancer, hematological cancer, digestive organ cancer, and others	Not restricted	Surgery, radiation therapy, chemotherapy, immunotherapy, stem cell transplant, hormone therapy, and other treatments	Intervention group: Untire mHealth app; control group: waiting list	Didactic format (smartphone–based 1-way communication)	3 months	QOL (EORTC QLQ-C30)
Sui et al [35], 2020, China	Unclear	100	100	Mean 61.37 (SD 11.21)	Mean 62.35 (SD 9.98)	Non–small cell lung cancer	Stage I to III	Surgery	Intervention group: WeChat app; control group: simple education and rehabilitation guidance	Interactive format (smartphone–based 2-way communication)	12 months	QOL (QLQ-C30), anxiety (HADS-A), and depression (HADS-D)

Author, year, and country	Intervention theory	Sample size, N		Patient characteristics					Intervention	Format of intervention delivery	Intervention duration	Outcomes and outcome measures
		Intervention	Control	Age (years), intervention	Age (years), control	Type of cancer	Stage of cancer	Treatment category				
Zhou et al [36], 2019, China	Roy Adaptation Model	66	66	Mean 44.62 (SD 7.89)	Mean 44.37 (SD 7.32)	Breast cancer	Stage I to III	Surgery	Intervention group: CAT[t]+routine care; control group: routine care	Interactive format (smartphone–based 2-way communication)	3 months	Anxiety (SAS[u]) and depression (SDS[v])
Zhu et al [37], 2018, China	The Bandura self-efficacy theory and the social exchange theory	57	57	Mean 46.2 (SD 8.5)	Mean 48.2 (SD 8.1)	Breast cancer	Stage 0 to III	Chemotherapy	Intervention group: BCS[w]+CAU[x]; control group: CAU	Interactive format (smartphone–based 2-way communication)	6 months	QOL (FACT-B[y]), anxiety (HADS-A), depression (HADS-D), self-efficacy (SICPA[z]), and distress (MDASI[aa])
Di and Li [38], 2018, China	Unclear	65	67	Mean 44.32 (SD 11.03)	Mean 42.28 (SD 10.37)	Nasopharyngeal carcinoma	Stage 0 to IV	Radiotherapy and chemotherapy	Intervention group: smartphone medical app; control group: conventional follow-up visit	Interactive format (smartphone–based 2-way communication)	6 months	QOL (QLQ-C30)
Dong et al [39], 2019, China	Unclear	26	24	Mean 48.00 (SD 5.54)	Mean 51.63 (SD 7.49)	Breast cancer	Stage I to III	Surgery	Intervention group: social media apps; control group: traditional treatment and rehabilitation	Interactive format (smartphone–based 2-way communication)	3 months	QOL (SF-36)
Hou et al [40], 2020, China	Unclear	53	59	N/A[ab]	N/A	Breast cancer	Stage 0 to III	Not restricted	Intervention group: BC-SMS[ac] app + health care; control group: health care	Didactic format (smartphone–based 1-way communication)	3 months	QOL (QLQ-C30)
Lei [41], 2016, China	Orem self-care theory	58	58	N/A	N/A	Laryngeal cancer	Stage 0 to IV	Surgery	Intervention group: Rehab assistant app; control group: usual care	Interactive format (smartphone–based 2-way communication)	3 months	QOL (QLQ-C30)
Rosen et al [42], 2018, United States	Mindfulness training	57	55	Mean 51.40 (SD 10.73)	Mean 53.22 (SD 9.91)	Breast cancer	Stage 0 to IV	Not restricted	Intervention group: app-delivered mindfulness training; control group: waitlist control	Didactic format (smartphone–based 1-way communication)	3 months	QOL (FACT-B)

Author, year, and country	Intervention theory	Sample size, N		Patient characteristics					Intervention	Format of intervention delivery	Intervention duration	Outcomes and outcome measures
		Intervention	Control	Age (years), intervention	Age (years), control	Type of cancer	Stage of cancer	Treatment category				
Zha [43], 2020, China	Unclear	41	41	Mean 45.14 (SD 11.14)	Mean 46.38 (SD 11.57)	Breast cancer	Stage I to II	Surgery	Intervention group: WeChat app care; control group: routine care	Interactive format (smartphone–based 2-way communication)	3 months	QOL (SF-36) and anxiety (STAI)
Absolom et al [44], 2021, United Kingdom	Unclear	256	252	Mean 55.9 (SD 12.2)	Mean 56.0 (SD 11.3)	Breast cancer, colon cancer, and gynecological cancer	Primary or local disease, metastatic	Chemotherapy	Intervention group: eRAPID; control group: routine care	Interactive format (mobile device–based 2-way communication)	18 weeks	QOL (FACT-G) and self-efficacy (Self-Efficacy Scale)
Berg et al [45], 2019, United States	Unclear	38	18	Mean 32.63 (SD 5.87)	Mean 32.39 (SD 4.60)	Breast cancer, lymphoma, and others	Stage 0 to IV	Not restricted	Intervention group: AWAKE; control group: attention control	Interactive format (mobile device–based 2-way communication)	6 months	QOL (QLQ-C30), depression (HADS-D), and self-efficacy (Self-Efficacy Scale)
Chen et al [46], 2021, China	Unclear	40	40	Mean 59.6 (SD 6.5)	Mean 59.8 (SD 7.0)	Esophageal cancer	Stage I to IIIa	Surgery	Intervention group: WeChat; control group: routine care	Interactive format (WeChat group–based 2-way communication)	3 months	QOL (QLQ-C30)
Huggins et al [47], 2022, Australia	CBT	36	37	Mean 66.6 (SD 9.7)	Mean 63.2 (SD 9.9)	Nasopharyngeal carcinoma	Not restricted	Not restricted	Intervention group: myPace; control group: routine care	Interactive format (app-based 2-way communication)	12 months	QOL (QLQ-C30)
Maguire et al [48], 2021, United Kingdom	Unclear	415	414	Mean 51.9 (SD 12.4)	Mean 52.9 (SD 12.1)	Breast cancer and colon cancer	Not restricted	Chemotherapy	Intervention group: ASyMS; control group: standard care	Didactic format (smartphone–based 1-way communication)	12 weeks	QOL (QLQ-C30), self-efficacy (CASE-cancer[ad]), and anxiety (STAI)

Author, year, and country	Intervention theory	Sample size, N		Patient characteristics					Intervention	Format of intervention delivery	Intervention duration	Outcomes and outcome measures
		Intervention	Control	Age (years), intervention	Age (years), control	Type of cancer	Stage of cancer	Treatment category				
Seib et al [49], 2022, Australia	CBT	175	176	Mean 52.6 (SD 9.4)	Mean 53.7 (SD 8.1)	Breast cancer, gynecological cancer, and blood cancer	Not restricted	Not restricted	Intervention group: WWACP; control group: standard care	Interactive format (app-based 2-way communication)	12 weeks	QOL (SF-36)

[a]HADS-A: Hospital Anxiety and Depression Scale-Anxiety subscale.

[b]HADS-D: Hospital Anxiety and Depression Scale-Depression subscale.

[c]HRQOL: health-related quality of life.

[d]SF-36: 36-item Short Form Health Survey.

[e]mHealth: mobile health.

[f]QOL: quality of life.

[g]FACT-ES QLS: Functional Assessment of Cancer Therapy-Endocrine Symptoms Quality of Life Scale.

[h]NCCN-DT: National Comprehensive Cancer Network Distress Thermometer.

[i]MSAS-GDI: Memorial Symptom Assessment Scale-General Distress Index.

[j]EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30.

[k]STAI: State-Trait Anxiety Inventory.

[l]CBT: cognitive behavioral therapy.

[m]PHQ-9: Patient Health Questionnaire-9.

[n]FACT-G: Functional Assessment of Cancer Therapy-General.

[o]BDI-II: Beck Depression Inventory-Second Edition.

[p]WHOQOL-BREF: World Health Organization Quality of Life-BREF questionnaire.

[q]BDI: Beck Depression Inventory.

[r]FACIT - Pal: Functional Assessment of Chronic Illness Therapy-Palliative Care.

[s]GAD-7: Generalized Anxiety Disorder-7.

[t]CAT: cyclic adjustment training.

[u]SAS: Self-Rating Anxiety Scale.

[v]SDS: Self-Rating Depression Scale.

[w]BCS: breast cancer e-support.

[x]CAU: care as usual.

[y]FACT-B: Functional Assessment of Cancer Therapy-B.

[z]SICPA: Stanford Inventory of Cancer Patient Adjustment.

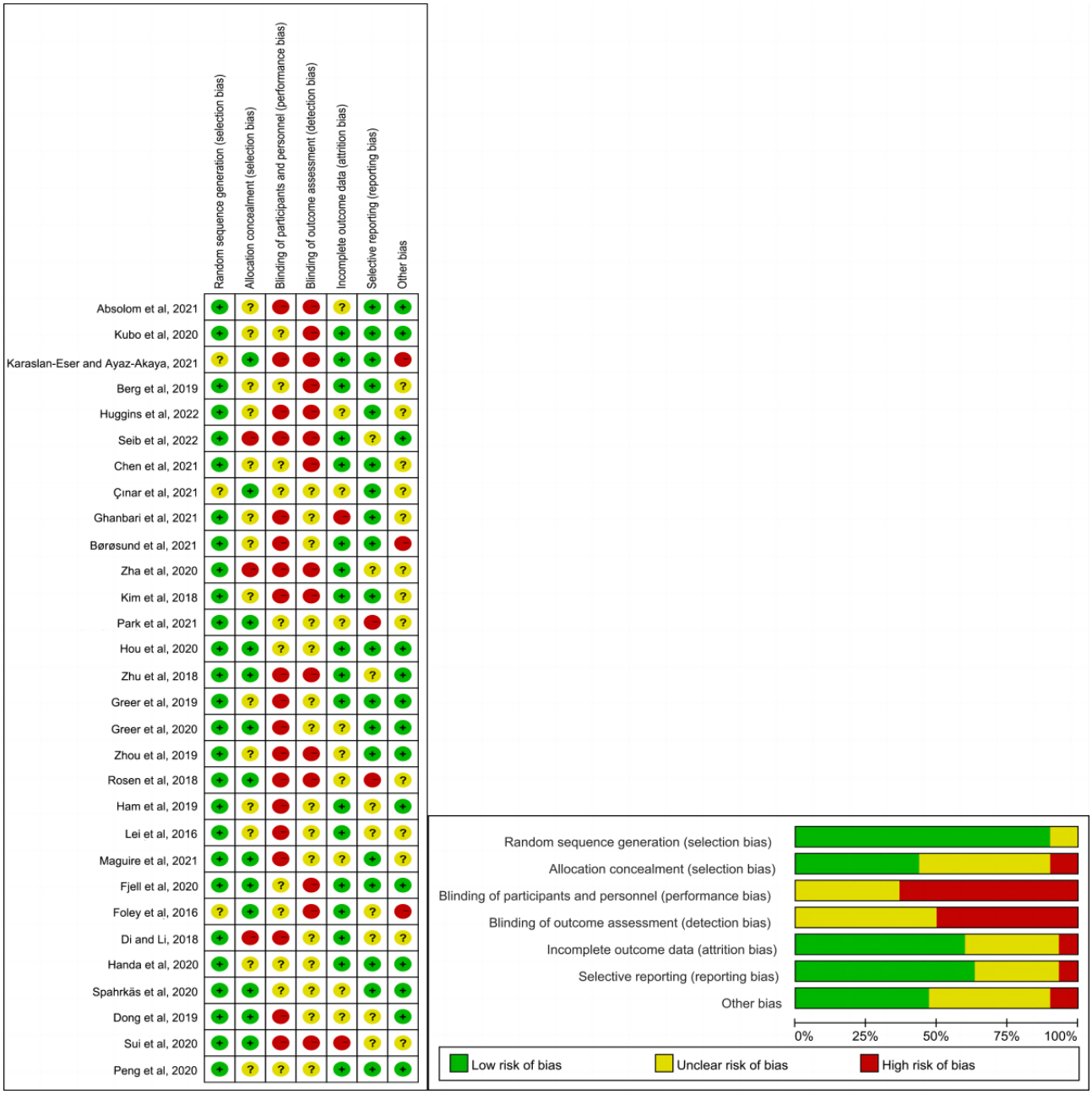
[aa]MDASI: MD Anderson Symptom Inventory.

[ab]N/A: not applicable.

[ac]BCSMS: Breast Cancer Self-Management Support.

[ad]CASE-cancer: Communication and Attitudinal Self-Efficacy scale for cancer.

Figure 2. Risk-of-bias summary and graph [16-20,25-49].



Functions of Smartphone Apps

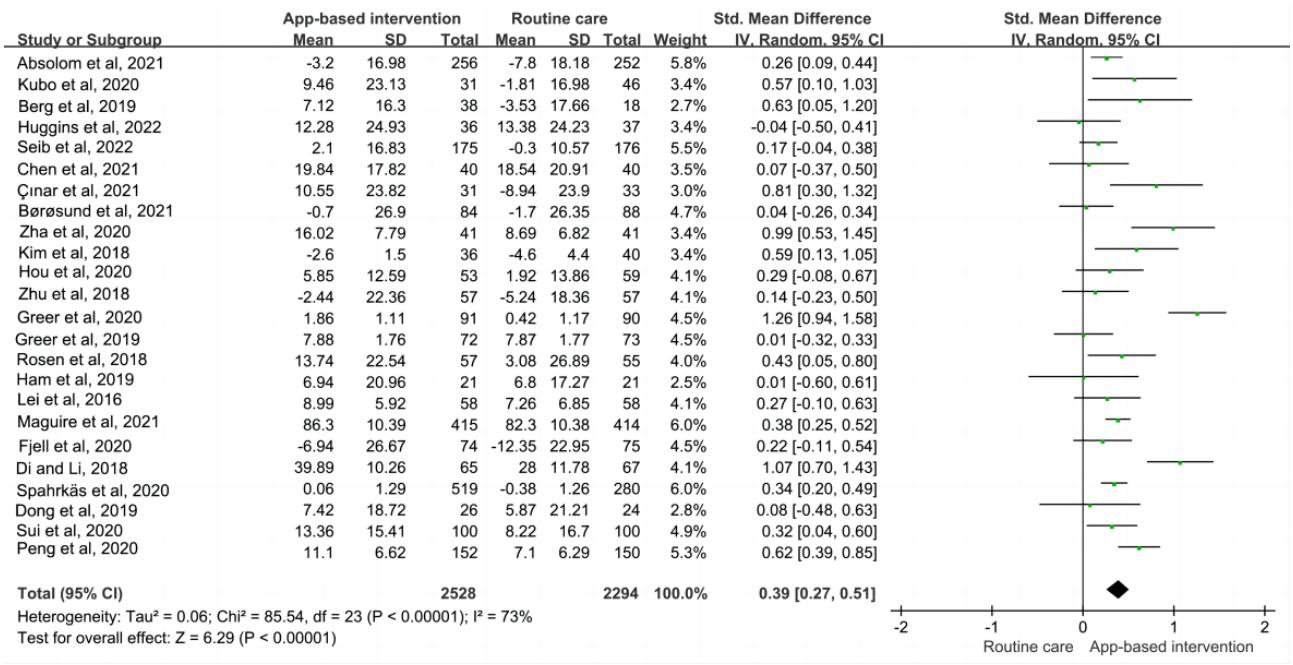
The functions of these apps can be classified as follows: provision of health education and advice, physician-patient communication via the mHealth app, and data management regarding self-management behaviors of patients with cancer (including data upload, visualization, and reminder services). Physicians and patients interact in 2 ways: the app generates automated feedback based on predesigned personalized feedback, and medical professionals issue interactive guidance based on patient-provided personalized data. Most (22/30, 73%) of these studies incorporated personalized guidance services provided by health care professionals who analyzed patient data and communicated with the patients via SMS text message, phone, or video.

Effects on QOL

A total of 80% (24/30) of the studies [16,19,25-30,33-35,37-49] involving 4822 participants used various scales to report the outcome of QOL. Of these 24 studies, 8 (33%) [19,26,27,37,39,40,42,43] focused on patients with breast cancer, and the other 16 (67%) included patients with multiple types of cancer (such as breast cancer, brain cancer, and prostate cancer). A total of 62% (15/24) of the studies had an intervention duration of <3 months, and the remainder had an intervention duration of 3 to 12 months. The apps used different intervention theories (including cognitive behavioral therapy, psychoeducation, and mindfulness-based stress reduction); 25% (6/24) of the studies used cognitive behavioral therapy interventions, and 8% (2/24) of the studies were based on mindfulness-based therapy. In these studies, patients with cancer received different treatment strategies; 29% (7/24) of the studies were conducted only among patients under chemotherapy, and

25% (6/24) were conducted only in patients undergoing surgery. Owing to the significant heterogeneity among the studies (P<.001; I^2=77%), the results were pooled using a random-effects model. Overall, the mHealth app interventions significantly improved cancer-related QOL scores (SMD=0.39, 95% CI 0.27-0.51; P<.001; Figure 3 [16,19,25-30,33-35,37-49]).

Figure 3. Meta-analysis on quality of life [16,19,25-46]. IV: inverse variance; Std: standardized.



We conducted subgroup analyses according to intervention duration, type of cancer, intervention theory, treatment strategy, and intervention delivery format to investigate potential sources of heterogeneity. Pooled results for the short-term (≤3 months) follow-up period suggested that mHealth app medical interventions were effective in improving QOL ($SMD_{<3\ months}$=0.41, 95% CI 0.26-0.57; P=.001; $SMD_{3\ to\ 12\ months}$=0.36, 95% CI 0.14-0.57; P=.001; Table 2 and Multimedia Appendix 3 [16,19,25-30,33-35,37-49]). When studies were grouped by type of cancer, the results showed that mHealth app interventions may improve cancer-related QOL scores across cancer types ($SMD_{Breast\ cancer}$=0.42, 95% CI 0.21-0.63; P<.001; $SMD_{Various\ cancers}$=0.38, 95% CI 0.23-0.53; P=.001; Table 2 and Multimedia Appendix 4 [16,19,25-30,33-35,37-49]). Subgroup analyses of different intervention theories revealed low heterogeneity for cognitive behavioral theory (35%) and mindfulness-based theory (0%), implying that different intervention theories may be an important source of heterogeneity (Table 2 and Multimedia Appendix 5 [16,19,25-30,33-35,37-49]). Studies grouped by intervention delivery format revealed that these interventions significantly improved cancer-related QOL scores across different intervention delivery formats ($SMD_{Interactive\ format}$=0.36, 95% CI 0.22-0.50; P<.001; $SMD_{Didactic\ format}$=0.48, 95% CI 0.22-0.73; P<.001; Table 2 and Multimedia Appendix 6 [16,19,25-30,33-35,37-49]). There were no significant differences in QOL scores, but there was a high heterogeneity among patients with cancer receiving different treatment modalities (Table 2 and Multimedia Appendix 7 [16,19,25-30,33-35,37-49]).

In the sensitivity analysis, switching from a random-effects model to a fixed-effects model confirmed the effect of the app interventions (SMD=0.43, 95% CI 0.35-0.50; P<.001). Furthermore, when each study was excluded sequentially, the pooled estimates remained robust, ranging from 0.38 (95% CI 0.30-0.45) to 0.46 (95% CI 0.37-0.54). There was no evidence of publication bias (Begg test: P=.65; Egger test: P=.67; Multimedia Appendix 8). Therefore, the pooled estimate for QOL was robust.

Table 2. Subgroup analyses of quality of life (N=24).

Stratification	Studies, n (%)	*P* value for heterogeneity	I^2 (%)	Pooled standardized mean difference (95% CI)	*P* value for pooled results
Intervention duration (months)					
<3	15 (62)	<.001	75	0.41 (0.26-0.57)	.001[a]
3 to 12	9 (38)	<.001	72	0.36 (0.14-0.57)	.001[a]
Types of cancer					
Breast cancer	8 (33)	.05	51	0.42 (0.21-0.63)	.001[a]
Various cancers	16 (67)	<.001	79	0.38 (0.23-0.53)	.001[a]
Intervention theory					
Cognitive behavioral theory	6 (25)	.17	35	0.16 (0.01-0.30)	.03[a]
Mindfulness-based theory	2 (8)	.65	0	0.48 (0.19-0.77)	.01[a]
Other theories	16 (67)	<.001	76	0.49 (0.33-0.66)	.001[a]
Format of intervention delivery					
Interactive format (smartphone–based 2-way communication)	18 (75)	<.001	68%	0.36 (0.22-0.50)	<.001[a]
Didactic format (smartphone–based 1-way communication)	6 (25)	<.001	83%	0.48 (0.22-0.73)	<.001[a]
Treatment category					
Patients for chemotherapy	7 (29)	<.001	87%	0.55 (0.27-0.82)	<.001[a]
Patients for surgery	6 (25)	.02	62%	0.41 (0.13-0.69)	.004[a]
Patients for various treatments	11 (46)	.02	53%	0.28 (0.14-0.42)	.007[a]

[a]*P*<.05.

Effects on Anxiety

A total of 47% (14/30) of the studies [16-20,25,28,30,33,35-37,43,48] measured anxiety scores using different scales. Overall, the mHealth app interventions significantly alleviated anxiety among cancer survivors, but there was high heterogeneity (SMD=−0.64, 95% CI −0.73 to −0.56; *P*<.001; I^2=97%; Figure 4 [16-20, 25-28, 30-33, 35-37, 43-45, 48]).

On the basis of groups of intervention duration, 79% (11/14) of the studies had an intervention duration of <3 months, and 21% (3/14) had an intervention duration of 3 to 12 months. Subgroup analyses showed that these app-based interventions were still effective with different intervention durations. A total of 50% (7/14) of the studies [17-20,37,43,48] compared anxiety scores among breast cancer survivors and showed poor app intervention outcomes (SMD=−0.87, 95% CI −1.79 to 0.05; *P*=.06; I^2=96%). Subgroup analyses by different intervention theories revealed high heterogeneity among interventions based on cognitive behavioral theory, but these were still effective in alleviating anxiety. Furthermore, subgroup analyses revealed that mHealth app interventions with an interactive format significantly reduced cancer-related anxiety scores (SMD=−1.27, 95% CI −1.99 to −0.56; *P*=.001; I^2=97%). When studies were grouped by treatment strategy, app interventions did not alleviate anxiety in patients in chemotherapy (SMD=−0.06, 95% CI −0.32 to 0.19; *P*=.62; I^2=60%) but could alleviate anxiety in patients undergoing surgery or comprehensive treatment (Table 3).

We found no significant change in the pooled estimates when single studies were excluded sequentially and the pooled model was changed. No evidence of publication bias was found (Begg test: *P*=.69; Egger test: *P*=.30; Multimedia Appendix 8). Therefore, the pooled estimate for anxiety was robust.

Figure 4. Meta-analysis on (A) anxiety, (B) depression, (C) distress, and (D) self-efficacy [16-20,25-28,30,31,33,34,40-42,45,47-49]. IV: inverse variance; Std: standardized.

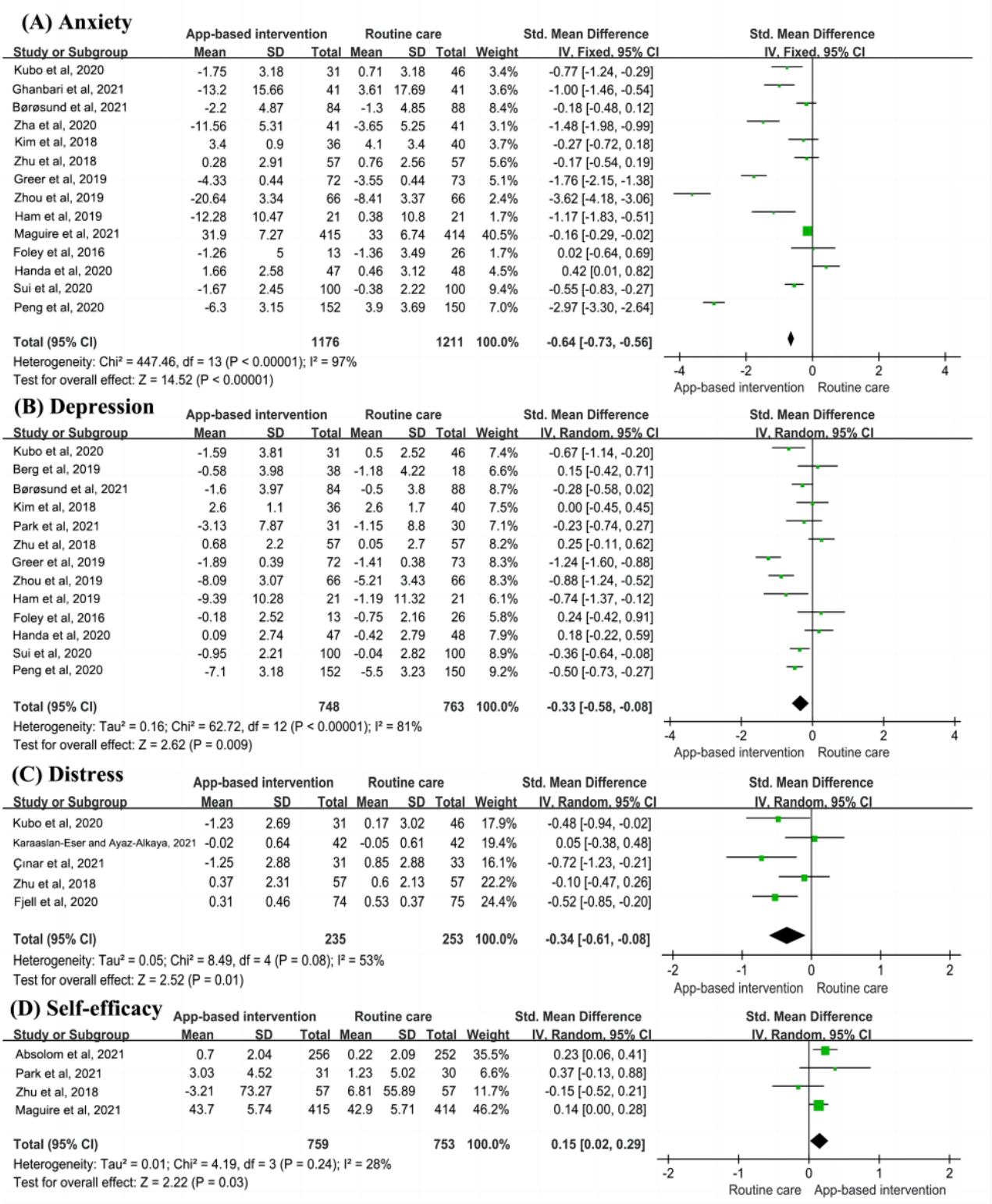


Table 3. Subgroup analyses of anxiety (N=14).

Stratification	Studies, n (%)	*P* value for heterogeneity	I^2 (%)	Pooled standardized mean difference (95% CI)	*P* value for pooled results
Intervention duration (months)					
<3	11 (79)	<.001	98	−1.16 (−1.91 to −0.41)	.002[a]
3 to 12	3 (21)	.14	49	−0.32 (−0.57 to −0.06)	.01[a]
Types of cancer					
Breast cancer	7 (50)	<.001	96	−0.87 (−1.79 to 0.05)	.06
Various cancers	7 (50)	<.001	97	−1.07 (−1.86 to −0.29)	.006[a]
Intervention theory					
Cognitive behavioral theory	4 (29)	<.001	93	−1.02 (−1.81 to −0.23)	.01[a]
Other theories	10 (71)	<.001	98	−0.95 (−1.68 to −0.23)	.01[a]
Format of intervention delivery					
Interactive format (smartphone–based 2-way communication)	10 (71)	<.001	97	−1.27 (−1.99 to −0.56)	.001[a]
Didactic format (smartphone–based 1-way communication)	3 (21)	<.001	82	−0.17 (−0.67 to 0.32)	.49
Treatment category					
Patients for chemotherapy	4 (29)	.06	60	−0.06 (−0.32 to 0.19)	.62
Patients for surgery	4 (29)	<.001	97	−1.41 (−2.81 to −0.01)	.04[a]
Patients for various treatments	6 (43)	<.001	97	−1.31 (−2.26 to −0.36)	.007[a]

[a]*P*<.05.

Effects on Depression

The meta-analysis for depression included 1511 patients from 43% (13/30) of the studies [16,18-20,25,28,30,32,33,35-37,45]. A random-effects model was chosen for analysis owing to the significant heterogeneity among the 43% (13/30) of the studies (*P*<.001; I^2=81%). The pooled results indicated that the mHealth app intervention group had a lower depression score than the routine care group (SMD=−0.33, 95% CI −0.58 to −0.08; *P*=.009; Figure 4).

Grouping by intervention duration, 69% (9/13) of the studies had an intervention duration of <3 months, and 31% (4/13) had an intervention duration of 3 to 12 months. Subgroup analyses showed that these app-based interventions were effective with durations of <3 months but not with a duration of 3 to 12 months (SMD=−0.25, 95% CI −0.51 to 0.02; *P*=.07; I^2=53%). When studies were grouped by type of cancer, 46% (6/13) of the studies involved breast cancer survivors [21-23,43,47,48], and mHealth app interventions did not alleviate depression in these survivors (SMD=−0.11, 95% CI −0.27 to 0.06; *P*=.21). Subgroup analyses according to intervention theory revealed that cognitive behavioral theory–based interventions could effectively relieve depression in cancer survivors (SMD=−0.75, 95% CI −1.42 to 0.09; *P*=.03), but there was high heterogeneity. A subgroup analysis revealed that mHealth app interventions with an interactive format significantly reduced cancer-related depression (SMD=−0.41, 95% CI −0.70 to −0.12; *P*=.006), but didactic format interventions were not effective in improving depression scores (SMD=−0.12, 95% CI −0.54 to 0.30; *P*=.58). When studies were grouped by treatment strategy, researchers found that app interventions did not alleviate depression in survivors who were treated with chemotherapy and surgery but could alleviate depression in survivors with comprehensive treatment (SMD=−0.56, 95% CI −0.90 to −0.21; *P*=.001; I^2=79%; Table 4).

The fixed-effects model produced the same outcome as the random-effects model in the sensitivity analysis. In addition, when using a single-study approach, we found no studies that significantly altered the pooled results. No significant publication bias was found (Begg test: *P*=.58; Egger test: *P*=.49; Multimedia Appendix 8).

Table 4. Subgroup analyses of depression (N=13).

Stratification	Studies, n (%)	*P* value for heterogeneity	I^2 (%)	Pooled standardized mean difference (95% CI)	*P* value for pooled results
Intervention duration (months)					
<3	9 (69)	<.001	82	−0.45 (−0.77 to −0.13)	.006[a]
3 to 12	4 (31)	.09	53	−0.25 (−0.51 to 0.02)	.07
Types of cancer					
Breast cancer	7 (54)	<.001	77	−0.11 (−0.27 to 0.06)	.21
Various cancers	6 (46)	.001	75	−0.55 (−0.68 to −0.42)	.006[a]
Intervention theory					
Cognitive behavioral theory	3 (23)	<.001	88	−0.75 (−1.42 to −0.09)	.03[a]
Other theories	10 (77)	<.001	75	−0.21 (−0.46 to 0.04)	.10
Format of intervention delivery					
Interactive format (smartphone–based 2-way communication)	9 (69)	<.001	84	−0.41 (−0.70 to −0.12)	.006[a]
Didactic format (smartphone–based 1-way communication)	4 (31)	.07	58	−0.12 (−0.54 to 0.30)	.58
Treatment category					
Patients for chemotherapy	3 (23)	.69	0	0.16 (−0.07 to 0.40)	.17
Patients for surgery	4 (31)	.01	72	−0.37 (−0.77 to 0.03)	.07
Patients for various treatments	6 (46)	.002	79	−0.56 (−0.90 to −0.21)	.001[a]

[a]*P*<.05.

Effects on Distress

The meta-analysis of distress included 17% (5/30) of the studies [16,26,27,31,43] with a total of 488 cancer survivors. As there was heterogeneity among the studies (*P*=.08; I^2=53%), a random-effects model was used to pool the results. Overall, the mHealth app interventions significantly alleviated distress among cancer survivors (SMD=−0.34, 95% CI −0.61 to −0.08; *P*=.01; Figure 4). To assess the robustness of the pooled results, we performed sensitivity analyses using various pooled models. The pooled results of the fixed-effects model also showed that the app intervention group had lower distress scores than the usual care group (SMD=−0.34, 95% CI −0.52 to −0.16; *P*=.006), indicating that the pooled effect size was robust. Publication bias was not examined as <10 studies were included.

Effects on Self-efficacy

A total of 13% (4/30) of the studies [32,37,44,45] reported self-efficacy as an outcome. Pooling of studies showed a statistically significant effect size favoring the intervention group (SMD=0.15, 95% CI 0.02-0.29; *P*=.03; I^2=28%). The fixed-effect model also showed that app interventions had higher self-efficacy scores than usual care (SMD=0.16, 95% CI 0.06-0.26; *P*=.008; Figure 4).

Discussion

Principal Findings

Currently, the medical pattern is changing from a biomedical pattern (the treatment of disease only focusing on the patient's physical function) to a biopsychosocial medical pattern (the treatment of disease with comprehensive consideration of the patient's physical function, mental health, and social environment). Thus, greater attention is being paid to patients' mental health and social functioning. Among cancer survivors, symptoms such as depression, anxiety, distress, and pain are prevalent and undertreated, which may negatively affect their QOL and self-efficacy. However, smartphone users are increasing worldwide and are expected to reach 6.8 billion by 2023, with a smartphone penetration rate of 53.8% [50]. Furthermore, smartphone apps have natural advantages over websites, SMS text messages, and other similar communication methods owing to their personalized design, rich mobile device features (such as cameras, phones, GPS, and contact lists), and timely push features. Therefore, the use of smartphone health apps could be a potentially effective way to improve mental health and social functioning among patients with cancer.

We included 30 RCTs in this meta-analysis, and all studies (30/30, 100%) provided smartphone app interventions for cancer survivors. The pooled results showed that smartphone app–based interventions improved QOL (SMD=0.39; *P*<.001) and self-efficacy (SMD=0.15; *P*=.03) in cancer survivors compared with conventional care education and significantly reduced

adverse psychological outcomes (anxiety, depression, and distress). In particular, short-term interventions (duration of ≤3 months), physician-patient interaction interventions (2-way communication using a smartphone app), and cognitive behavioral therapy–based interventions might be most effective for improving QOL and alleviating adverse psychological effects.

Interpretation of Findings

The effect of mHealth app interventions on QOL, anxiety, depression, distress, and self-efficacy in adult cancer survivors over a median follow-up time of 2.8 months was consistent with recent results regarding cell phone, SMS text message, and web-based interventions [21,22,26]. This effect can be attributed to the prevalence and inherent advantages of smartphones. Compared with routine care, app-based interventions can provide more visually based and vivid educational counseling, enabling patients to establish close and ongoing contact with their treatment team [51,52]. Furthermore, with such an intervention, cancer survivors may become more aware of their condition and learn to cope with some of the problems associated with cancer [51]; as a result, patients may have a greater sense of empowerment and willingness to care for themselves, thereby improving their QOL and alleviating adverse psychological effects [53]. In addition, as a high financial burden is associated with a low QOL and high anxiety in cancer survivors [54], app interventions can help reduce health care costs, further improving patients' QOL and alleviating adverse psychological effects [55].

In this review, we conducted subgroup analyses according to intervention duration, type of cancer, intervention theory, treatment category, and intervention delivery format. We found that the short-term effects of app interventions on QOL and psychological outcomes (median follow-up period of 2.8 months) were superior to the long-term effects, which were inconsistent for QOL, anxiety, and depression. This may be influenced by the progression, vulnerability, and persistence of cancer itself. However, this highlights the need for further research to test the effectiveness of mHealth interventions over the long term. Pooled results from studies on patients with breast cancer found that, although tending to alleviate anxiety and depression (SMD <0), app interventions did not significantly improve patients' anxiety and depression status. In female patients, the rich cancer information within an app may remind them of what they are experiencing, leading to increased anxiety and depression [18]. Therefore, clinical practitioners should further explore appropriate care for patients with breast cancer based on evidence-based research and cognitive behavioral therapy. Among the different formats of intervention delivery, most (22/30, 73%) studies used app monitoring combined with feedback interventions, which significantly improved patients' anxiety and depression. On the one hand, cancer survivors may become more aware of their condition through disease self-monitoring and learning to cope with some cancer-related problems [51]. By contrast, by conducting physician-patient communication via an app, patients with cancer may develop a close and ongoing partnership with their treatment team and communicate more effectively regarding disease progression or treatment complications. However, the effectiveness of educational message delivery may depend on how easily the patient understands the content and the importance of the message. Therefore, interventions in a didactic format to deliver educational messages have not been effective in alleviating anxiety and depression. Our review showed that cognitive behavioral therapy was effective in improving QOL and alleviating adverse psychological effects among cancer survivors. This result is consistent with those of other studies [56,57]. A possible explanation is that cognitive behavioral therapy interventions for patients address a broad range of aspects, such as physical, psychological, and social aspects, which can improve QOL and alleviate adverse psychological effects. However, relevant studies have been conducted among patients with cancer using an app, which cannot be compared directly with breast cancer treatment. Therefore, these results should be interpreted with caution.

The results of this meta-analysis indicated a significant improvement in QOL among adult cancer survivors who received chemotherapy. This was similar to the findings of 2 previous meta-analyses, which also found a significant improvement in QOL [21,24]. However, the intervention effects on anxiety and depression remain unclear as there was no significant difference between the intervention and control groups for both outcomes. A total of 2 previous meta-analyses regarding the effects of mobile phone–based interventions on anxiety and depression in this patient population also yielded contrasting results [21,24]. One study found that anxiety but not depression was significantly reduced [24], whereas the other study reported inverse findings [21]. These inconsistencies point to the need for further research to test the effectiveness of mHealth interventions on anxiety and depression in patients with cancer.

Study Limitations

This study had some limitations. First, the included studies had qualitative and methodological weaknesses. Most studies failed to elucidate the processes of allocation concealment (17/30, 57%), researcher or participant blinding (19/30, 63%), and strategies for handling incomplete outcome data. Therefore, the design of allocation concealment, participant blinding, and outcome assessment should be emphasized in future studies to draw more credible conclusions. Second, there is a huge variation in the conceptualization and operationalization of patient participation, which makes data synthesis extremely difficult. The effects of app interventions should be interpreted with caution owing to the high heterogeneity in the operational definitions of measurement instruments and instrument scoring systems. However, this meta-analysis included only RCTs and used random-effects models to pool results when appropriate to yield the most conservative estimates. Subgroup and sensitivity analyses were also performed, and the results showed that the pooled estimates were relatively robust. In addition, because of the limitations of the included studies, we did not conduct subgroup analyses on the frequency of physician-patient interactions via apps; previous studies suggested that app interaction frequency leads to different effects [58]. Therefore, further studies should be conducted on interaction frequency. Finally, the extraction and classification of interventions is challenging because of considerable heterogeneity in the design

of the interventions. The risk of misclassification of intervention characteristics and the exploratory nature of our subgroup analyses prevented us from drawing reliable conclusions about the characteristics of effective interventions.

Implications

Our findings have several important implications. First, at a median follow-up time of 2.8 months, mobile app interventions may have a significant effect on enhancing QOL in cancer survivors and alleviating anxiety, depression, and distress in these patients. However, there is an urgent need to assess the long-term effects of these interventions on QOL and psychological outcomes. Second, using a physician-patient interaction intervention is more likely to significantly improve QOL and psychological effects. Future clinical research should further explore care modalities of patients with cancer based on the physician-patient interaction format. Third, cognitive behavioral therapy interventions address many aspects, such as physical, psychological, and social aspects, which improves QOL and alleviates adverse psychological effects. In the future, the development of mHealth apps that are based on cognitive behavioral theory should be encouraged. Fourth, clinical practitioners should further explore appropriate care strategies for breast cancer survivors. Fifth, it is difficult to identify patterns of patient engagement with smartphone app–based interventions because of the wide variability in intervention design and measurement tool scoring systems among the studies. By exploring factors such as participant characteristics and active engagement, further insights can be gained into strategies that can help increase patients' motivation to participate and maintain intervention integrity.

Conclusions

This review showed that smartphone app–based interventions might help address certain psychological issues experienced by cancer survivors. In particular, short-term interventions (duration of ≤3 months), physician-patient interaction interventions (2-way communication using a smartphone app), and cognitive behavioral therapy–based interventions might be more effective in improving QOL and alleviating adverse psychological effects. However, the evidence supporting these interventions is still being gathered and is not yet fully conclusive. Further rigorous and well-designed studies are warranted to address the methodological flaws identified in this review. In conclusion, mHealth interventions may be effective in providing psychological support for adult cancer survivors.

Acknowledgments

The study was partly supported by the Science and Technology Plan Project of Xiangyang (grant 0001-36) and also partly supported by the clinical trial center of Xiangyang Central Hospital is affiliated with Hubei University of Arts and Sciences (grant 2022YB19). The funding body had no role in the study design, data collection and analysis, and interpretation of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Searching strategy.
[DOCX File , 31 KB - jmir_v24i12e39799_app1.docx]

Multimedia Appendix 2
Risk of bias.
[DOCX File , 22 KB - jmir_v24i12e39799_app2.docx]

Multimedia Appendix 3
Pooled results for quality of life grouped by intervention duration. IV: inverse variance; Std: standardized.
[PNG File , 480 KB - jmir_v24i12e39799_app3.png]

Multimedia Appendix 4
Pooled results for quality of life grouped by type of cancer. IV: inverse variance; Std: standardized.
[PNG File , 481 KB - jmir_v24i12e39799_app4.png]

Multimedia Appendix 5
Pooled results for quality of life grouped by intervention theory. IV: inverse variance; Std: standardized.
[PNG File , 520 KB - jmir_v24i12e39799_app5.png]

Multimedia Appendix 6
Pooled results for quality of life grouped by format of intervention delivery. IV: inverse variance; Std: standardized.
[PNG File , 481 KB - jmir_v24i12e39799_app6.png]

Multimedia Appendix 7
Pooled results for quality of life grouped by treatment strategy. IV: inverse variance; Std: standardized.
[PNG File , 523 KB - jmir_v24i12e39799_app7.png]

Multimedia Appendix 8
Funnel plots of quality of life, anxiety, and depression. SMD: standardized mean difference.
[PNG File , 71 KB - jmir_v24i12e39799_app8.png]

References

1. Cancer prevention and control in the context of an integrated approach: report by the Secretariat. World Health Organisation. 2016. URL: https://apps.who.int/gb/ebwha/pdf_files/EB140/B140_31-en.pdf [accessed 2022-11-30]
2. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, CONCORD Working Group. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 2018 Mar 17;391(10125):1023-1075 [FREE Full text] [doi: 10.1016/S0140-6736(17)33326-3] [Medline: 29395269]
3. Grabsch B, Clarke DM, Love A, McKenzie DP, Snyder RD, Bloch S, et al. Psychological morbidity and quality of life in women with advanced breast cancer: a cross-sectional survey. Palliat Support Care 2006 Mar 29;4(1):47-56. [doi: 10.1017/s1478951506060068] [Medline: 16889323]
4. Linden W, Vodermaier A, Mackenzie R, Greig D. Anxiety and depression after cancer diagnosis: prevalence rates by cancer type, gender, and age. J Affect Disord 2012 Dec 10;141(2-3):343-351. [doi: 10.1016/j.jad.2012.03.025] [Medline: 22727334]
5. Mitchell AJ, Chan M, Bhatti H, Halton M, Grassi L, Johansen C, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. Lancet Oncol 2011 Feb;12(2):160-174. [doi: 10.1016/s1470-2045(11)70002-x]
6. Meeting Psychosocial Needs of Women with Breast Cancer. Washington (DC): National Academies Press (US); 2004.
7. Omran S, Saeed A, Simpson J. Symptom distress of Jordanian patients with cancer receiving chemotherapy. Int J Nurs Pract 2012 Apr;18(2):125-132. [doi: 10.1111/j.1440-172X.2012.02012.x] [Medline: 22435975]
8. Khan F, Amatya B, Pallant JF, Rajapaksa I. Factors associated with long-term functional outcomes and psychological sequelae in women after breast cancer. Breast 2012 Jun;21(3):314-320 [FREE Full text] [doi: 10.1016/j.breast.2012.01.013] [Medline: 22342676]
9. Furness K, Sarkies MN, Huggins CE, Croagh D, Haines TP. Impact of the method of delivering electronic health behavior change interventions in survivors of cancer on engagement, health behaviors, and health outcomes: systematic review and meta-analysis. J Med Internet Res 2020 Jun 23;22(6):e16112 [FREE Full text] [doi: 10.2196/16112] [Medline: 32574147]
10. Ashing K, Rosales M. A telephonic-based trial to reduce depressive symptoms among Latina breast cancer survivors. Psychooncology 2014 May 11;23(5):507-515. [doi: 10.1002/pon.3441] [Medline: 24217994]
11. Smith SK, Kuhn E, O'Donnell J, Koontz BF, Nelson N, Molloy K, et al. Cancer distress coach: pilot study of a mobile app for managing posttraumatic stress. Psychooncology 2018 Jan 02;27(1):350-353. [doi: 10.1002/pon.4363] [Medline: 28032921]
12. Lee H, Ghebre R, Le C, Jang YJ, Sharratt M, Yee D. Mobile phone multilevel and multimedia messaging intervention for breast cancer screening: pilot randomized controlled trial. JMIR Mhealth Uhealth 2017 Nov 07;5(11):e154 [FREE Full text] [doi: 10.2196/mhealth.7091] [Medline: 29113961]
13. Keum J, Chung MJ, Kim Y, Ko H, Sung MJ, Jo JH, et al. Usefulness of smartphone apps for improving nutritional status of pancreatic cancer patients: randomized controlled trial. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e21088 [FREE Full text] [doi: 10.2196/21088] [Medline: 34463630]
14. Galiano-Castillo N, Arroyo-Morales M, Lozano-Lozano M, Fernández-Lao C, Martín-Martín L, Del-Moral-Ávila R, et al. Effect of an internet-based telehealth system on functional capacity and cognition in breast cancer survivors: a secondary analysis of a randomized controlled trial. Support Care Cancer 2017 Nov 22;25(11):3551-3559. [doi: 10.1007/s00520-017-3782-9] [Medline: 28639097]
15. Okunade KS, Salako O, Adejimi AA, Akinsola OJ, Fatiregun O, Adenekan MA, et al. Impact of mobile technologies on cervical cancer screening practices in Lagos, Nigeria (mHealth-Cervix): protocol for a randomised controlled trial. F1000Res 2020 May 4;9:322 [FREE Full text] [doi: 10.12688/f1000research.22991.1] [Medline: 32528665]
16. Kubo A, Kurtovich E, McGinnis M, Aghaee S, Altschuler A, Quesenberry C, et al. Pilot pragmatic randomized trial of mHealth mindfulness-based intervention for advanced cancer patients and their informal caregivers. Psychooncology 2020 Sep 26 (forthcoming). [doi: 10.1002/pon.5557] [Medline: 32979294]
17. Ghanbari E, Yektatalab S, Mehrabi M. Effects of psychoeducational interventions using mobile apps and mobile-based online group discussions on anxiety and self-esteem in women with breast cancer: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 18;9(5):e19262 [FREE Full text] [doi: 10.2196/19262] [Medline: 34003138]

18. Handa S, Okuyama H, Yamamoto H, Nakamura S, Kato Y. Effectiveness of a smartphone application as a support tool for patients undergoing breast cancer chemotherapy: a randomized controlled trial. Clin Breast Cancer 2020 Jun;20(3):201-208. [doi: 10.1016/j.clbc.2020.01.004] [Medline: 32201165]
19. Kim HJ, Kim SM, Shin H, Jang J, Kim YI, Han DH. A mobile game for patients with breast cancer for chemotherapy self-management and quality-of-life improvement: randomized controlled trial. J Med Internet Res 2018 Oct 29;20(10):e273 [FREE Full text] [doi: 10.2196/jmir.9559] [Medline: 30578205]
20. Foley N, O'Connell E, Lehane E, Livingstone V, Maher B, Kaimkhani S, et al. PATI: patient accessed tailored information: a pilot study to evaluate the effect on preoperative breast cancer patients of information delivered via a mobile application. Breast 2016 Dec;30:54-58. [doi: 10.1016/j.breast.2016.08.012] [Medline: 27611236]
21. Chen Y, Guan B, Li Z, Li X. Effect of telehealth intervention on breast cancer patients' quality of life and psychological outcomes: a meta-analysis. J Telemed Telecare 2017 Jan 12;24(3):157-167. [doi: 10.1177/1357633x16686777]
22. Li J, Liu Y, Jiang J, Peng X, Hu X. Effect of telehealth interventions on quality of life in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2021 Oct;122:103970. [doi: 10.1016/j.ijnurstu.2021.103970] [Medline: 34303269]
23. Akingbade O, Nguyen K, Chow K. Effect of mHealth interventions on psychological issues experienced by women undergoing chemotherapy for breast cancer: a systematic review and meta-analysis. J Clin Nurs 2022 (forthcoming). [doi: 10.1101/2022.02.17.22271140]
24. Zhang Q, Zhang L, Yin R, Fu T, Chen H, Shen B. Effectiveness of telephone-based interventions on health-related quality of life and prognostic outcomes in breast cancer patients and survivors-A meta-analysis. Eur J Cancer Care (Engl) 2018 Jan 16;27(1):e12632. [doi: 10.1111/ecc.12632] [Medline: 28090704]
25. Børøsund E, Ehlers SL, Clark MM, Andrykowski MA, Cvancarova Småstuen M, Solberg Nes L. Digital stress management in cancer: testing StressProffen in a 12-month randomized controlled trial. Cancer 2022 Apr 01;128(7):1503-1512. [doi: 10.1002/cncr.34046] [Medline: 34855212]
26. Çınar D, Karadakovan A, Erdoğan AP. Effect of mobile phone app-based training on the quality of life for women with breast cancer. Eur J Oncol Nurs 2021 Jun;52:101960. [doi: 10.1016/j.ejon.2021.101960] [Medline: 33882446]
27. Fjell M, Langius-Eklöf A, Nilsson M, Wengström Y, Sundberg K. Reduced symptom burden with the support of an interactive app during neoadjuvant chemotherapy for breast cancer - a randomized controlled trial. Breast 2020 Jun;51:85-93 [FREE Full text] [doi: 10.1016/j.breast.2020.03.004] [Medline: 32247251]
28. Greer J, Jacobs J, Pensak N, MacDonald J, Fuh C, Perez G, et al. Randomized trial of a tailored cognitive-behavioral therapy mobile application for anxiety in patients with incurable cancer. Oncologist 2019 Aug;24(8):1111-1120 [FREE Full text] [doi: 10.1634/theoncologist.2018-0536] [Medline: 30683710]
29. Greer J, Jacobs JM, Pensak N, Nisotel LE, Fishbein J, MacDonald J, et al. Randomized trial of a smartphone mobile app to improve symptoms and adherence to oral therapy for cancer. J Natl Compr Canc Netw 2020 Feb;18(2):133-141. [doi: 10.6004/jnccn.2019.7354] [Medline: 32023526]
30. Ham K, Chin S, Suh YJ, Rhee M, Yu E, Lee HJ, et al. Preliminary results from a randomized controlled study for an app-based cognitive behavioral therapy program for depression and anxiety in cancer patients. Front Psychol 2019 Jul 25;10:1592 [FREE Full text] [doi: 10.3389/fpsyg.2019.01592] [Medline: 31402881]
31. Karaaslan-Eşer A, Ayaz-Alkaya S. The effect of a mobile application on treatment adherence and symptom management in patients using oral anticancer agents: a randomized controlled trial. Eur J Oncol Nurs 2021 Jun;52:101969. [doi: 10.1016/j.ejon.2021.101969] [Medline: 33991868]
32. Park HR, Kang HS, Kim SH, Singh-Carlson S. Effect of a smart pill bottle reminder intervention on medication adherence, self-efficacy, and depression in breast cancer survivors. Cancer Nurs 2021 Oct 12;45(6):E874-E882. [doi: 10.1097/ncc.0000000000001030]
33. Peng Z, Li L, Chen Y, Feng Z, Fang X. WeChat app-based reinforced education improves the quality of opioid titration treatment of cancer-related pain in outpatients: a randomized control study. BMC Cancer 2020 Sep 04;20(1):852 [FREE Full text] [doi: 10.1186/s12885-020-07270-w] [Medline: 32887560]
34. Spahrkäs SS, Looijmans A, Sanderman R, Hagedoorn M. Beating cancer-related fatigue with the Untire mobile app: results from a waiting-list randomized controlled trial. Psychooncology 2020 Nov 11;29(11):1823-1834 [FREE Full text] [doi: 10.1002/pon.5492] [Medline: 33393199]
35. Sui Y, Wang T, Wang X. The impact of WeChat app-based education and rehabilitation program on anxiety, depression, quality of life, loss of follow-up and survival in non-small cell lung cancer patients who underwent surgical resection. Eur J Oncol Nurs 2020 Apr;45:101707. [doi: 10.1016/j.ejon.2019.101707] [Medline: 32078926]
36. Zhou K, Li J, Li X. Effects of cyclic adjustment training delivered via a mobile device on psychological resilience, depression, and anxiety in Chinese post-surgical breast cancer patients. Breast Cancer Res Treat 2019 Nov 24;178(1):95-103. [doi: 10.1007/s10549-019-05368-9] [Medline: 31342310]
37. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]

38. Di R, Li G. Use of a smartphone medical app improves complications and quality of life in patients with nasopharyngeal carcinoma who underwent radiotherapy and chemotherapy. Med Sci Monit 2018 Sep 04;24:6151-6156. [doi: 10.12659/msm.908146]
39. Dong X, Yi X, Gao D, Gao Z, Huang S, Chao M, et al. The effects of the combined exercise intervention based on internet and social media software (CEIBISMS) on quality of life, muscle strength and cardiorespiratory capacity in Chinese postoperative breast cancer patients: a randomized controlled trial. Health Qual Life Outcomes 2019 Jun 26;17(1):109 [FREE Full text] [doi: 10.1186/s12955-019-1183-0] [Medline: 31242926]
40. Hou I, Lin H, Shen S, Chang K, Tai H, Tsai A, et al. Quality of life of women after a first diagnosis of breast cancer using a self-management support mHealth app in Taiwan: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 04;8(3):e17084 [FREE Full text] [doi: 10.2196/17084] [Medline: 32130181]
41. Lei L. The Application of APP in the Transitional Care Among the post-operative patients with laryngeal cancer. In: Dissertation of nursing school in Zhengzhou University. Zhengzhou: Zhengzhou University; 2016.
42. Rosen KD, Paniagua SM, Kazanis W, Jones S, Potter JS. Quality of life among women diagnosed with breast Cancer: a randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology 2018 Aug 01;27(8):2023-2030. [doi: 10.1002/pon.4764] [Medline: 29766596]
43. Zha G. Effect of WeChat continuity of care on psychological stress, self-care ability and quality of life of chemotherapy patients after radical breast cancer surgery. J Practical Cardiopulmonary Vascular Disease 2020 Jul 30:148-150 [FREE Full text]
44. Absolom K, Warrington L, Hudson E, Hewison J, Morris C, Holch P, et al. Phase III randomized controlled trial of eRAPID: eHealth intervention during chemotherapy. J Clin Oncol 2021 Mar 01;39(7):734-747. [doi: 10.1200/jco.20.02015]
45. Berg CJ, Vanderpool RC, Getachew B, Payne JB, Johnson MF, Sandridge Y, et al. A hope-based intervention to address disrupted goal pursuits and quality of life among young adult cancer survivors. J Cancer Educ 2020 Dec 11;35(6):1158-1169 [FREE Full text] [doi: 10.1007/s13187-019-01574-7] [Medline: 31297743]
46. Chen K, Yao F, Chen X, Lin Y, Kang M. Effectiveness of telerehabilitation on short-term quality of life of patients after esophageal cancer surgery during COVID-19: a single-center, randomized, controlled study. J Gastrointest Oncol 2021 Aug;12(4):1255-1264 [FREE Full text] [doi: 10.21037/jgo-21-385] [Medline: 34532085]
47. Huggins CE, Hanna L, Furness K, Silvers MA, Savva J, Frawley H, et al. Effect of early and intensive telephone or electronic nutrition counselling delivered to people with upper gastrointestinal cancer on quality of life: a three-arm randomised controlled trial. Nutrients 2022 Aug 07;14(15):3234 [FREE Full text] [doi: 10.3390/nu14153234] [Medline: 35956410]
48. 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]
49. Seib C, Anderson D, McGuire A, Porter-Steele J, McDonald N, Balaam S, et al. Improving health-related quality of life in women with breast, blood, and gynaecological cancer with an eHealth-enabled 12-week lifestyle intervention: the women's wellness after Cancer program randomised controlled trial. BMC Cancer 2022 Jul 08;22(1):747 [FREE Full text] [doi: 10.1186/s12885-022-09797-6] [Medline: 35804322]
50. How many people have smartphones in 2022? OBERLO. URL: https://www.oberlo.com/statistics/how-many-people-have-smartphones [accessed 2022-10-24]
51. Ryhänen AM, Rankinen S, Siekkinen M, Saarinen M, Korvenranta H, Leino-Kilpi H. The impact of an empowering internet-based breast cancer patient pathway program on breast cancer patients' clinical outcomes: a randomised controlled trial. J Clin Nurs 2013 Apr 11;22(7-8):1016-1025. [doi: 10.1111/jocn.12007] [Medline: 23480498]
52. Diwakar V, Ertmer PA, Nour AY. Developing interactive course Web sites for distance education and characteristics of students enrolled in distance learning courses. J Vet Med Educ 2003 Dec;30(4):351-357. [doi: 10.3138/jvme.30.4.351] [Medline: 14976621]
53. McLean S, Protti D, Sheikh A. Telehealthcare for long term conditions. BMJ 2011 Feb 03;342(feb03 2):d120. [doi: 10.1136/bmj.d120] [Medline: 21292710]
54. Park J, Look KA. Relationship between objective financial burden and the health-related quality of life and mental health of patients with cancer. J Oncol Pract 2018 Feb;14(2):e113-e121. [doi: 10.1200/jop.2017.027136]
55. Bradford N, Caffery L, Smith A. Telehealth services in rural and remote Australia: a systematic review of models of care and factors influencing success and sustainability. Rural Remote Health 2016 Oct 17:3808. [doi: 10.22605/rrh3808]
56. Ye M, Du K, Zhou J, Zhou Q, Shou M, Hu B, et al. A meta-analysis of the efficacy of cognitive behavior therapy on quality of life and psychological health of breast cancer survivors and patients. Psychooncology 2018 Jul 26;27(7):1695-1703. [doi: 10.1002/pon.4687] [Medline: 29500842]
57. Getu MA, Chen C, Panpan W, Mboineki JF, Dhakal K, Du R. The effect of cognitive behavioral therapy on the quality of life of breast cancer patients: a systematic review and meta-analysis of randomized controlled trials. Qual Life Res 2021 Feb 17;30(2):367-384. [doi: 10.1007/s11136-020-02665-5] [Medline: 33068239]
58. He Q, Zhao X, Wang Y, Xie Q, Cheng L. Effectiveness of smartphone application-based self-management interventions in patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. J Adv Nurs 2022 Feb 29;78(2):348-362. [doi: 10.1111/jan.14993] [Medline: 34324218]

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
QOL: quality of life
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by R Kukafka; submitted 24.05.22; peer-reviewed by H Mehdizadeh, P Anderson; comments to author 30.09.22; revised version received 08.11.22; accepted 13.11.22; published 19.12.22.

Please cite as:
Qin M, Chen B, Sun S, Liu X
Effect of Mobile Phone App–Based Interventions on Quality of Life and Psychological Symptoms Among Adult Cancer Survivors: Systematic Review and Meta-analysis of Randomized Controlled Trials
J Med Internet Res 2022;24(12):e39799
URL: https://www.jmir.org/2022/12/e39799
doi:10.2196/39799
PMID:36534460

Review

Digital Biomarker–Based Interventions: Systematic Review of Systematic Reviews

Hossein Motahari-Nezhad[1,2], MSc; Hana Al-Abdulkarim[3,4], MSc; Meriem Fgaier[3], MSc; Mohamed Mahdi Abid[5], MSc; Márta Péntek[1], MD, DSc; László Gulácsi[1,6], MD, DSc; Zsombor Zrubka[1,6], MBA, MD, PhD

[1]Health Economics Research Center, University Research and Innovation Center, Óbuda University, Budapest, Hungary

[2]Doctoral School of Business and Management, Corvinus University of Budapest, Budapest, Hungary

[3]Doctoral School of Applied Informatics and Applied Mathematics, Óbuda University, Budapest, Hungary

[4]Drug Policy and Economic Center, National Guard Health Affairs, King Abdullah International Medical Research Center, Riyadh, Saudi Arabia

[5]Research Center of Epidemiology and Statistics, Université Sorbonne Paris Cité, Paris, France

[6]Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary

Corresponding Author:
Hossein Motahari-Nezhad, MSc
Health Economics Research Center, University Research and Innovation Center, Óbuda University
Budapest, Bécsi út 96/B
Budapest, 1034
Hungary
Phone: 36 702097967
Email: h.motahari.lib@gmail.com

Abstract

Background: The introduction of new medical technologies such as sensors has accelerated the process of collecting patient data for relevant clinical decisions, which has led to the introduction of a new technology known as digital biomarkers.

Objective: This study aims to assess the methodological quality and quality of evidence from meta-analyses of digital biomarker–based interventions.

Methods: This study follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline for reporting systematic reviews, including original English publications of systematic reviews reporting meta-analyses of clinical outcomes (efficacy and safety endpoints) of digital biomarker–based interventions compared with alternative interventions without digital biomarkers. Imaging or other technologies that do not measure objective physiological or behavioral data were excluded from this study. A literature search of PubMed and the Cochrane Library was conducted, limited to 2019-2020. The quality of the methodology and evidence synthesis of the meta-analyses were assessed using AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews 2) and GRADE (Grading of Recommendations, Assessment, Development, and Evaluations), respectively. This study was funded by the National Research, Development and Innovation Fund of Hungary.

Results: A total of 25 studies with 91 reported outcomes were included in the final analysis; 1 (4%), 1 (4%), and 23 (92%) studies had high, low, and critically low methodologic quality, respectively. As many as 6 clinical outcomes (7%) had high-quality evidence and 80 outcomes (88%) had moderate-quality evidence; 5 outcomes (5%) were rated with a low level of certainty, mainly due to risk of bias (85/91, 93%), inconsistency (27/91, 30%), and imprecision (27/91, 30%). There is high-quality evidence of improvements in mortality, transplant risk, cardiac arrhythmia detection, and stroke incidence with cardiac devices, albeit with low reporting quality. High-quality reviews of pedometers reported moderate-quality evidence, including effects on physical activity and BMI. No reports with high-quality evidence and high methodological quality were found.

Conclusions: Researchers in this field should consider the AMSTAR-2 criteria and GRADE to produce high-quality studies in the future. In addition, patients, clinicians, and policymakers are advised to consider the results of this study before making clinical decisions regarding digital biomarkers to be informed of the degree of certainty of the various interventions investigated in this study. The results of this study should be considered with its limitations, such as the narrow time frame.

International Registered Report Identifier (IRRID): RR2-10.2196/28204

(J Med Internet Res 2022;24(12):e41042) doi:10.2196/41042

KEYWORDS

digital biomarker; digital health; digital devices; AMSTAR-2; GRADE; methodological quality; evidence synthesis; publication bias; imprecision; implantable; wearable

Introduction

The introduction of new medical technologies such as sensors has accelerated the process of collecting patient data for relevant clinical decisions [1], which has led to the introduction of a new technology known as digital biomarkers (DBMs). "Digital biomarkers are objective, measurable, physiological, and behavioral parameters collected using wearable, portable, implantable, or digestible digital devices" [2]. DBMs can play an important role in daily clinical practice and clinical trials [3]. By providing timely and reliable disease-related information, DBMs can increase diagnostic accuracy, improve treatment decisions and help minimize clinical errors, and contribute to better patient outcomes [4-6]. Digital biomarkers can provide more reliable results than cross-sectional surveillance or prospective follow-up, allowing fewer patient visits [7]. Because of their growing importance in the health care value chain, the market of DBMs is expected to grow at a compound annual growth rate of 40.4% between 2019 and 2025, reaching a global revenue of US $5.64 billion by 2025 [8,9].

The rapid development of digital health technologies such as software [10], sensors [11], or robots [12,13] requires thorough examination and demonstration of their clinical effectiveness and economic benefits before they are widely deployed in publicly funded health systems. Assessing the value of digital health technologies is complex, with considerations beyond normal health economic analyses [14-18]. The evidence required for the value assessment of digital health technologies usually reflects their risk category ranging from basic consumer health monitoring to interventions impacting therapy or diagnosis. For high-risk technologies, it is essential to demonstrate the clinical benefit of randomized clinical trials conducted in a relevant health system or meta-analyses of randomized controlled trials [17,18].

In recent years, the clinical outcomes of DBMs have been extensively synthesized in systematic reviews and meta-analyses with inconsistent results, calling for a more systematic approach to evaluating the evidence concerning DBM interventions [19]. When interpreting systematic reviews, it is essential to appraise the quality of evidence and estimates of the effect size. Among the several methods for assessing the quality of evidence [20], the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) approach is used most commonly in systematic reviews, health technology assessments, and treatment guidelines [19]. GRADE classifies the quality of evidence into 4 categories from high to very poor [19]. However, poor reporting may limit the assessment of the quality of the evidence presented in systematic reviews. The AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews 2) tool was developed to assess the methodological quality of systematic reviews [21].

Our goal, therefore, is to provide innovators and policymakers with actionable guidance on the level of evidence generation for DBMs, a rapidly growing area of medicine [2]. This systematic review of systematic reviews assesses the overall strength of evidence and methodological quality of systematic reviews that present a quantitative synthesis of the effects of digital biomarkers on health outcomes compared with interventions that do not include digital biomarkers. The AMSTAR-2 technique examines the methodological quality of studies, while GRADE assesses the overall quality of evidence based on digital biomarker technologies and reported outcomes.

Methods

Design and Protocol

This study follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for reporting systematic reviews (Multimedia Appendix 1) [22]. The protocol of the current systematic review was published in *JMIR Research Protocols* [23].

Eligibility for Inclusion

DBMs are "objective, measurable, physiological, and behavioral parameters collected using wearable, portable, implantable, or digestible digital devices" [2]. In this research, we defined DBMs as either behavioral/physiological data or the digital devices used to collect these data. Wearable, implantable, or digestible medical devices or sensors that generate physiologic or behavioral data were considered digital biomarkers (eg, fitness trackers and defibrillators). Imaging or other technologies that do not measure physiological or behavioral data were excluded from this study. We interpret portable as "portable by patients or consumers"; therefore, portable devices operated by health care professionals (eg, digital stethoscopes) were excluded. We note that the definition of DBMs may overlap with sensor applications in the general population, such as citizen sensing [24]. In this search, we only considered systematic reviews that use digital devices deployed by clinicians or patients to collect clinical data in the context of treatment.

We included systematic reviews reporting meta-analyses of clinical outcomes of DBM-based interventions compared with alternative interventions without DBMs. In particular, we considered systematic reviews summarizing DBM-related evidence in a human population for any condition, age group, or sex. All interventions that use DBMs for any purpose related to diagnosing patients, monitoring outcomes, or influencing a therapeutic intervention were considered. There were no restrictions on comparators as long as the comparator arm did not involve using DBMs for the aforementioned purposes. Only meta-analyses of clinical outcomes were considered (ie, intended or unintended change in participants' health status due to an intervention). Systematic reviews focused on the measurement properties, or other technical or utilization characteristics of DBMs that do not result in a change in participants' health status were not eligible for this review. We considered full-text articles

published in English in peer-reviewed journals between January 1, 2019, and December 31, 2020.

Search Strategies

A literature search was conducted in PubMed and the Cochrane Library, with a time frame limited to 2019 and 2020. In addition, we checked the reference lists of systematic reviews potentially relevant to our research. The literature search used keywords related to "digital biomarkers" [2] in conjunction with The National Library of Medicine's filter for "systematic reviews" [25] and the publication date. Multimedia Appendix 2 contains the complete search syntax.

Screening and Selection

After removing duplicates, 2 reviewers (HM-N and MMA) independently screened titles and abstracts using 2 main criteria: (1) systematic reviews and (2) interventions that included DBMs. Reviewer calibration was performed after screening the titles/abstracts of the first 100 records using the following method. Both screening criteria were scored as either 1 (criterion not met) or 0 (criterion met or uncertain). Therefore, reviewers can evaluate each record with a score of 1, 2, 3, and 4, corresponding to the response patterns (0,0), (1,0), (0,1), and (1,1), respectively. Interrater agreement and κ statistics were calculated for scoring, and reviewers were retrained if worse than substantial agreement (κ=0.6) was observed [26]. A third reviewer (ZZ) made the decision in the case of nonmatching scores.

Full-text articles were assessed by 2 independent reviewers against all eligibility criteria: (1) English language; (2) human research; (3) publication date; (4) meta-analysis of clinical outcomes; (5) the intervention involved a DBM used for diagnosis, patient monitoring, or influencing therapy; (6) the comparator arm lacked a DBM for the same purposes. All 6 criteria had to be answered "yes" for inclusion. Discrepancies were resolved by the 2 reviewers. In case of disagreement, a third reviewer took a decision.

Data Extraction and Quality Assessment

Data extraction and the assessments of methodological quality and the quality of evidence were performed by 2 independent researchers (HM-N, HA-A, or MF). Interrater agreement was assessed after completing data extraction from 20% of the included studies. Disagreements between reviewers were resolved by consensus, and a third reviewer (ZZ) resolved the remaining differences.

Study-Level Variables

The following study-level variables were recorded: Year of publication; country of the first author; number of included studies in the qualitative/quantitative synthesis overall and separately for each outcome; study designs of the included studies (randomized controlled trial/nonrandomized controlled trial/cohort study/case-control study/cross-sectional study) [27]; population and its age range; the disease condition evaluated using the International Classification of Diseases 11th Revision (ICD-11) coding [28]; the number of included studies; intervention; type of intervention using the International Classification of Health Interventions (ICHI) coding [29]; comparator; type of comparator; the DBM; role of the DBM (diagnosis/patient monitoring/influencing intervention); body function quantified by the digital biomarker using the International Classification of Functioning, Disability and Health (ICF) coding [30]; and the list of synthesized outcomes.

Outcome-Level Variables

We extracted the outcome measured, the total number of studies that examined that outcome, the total number of patients and the number receiving the intervention, the effect size and its 95% CI (upper and lower limits), and the type of effect size (eg, standardized mean difference/odds ratio/risk ratio).

Assessment of the Methodological Quality of the Systematic Reviews

The methodological quality of the included systematic reviews was assessed using the AMSTAR-2 tool [21]. AMSTAR-2 is a recognized and reliable 16-item tool for evaluating the methodological quality of systematic reviews of health care treatments [21,31]. We performed a consistent assessment [32] using the AMSTAR-2 website and categorized the reporting quality of reviews accordingly as critically low, low, medium, and high [21].

Assessing the Quality of the Evidence

We assessed the quality of evidence for each outcome using the GRADE system [19,33]. By default, GRADE classifies evidence from randomized controlled trials as high quality. However, this rating can be downgraded based on the assessment of the following 5 quality domains: (1) risk of bias [34], (2) inconsistency [35], (3) imprecision [36], (4) publication bias [37], and (5) indirectness [38]. Depending on the severity of the quality concerns, a downgrade of 0, 1, or 2 can be proposed for each domain.

We assessed the risk of bias according to the following criteria: if 75% or more than 75% of the included studies had a low risk of bias for a given outcome, no downgrade was applied. If less than 75% of the included studies had a low risk of bias or risk of bias was not reported, 1 downgrade was used [39].

Inconsistency was assessed by the reported heterogeneity for each outcome. If the I^2 statistic was less than or equal to 75%, no downgrading was performed. If the I^2 statistic was greater than 75%, 1 downgrade was assigned. If only a single study was included for the outcome, no downgrade was applied. If heterogeneity was not reported, a downgrade was applied [39].

Imprecision was assessed by evaluating the sample size [40]. The evidence was not downgraded if the pooled sample size exceeded 2000 [33]. We applied 1 downgrade if the pooled sample size was less than 200. Between a pooled sample size of 200 and 2000, we evaluated the optimal information size by power analysis using Stata version 16 (StataCorp LLC) as follows [33]: assuming a weak effect size [41], we calculated the sample size for a randomized controlled trial assuming a balanced sample, a power of 0.8, and a significance level of .05. One downgrade was applied when the calculated sample size was larger than the pooled sample size [33,40]. The following procedure was used for the small effect size: a Cohen *d* of 0.2

for continuous measures and 1.68 for the odds ratio. A weak effect size of 1.68 was also estimated for the risk ratio and hazard ratio, assuming a nonexposed prevalence of 0% [41,42].

The potential effect of publication bias on the effect size estimates was assessed for each outcome using the trim-and-fill method proposed by Duval and Tweedie [43]. Potentially missing studies were imputed, and the pooled effect size of the full data set was recalculated. If the imputation changed the conclusions of the analysis (eg, a significant effect size became no longer significant or the magnitude of effect size changed), we applied a downgrade due to publication bias [43]. According to the recommendations of the Cochrane Handbook [42], we assessed publication bias only in meta-analyses involving at least ten studies due to the limited power of risk of bias tests when applied on fewer studies.

When assessing indirectness for each outcome, we considered discrepancies between the included studies and the research question of the meta-analysis [44]. If the population, interventions, or comparators of the studies did not match the main objectives of the meta-analysis, a downgrade of 1 or 2 was considered, depending on the severity of this nonmatch, based on the consensus of the 2 independent investigators involved in data extraction.

The overall grading of the quality of evidence for each outcome was based on consensus, following the recommendation of Pollock et al [39]. The evidence was considered as high quality if further research was very unlikely to change our confidence in the estimate of effect (0 downgrades); moderate quality if further research was likely to have an important effect on our confidence in the estimate of effect and might change the estimate (1-2 downgrades); low quality if further research was very likely to have an important effect on our confidence in the estimate of effect and might change the estimate (3-4 downgrades); and very low quality if any estimate of the effect was very uncertain (5-6 downgrades) [19,39].

Evidence Synthesis

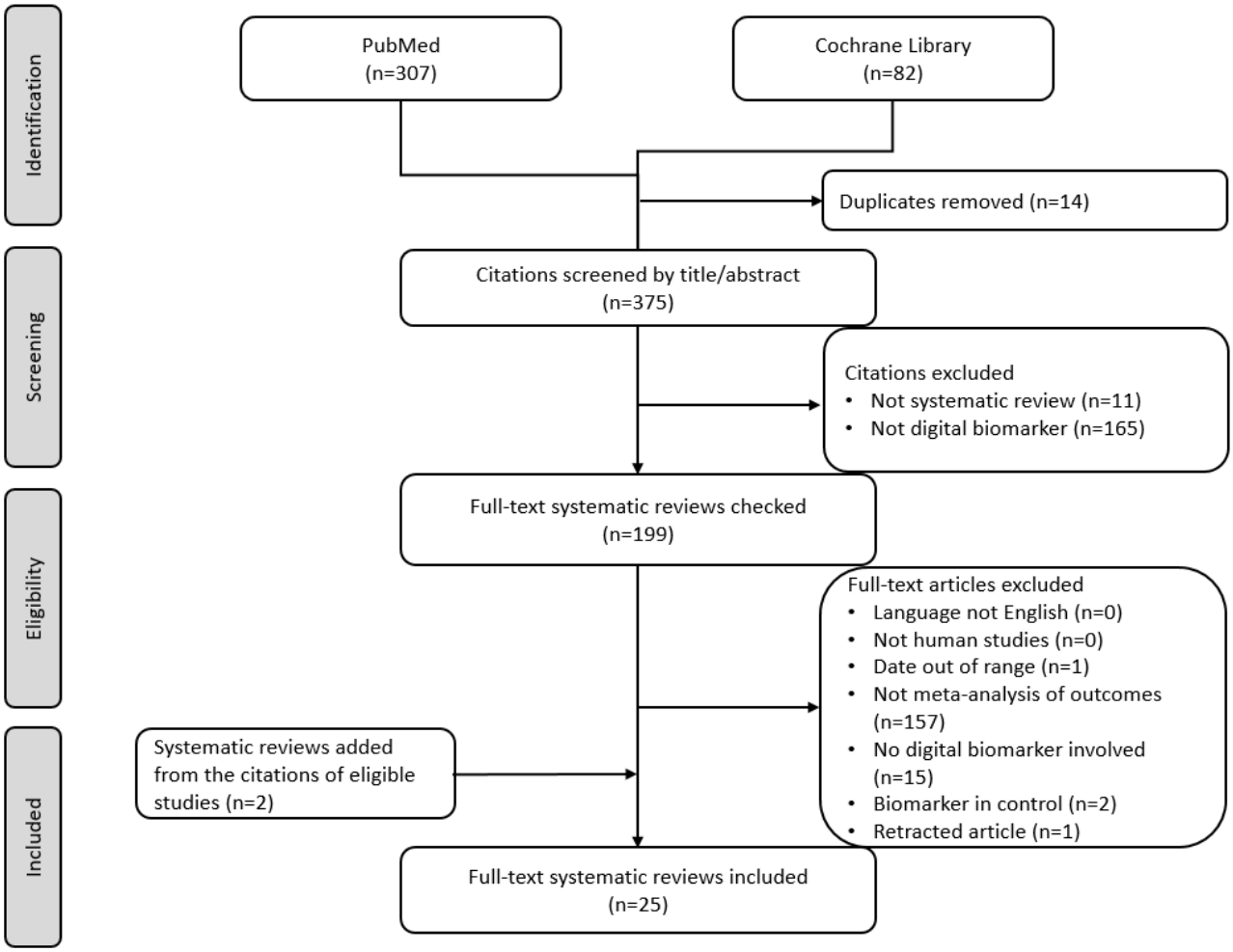
Descriptive statistics including frequency and percentage were used to describe the characteristics of the studies using Stata version 16 and MS Excel 2016. The graphs were designed using R programming language 4.1.3 (R Core Team/R Foundation for Statistical Computing). In the designed graphs (Figures 2 and 3), the letters on the horizontal axis correspond to the interventions because the types of interventions were heterogeneous; for example, in 1 study, the intervention was a single digital device (such as an implantable cardiac defibrillator [ICD]), whereas in another, it was a combination of devices (such as Fitbit, Jawbone UP24, combined heart rate monitor, and accelerometer [Actiheart], wrist-worn accelerometer, FIT Core, Body Media, Fitbug Orb, Polar FA20 accelerometer). Given the diversity of populations and treatments studied, we tabulated the GRADE evidence summary for each DBM by type of intervention and outcome.

Results

Screening and Selection of Studies

Searches of the PubMed and Cochrane Library electronic databases yielded 307 and 82 documents, respectively, bringing the total number of studies found to 389. After removing duplicates (n=14), 375 studies were considered eligible for title/abstract screening. In the screening phase, we removed 176 studies, of which 11 were not systematic reviews and 165 did not involve DBMs (87 disagreements between reviewers during title/abstract screening; Cohen κ=0.54). During the screening phases of the titles/abstracts, “digital biomarker” was associated with 82 disagreements and “systematic review” with 5. Therefore, 199 studies were included in the full-text screening. In accordance with the eligibility criteria, 176 full-text papers were excluded (between-reviewers κ=0.76) for the following reasons: publication date outside the acceptable range (n=1), no meta-analysis of results (n=157), studies without DBMs (n=15), retraction (n=1) [45], and DBMs in the control group (n=2). The list of excluded studies with reasons are presented in Multimedia Appendix 3. In addition, when reviewing the reference lists of the final eligible studies, 2 more reviews met the inclusion criteria. Therefore, 25 systematic reviews were included in the final analysis (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram selecting/screening process.



Characteristics of the Included Systematic Reviews

Most studies were published by authors from Australia (5/25, 20%) [46-50] followed by those from the United States (3/25, 12%) [51-53], Taiwan (3/25, 12%) [54-56], Canada (2/25, 8%) [57,58], Hong Kong (2/25, 8%) [59,60], and the United Kingdom of Great Britain and Northern Ireland (1/25, 4%) [61]. The other 9 reviews (36%) were published by researchers from Belgium [62], China [63], France [64], Greece [65], Japan [66], the Netherlands [67], Portugal [68], Saudi Arabia [69], and Thailand [70].

Populations

Using ICD-11, most participants in the included systemic reviews were assigned to circulatory system diseases [47,48,51,56,57,60,62,65,67-70], followed by patients with endocrine, nutritional, or metabolic disorders [58,60,62,67] and respiratory system diseases [57,60-62,67]. Patients with nutritional disorders [47,60,62], diseases of the nervous system [47,60,67], and problems associated with health behaviors [47,59,67] were included in 3 reviews each. The other populations were classified in the presence of device, implants, or grafts [55,67]; diseases of the musculoskeletal system or connective tissue [47,64]; causes of health care–related harm or injury [52,53]; diseases of the urinary system [63]; injury or harm arising from surgical or medical care [55]; neoplasms [47]; and injury, poisoning, or certain other consequences of external causes [66]. In addition, 4 reviews examined nonclinical populations [46,49,50,54]. In some reviews, nonclinical cases such as healthy individuals [57,60], employees [57], and students [57] were included in addition to patients with clinical diseases that could not be categorized using the ICD-11 tool.

Interventions

In accordance with the ICHI instrument, 14 diverse intervention categories were discovered, and the majority of digital biomarkers were used as interventions on physical activity behaviors (eg, Fitbit) [46-50,57-62,64,67], conversion of cardiac rhythm (eg, cardiac defibrillators) [51-53,63,68,69], cardiac electrophysiological monitoring (eg, iPhone-based rhythm monitoring device) [55,65,70], weight maintenance functions (eg, Garmin or Jawbone UP24) [49,54,57], and whole-body measurement (eg, wristbands and smartwatches) [50,54]. Other interventions identified were associated with cardiopulmonary resuscitation (eg, metronome with a siren) [56], assisting and leading exercise for exercise tolerance function (eg, GEx sensor of vital signs and smartphone) [48], body measurement of trunk (eg, wristbands, smartwatches) [54], pain (eg, accelerometer, pedometers) [64], test of functions (eg, YAMAX, Fitbit) [64], quality of life (eg, pedometers) [64], test of muscle endurance (eg, fitness trackers) [64], body measurement of lower limb (eg, accelerometer-based navigation system) [66], and test of maintaining body position (eg, accelerometer-based navigation system) [66].

Outcomes

According to the ICF system, the vast majority of reported outcomes concerned physical activity (looking after one's health; eg, moderate-to-vigorous physical activity, step counts) [46-50,58-62,64,67], followed by mortality (demographic change; eg, all-cause mortality, sudden cardiac death) [51-53,63,68-70], and heart functions (eg, return of spontaneous circulation, incidence of ventricular arrhythmia) [52,55,65,70]. A total of 11 studies also reported weight maintenance functions (eg, weight, BMI, and waist circumference) [49,54,57], health services, systems and policies (eg, quality of life and prevention) [55,63,64], maintaining one's health (such as hospitalization and readmission rate) [51,52,69], and managing one's own activity level (actions and behaviors to arrange the requirements in energy and time day-to-day procedures or duties; eg, sedentary behaviors) [46,57]. Because of the difference between sedentary behavior and physical activity, these 2 outcomes were considered different endpoints, as physical activity and sedentary behavior are measured differently and do not affect risks in the same way [71]. The other remaining reported outcomes were aerobic capacity [48], pain [64], fatigability [64], social security services, systems and policies (eg, disability) [64], body functions (eg, functional tests) [64], and mobility of joint functions (such as coronal femoral component alignment or coronal tibial component alignment) [66].

Bodily Functions Quantified by Digital Biomarkers

The most commonly used physiological/behavioral data captured by digital biomarkers to modify participants' health status were heart functions/rhythm [51-53,55,56,63,65,68,70] and physical activity (looking after one's health) [46,47,50,57-60,64,67], followed by walking [46-49,59-62,64,67], weight maintenance functions [49,54,57], gait pattern functions [57], running [59], aerobic capacity [48], and involuntary movement reaction functions [66]. For further information regarding population, intervention, outcome, and digital biomarkers, see Multimedia Appendix 4.

The Methodological Quality of Systematic Reviews

Most studies (23/25, 92%) [46-49,51-55,57-70] had critically low methodological quality according to the assessment using AMSTAR-2. The remaining studies also received high (1/25, 4%) [50] and low methodological quality (1/25, 4%) ratings [56]. The only study of high methodological quality was assigned to a review that investigated the effect of workplace pedometer interventions to increase physical activity [50]. Although all studies were able to meet criteria 3 (inclusion criteria), 9 (risk of bias assessment), and 11 (appropriate statistical methods) of AMSTAR-2, criteria 4 (comprehensive literature search), 7 (list of excluded studies), 10 (funding report), and 13 (account for risk of bias when reporting results) were met by only 2 [50,64], 2 [50,65], 2 [50,69], and 7 studies [46,50,54,56,58,60,66], respectively. Detailed information on the methodological quality of the studies for each criterion can be found in Multimedia Appendix 5.

Quality of Evidence Synthesis Results

The 25 reviews included in the study comprised a total of 91 outcomes. Of the 91 outcomes, only 6 (7%) were rated as high-quality evidence, whereas 80 (88%) were rated as moderate-quality and 5 (5%) as low-quality evidence. The results showed that the effect of an ICD on all-cause mortality received high-quality evidence for ICDs implanted after and with continuous flow left ventricular assist devices. Furthermore, based on the analyses, we are highly confident about the impact of the ICD on the probability of transplantation, the detection rate of atrial arrhythmias, and the incidence of stroke. By contrast, some outcomes were found to have low-quality evidence, including the effect of wearable activity trackers on steps in chronic respiratory disease as well as on steps in overweight and sedentary older adults. A total of 2 meta-analyses that examined the effect of wearable activity trackers on moderate-to-vigorous physical activity were also rated as low-quality evidence. Concerning the criteria of GRADE, risk of bias was found in most outcomes (85/91, 93%), followed by inconsistency (27/91, 30%) and imprecision (27/91, 30%). Publication bias was detected in a small number of outcomes (2/91, 2%). By contrast, no indirectness was revealed in the outcomes. In addition, 67 outcomes (74%) were not examined for publication bias because the minimum number of included studies was insufficient; 3 outcomes (3%) were also not assessed for inconsistency because only 1 study was included. See Multimedia Appendix 6 for more details.

Discussion

Principal Findings

To our knowledge, this study is the first to analyze the methodological and evidence-based quality of systematic reviews providing meta-analyses of digital biomarker–based interventions' effect on human populations' health-related outcomes. A total of 25 systematic reviews evaluating the clinical impact of digital biomarkers on human health were included in our study, comprising a total of 91 outcomes. There were no reviews of high methodological quality on digital biomarker–based interventions with high quality of evidence. Most outcomes had moderate-quality evidence synthesis. All implantable cardiac devices and monitors had significant results with moderate-quality evidence and critically low methodological quality. Most activity trackers also had significant effects on steps and weight with moderate certainty of evidence and critically low methodological quality. By contrast, the evidence synthesis and methodological quality of activity trackers were rated moderate and critically low, respectively, for quality of life, pain, fatigue, and disability. Still, the results of the meta-analyses showed a nonsignificant effect of activity trackers on the aforementioned endpoints.

The Methodological Quality of Systematic Reviews

The results of the methodological quality of the studies using the AMSTAR-2 tool showed that most studies had critically low methodological quality, mainly due to factor numbers 7 (excluded studies) and 10 (source of funding) of the AMSTAR-2 tool, leaving concerns about the unbiasedness of results and indicating the need for quality improvement. Researchers in this field need to follow the AMSTAR-2 guidelines and criteria to produce high-quality systematic reviews. The list of excluded studies and the rationale for deleting each study are critical parts

of the AMSTAR-2 tool for assessment [21]. This limitation is included in the majority of some previously published systematic reviews in digital interventions for reducing behavioral risks [72], synchronous digital mental health systematic reviews [73], and interventions involving antibacterial envelopes to reduce cardiac implantable electronic device–related infections [74].

As listing excluded studies and the rationale for their deletion are critical components of the methodology of systematic reviews according to the AMSTAR-2 criteria [21], researchers are advised to provide excluded studies with rationale for their exclusion when conducting systematic reviews. In addition, the source of funding for the research included in the systematic reviews should be indicated. Most systematic reviews included in this study could not meet this criterion. The results of this study are consistent with those of many previous studies [72,73,75]. Prior studies on digital interventions for reducing behavioral risks [72] and systematic review of synchronous digital mental health reviews [73] also rated the methodological quality of most systematic reviews as critically low. By contrast, the methodological quality of most systematic reviews on digital health interventions on palliative care [75] and the use of eHealth with immunizations [76] was rated low and moderate, respectively.

Quality of Evidence

Of the 91 outcomes assessed, only 6 had high-quality evidence, meaning that we can be highly confident that the actual effect is close to the estimated effect and that further studies are unlikely to change our confidence in the estimate of the effect [77]. Considering that a substantial proportion of digital biomarker–based outcomes had evidence of moderate quality, we have moderate confidence in the effect estimate. Although the actual effect is likely to be similar to the estimated effect, there is a possibility that it will be significantly different, and additional research is expected to have a significant impact on our confidence in the effect estimate and alter the estimate [77]. In addition, some outcomes were of low quality, suggesting that our confidence in the impact estimate is limited and that the actual effect may differ substantially from the impact estimate [77].

Most outcomes were downgraded mainly because of the risk of bias in the included studies. In addition, the analysis revealed that most of the included systematic reviews did not assess and discuss the impact of risk of bias on the measured outcomes. Therefore, clinical researchers in this field are advised first to determine the impact of risk of bias on their effect estimates and then discuss the likely impact of risk of bias on outcomes to produce high-quality results. High heterogeneity was another detrimental factor observed in nearly one-third of the outcomes. However, most of the included systematic reviews were able to meet AMSTAR criterion 14, investigated the sources of any heterogeneity in the results, and discussed this criterion's impact on the review results. Researchers can study heterogeneity in several ways, such as by performing subgroup analyses or meta-regressions, using a fixed-effects or random-effects model [42], changing the statistical measure from risk difference to relative risk, and deleting studies [78]. Another critical factor in the deterioration of the quality of some outcomes was imprecision. Clinical researchers should consider the optimal information size for their measured outcomes using power calculations to obtain a high-quality effect estimate without imprecision.

Some previous studies also assessed the quality of evidence in some research areas. A study evaluating the quality of evidence of systematic reviews of acupuncture for stroke rehabilitation concluded that the quality of evidence for almost all outcomes was low, mainly because of inconsistency, imprecision, and risk of bias, respectively [79]. Another study that assessed the quality of meta-analyses of Chinese herbal preparations for the treatment of rheumatoid arthritis concluded that most outcomes (55%) were of low quality. In comparison, 25% and 20% were of moderate and very low quality, respectively, primarily because of the risk of bias and inconsistency [80]. Quality assessment of the evidence on the role of the dietary supplement curcumin in the treatment of ulcerative colitis yielded 10 moderate, 6 low, and 3 very low certainties of the evidence. The most deteriorating reasons were imprecision and publication bias [81]. The quality of evidence synthesis from meta-analyses on the effect of antibacterial envelopes in reducing infections associated with cardiac implantable electronic devices was found to be moderate in 60% of the outcomes in a recent paper, mostly due to the risk of bias and inconsistency [74].

As shown in Figure 2, all digital device interventions had significant effects on cardiac-related outcomes. According to the analyses results, we are highly confident that ICD has an impact on all-cause mortality (in 2 cases) and on the likelihood of transplantation. Moreover, we are highly confident about the impact of implantable and monitoring devices (ICD, iPhone-based rhythm monitoring device, and pacemakers) on the detection rate of atrial arrhythmias and stroke. Furthermore, the effect of some cardiac electronic devices (Metronome with a siren, HeartStart-MRx, Zoll AED, Cardio First Angel) on the return of spontaneous circulation created high-quality evidence but they come from studies with low and critically low methodological quality, which may raise some concerns about their results. The other interventions all have moderate-quality evidence synthesis, and we are moderately confident in the effect estimate. Furthermore, the actual effect is probably close to the effect estimate, but there is a possibility that it is substantially different. By contrast, these studies' low and critically low methodological quality raise concerns about the validity of the effect estimates. More than 263,000 electronic cardiac devices have been implanted annually in Germany, France, and the United Kingdom [82]. Device therapy has become increasingly important in treating life-threatening heart disease [83]. As a result, patients, clinicians, and policymakers are advised to consider the results of this study when making medical decisions.

Regarding the interventions with activity trackers (Figure 3), the vast majority had significant effects on human outcomes, whereas 16 outcomes were found to be ineffective in changing human health, including the effects of accelerometer, pedometers, YAMAX, Fitbit on disability, fatigue, functional tests, pain, and quality of life; the effects of activity monitor, portable tablet computers with touch screens, Fitbit, Jawbone UP24 wearable device, pedometer, and accelerometer on

moderate-to-vigorous physical activity; the effect of Fitbit on sedentary behavior; the effect of Fitbit, Jawbone UP, Polar Active, Misfit Flash, Gruve Solution, LUMOback, BodyMedia FIT, SenseWear, ActiveLink, InBodyBand on moderate-to-vigorous physical activity (in 1 case) and on steps (in 1 case); the effect of Fitbit, Jawbone UP24, combined heart rate monitor and accelerometer (Actiheart), wrist-worn accelerometer, FIT Core, Body Media, Fitbug Orb, and Polar FA20 accelerometer on physical activity (in one case) and on weight; the effect of Fitbit, Jawbone UP24, Gruve, LUMOback, Polar Active, Fitbug, Pebble+, Fitmeter, personal activity monitor, Withings Pulse on sedentary behavior; the effect of Garmin, Pedometer, Fitbit, Accelerometer, YAMAX Digi-walker, GEx sensor of vital signs, and smartphone on steps; the effect of wristbands and smartwatches on waist circumference; and the effect of pedometer on BMI. Most of these had moderate-quality evidence synthesis from studies with critically low methodological quality. By contrast, our confidence in some effect estimates is limited, and the actual effects may differ substantially from the estimated effects, including the effect of pedometer on steps in chronic respiratory diseases, obesity, and in sedentary older adults; the effect of Fitbit, YorBody, AiperMotion on moderate-to-vigorous physical activity; and the effect of activity monitor, portable tablet computers with touch screens, Fitbit, Jawbone UP24 wearable device, pedometer, accelerometer on moderate-to-vigorous physical activity, which did not have even significant effect. Our distrust increases when we find that these results come from critically low methodological quality studies. Evidence of moderate quality, as shown in Figure 3, suggests that the use of pedometers may increase physical activity; these results are from a study with high methodological quality [50]. Other reported outcomes had moderate-quality evidence with critically low methodological quality. According to our analysis, and as shown in Figure 3, there is no high-quality evidence of the impact of activity trackers on human health behavior change.

Figure 2. Cardiovascular-related interventions, outcomes, and methodological and evidence synthesis quality. HF: heart failure; ICD: implantable cardiac defibrillator. A: Cardiac resynchronization therapy, Implantable cardiac defibrillator [52], B: Fragmented QRS (fQRS) [70], C: Implantable cardiac defibrillator [53,63,68,69], D: Implantable cardiac defibrillator, iPhone-based rhythm monitoring device, pacemakers [55], E: Impedance devices [51], F: Implantable cardiac monitor, Holter-Electrocardiogram [65], G: Metronome with a siren, HeartStart-MRx, Zoll AED, Cardio First AngelTM [56], H: Pressure sensors [51], I: Pressure sensors and Impedance devices (Cardio MEMS, RVP sensor, Chronicle, ICD- OptiVol, InSync Sentry, lung impedance) [51].

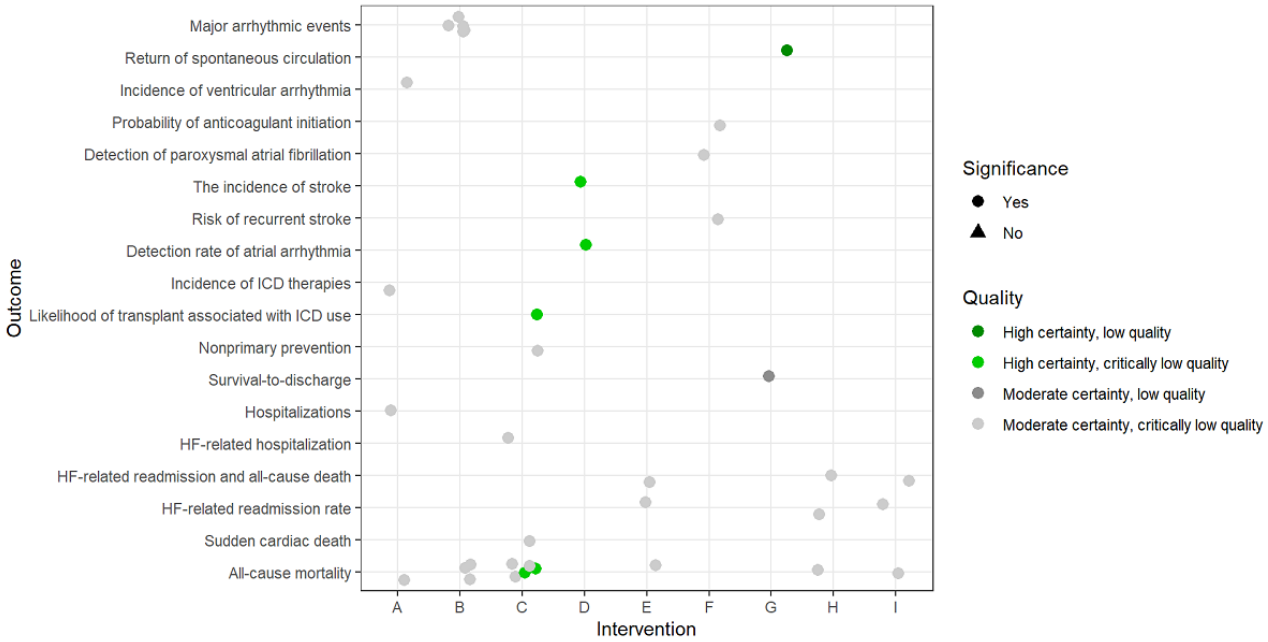
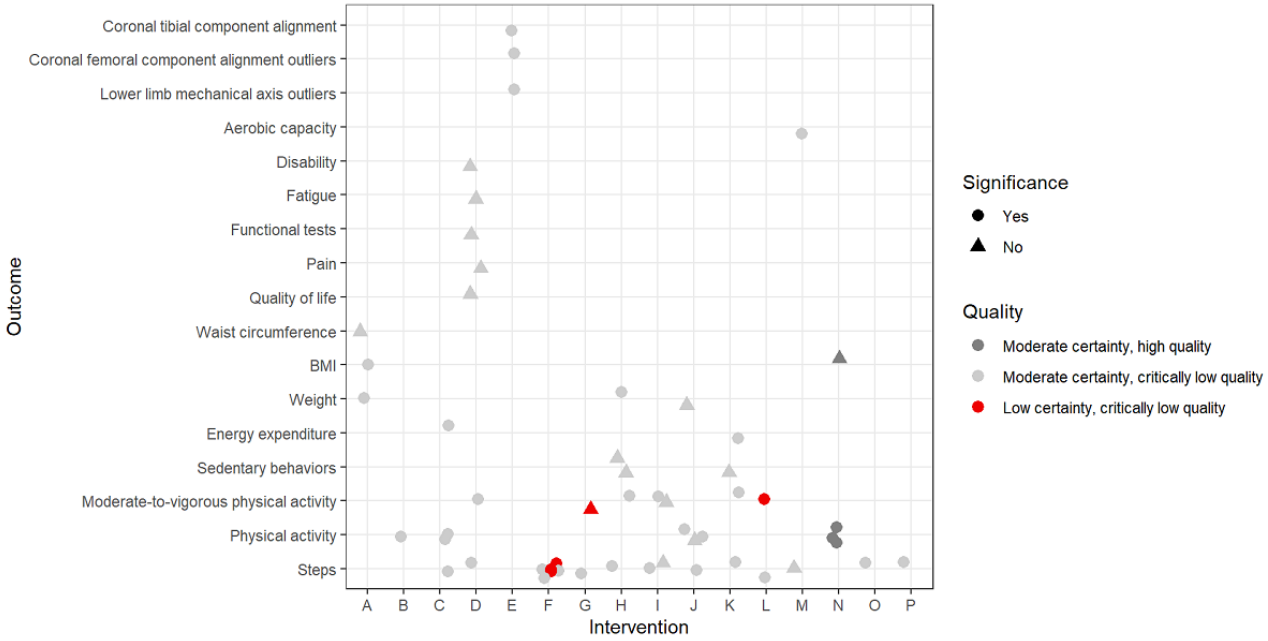


Figure 3. Activity trackers related to interventions, outcomes, and methodological and evidence synthesis quality. A: wristbands, smartwatches [54], B: Accelerometer, Dynaport MoveMonitor, Pedometer, Yamax Digi-walker CW700, ActivPal, ActiGraph, Personal Activity Monitor [67], C: Accelerometer, pedometer [60], D: Accelerometer, pedometers, Yamax, Fitbit [64], E: Accelerometer-based navigation system [66], F: wearable activity trackers (pedometer) [62], G: Activity monitor, portable tablet computers with touch screens, Fitbit, Jawbone UP24 wearable device, pedometer, accelerometer [59], H: Fitbit [57], I: Fitbit, Jawbone UP, Polar Active, Misfit Flash, Gruve Solution, LUMOback, BodyMedia Fit, SenseWear, ActiveLink, InBodyBand [47], J: Fitbit, Jawbone Up24, Combined heart rate monitor and accelerometer (Actiheart), Wrist-worn accelerometer, FIT Core, Body Media, Fitbug Orb, Polar FA20 accelerometer [49], K: Fitbit, Jawbone UP24, Gruve, LumoBack, Polar Active, Fitbug, Pebble+, Fitmeter, Personal Activity Monitor, Withings Pulse [46], L: Fitbit, Yorbody, AiperMotion [58], M: Garmin, Pedometer, Fitbit, Accelerometer, Yamax Digiwalker, Gex sensor of vital signs and smartphone [48], N: Pedometer [50], O: pedometer-based physical activity promotion [61], P: Pedometer physical activity promotion + pulmonary rehabilitation promotion [61].



Strengths

Most systematic review studies performed in the field of digital biomarkers in recent years have mainly been conducted with a specific focus on 1 or more disease areas or technologies, such as the effects of wearable fitness trackers on motivation and physical activity or ICD troubleshooting in patients with left ventricular assist devices. To our knowledge, no comprehensive systematic review of systematic reviews of all types of digital biomarkers has been published in all populations and in all diseases. Therefore, our review aims to assess the quality of methods and evidence of systematic reviews without limiting it to a specific domain or technology, using validated tools and standard methods. As a result, the strength of evidence can be compared between different types of interventions, providing practical guidance for clinicians and policymakers. To our knowledge, this is the first comprehensive study to address the methodological and evidence-based quality of systematic reviews of digital biomarker–based interventions. To categorize populations, interventions, outcomes, and behavioral/physiological data in digital biomarkers, we used World Health Organization (WHO) standard tools such as ICD-11, ICHI, and ICF. In addition, the most validated assessment tools, AMSTAR-2 and GRADE, were used to assess the methodological quality and quality of evidence synthesis of the systematic reviews.

Limitations

Despite the rigorous methodology, this study has some limitations, and readers are asked to consider the study's results in light of its limitations. One of the study's possible weaknesses is the short search duration (2019 and 2020). Only systematic reviews published in 2019 and 2020 were considered in this study according to the published protocol [23]. Because of the scope of the topic, we limited our assessment to a shorter period. However, given the new European Medical Devices Regulation (MDR) enacted in 2017 [84], we assumed this would be an exceptionally important period for evaluating clinical data collected before the regulations were implemented. While the 2-year period provides important insights into evidence syntheses published before MDR, longer periods would be needed to allow generalization of our findings.

As mentioned earlier, publication bias was assessed only in meta-analyses with at least ten studies. Of the 91 outcomes assessed, 67 included fewer than 10 studies, and we assessed publication bias in only 24 outcomes. In addition, the trim-and-fill approach, like any other method, may identify publication bias incorrectly in meta-analyses with a high degree of heterogeneity [85]. There were 2 outcomes where effect sizes were presented as a ratio of means. Thus, we interpreted the reported effect sizes as a mean difference to determine the optimal information size for assessing the imprecision. In 3 cases, the number of included studies in the meta-analyses was only 1. Therefore, an assessment of the quality of evidence was not possible for any of the GRADE criteria (risk of bias, publication bias, inconsistency, imprecision, and indirectness).

In our search, we operationalized the definition of digital biomarkers. However, we did not evaluate the sensitivity and

specificity of our search filter for articles on digital biomarkers. Besides the broad terms we used in our search strategy, digital biomarkers can be identified using terms related to the technology or type of data collected [3]. However, creating a complete list of appropriate search terms for all available technologies was beyond the scope of this study and remains an unresolved research topic. Specific sensor applications in the general population may raise health concerns (eg, COVID-19 contact–tracking apps [86]) that were not considered in this research. As recommended in the relevant guidelines for the systematic review of systematic reviews, we searched only the PubMed and Cochrane databases for reviews, and we did not search the Database of Abstracts of Reviews of Effectiveness (DARE) [87]. The DARE was not used in this study because it does not contain reviews from 2015. In addition, our published protocol required us to search gray literature; however, due to the large number of outcomes from peer-reviewed sources, we did not search gray literature.

In our search based on the definition of digital biomarkers and the inclusion criteria, we may have overlooked papers on digital biomarkers that were not defined by terms without the key adjectives used in the definition, as described earlier. Examples include thermometers and continuous glucose monitors. Thus, because of the ambiguity of definitions in digital health, more comprehensive keyword collections in this area are needed, as these were concluded in a recently accepted scoping review of digital biomarkers [88] and an ISPOR (International Society for Pharmacoeconomics and Outcomes Research) report [89].

Conclusion

In summary, we systematically reviewed the current evidence from systematic reviews on the use of digital biomarkers as interventions to change the health status of human populations. Overall, the 25 included current systematic reviews had critically low methodological quality, which may negatively affect the findings of the reported outcomes. In addition, most reported outcomes of interventions based on digital biomarkers had a moderate quality of evidence, implying that we have only moderate confidence in them. Only a small number of reported outcomes had high-quality evidence. Therefore, researchers in the field should consider the AMSTAR-2 criteria and GRADE to create future high-quality studies. Furthermore, patients, clinicians, and policymakers are advised to consider the results of this study before making clinical decisions relating to digital biomarkers.

Acknowledgments

This research under project no. TKP2020-NKA-02 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the Tématerületi Kiválósági Program funding scheme. During this research, HM-N, ZZ, MP, and LG received grant support from National Research, Development, and Innovation Fund of Hungary (TKP2021-NKTA-36 funding scheme).

Authors' Contributions

HM-N, LG, MP, and ZZ developed the concept. HM-N wrote the first manuscript draft. HA-A, HM-N, MF, and MMA performed the screening and data extraction. Data analysis was performed by HM-N and ZZ. All authors have commented on and approved the final manuscript. ZZ supervised the research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 18 KB - jmir_v24i12e41042_app1.docx]

Multimedia Appendix 2
Search strategies.
[DOCX File , 20 KB - jmir_v24i12e41042_app2.docx]

Multimedia Appendix 3
List of excluded studies.
[DOCX File , 70 KB - jmir_v24i12e41042_app3.docx]

Multimedia Appendix 4
Characteristics of the included studies.
[DOCX File , 47 KB - jmir_v24i12e41042_app4.docx]

Multimedia Appendix 5

Assessment of the methodological quality of reviews using the AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews 2) tool.
[DOCX File , 29 KB - jmir_v24i12e41042_app5.docx]

Multimedia Appendix 6
Evidence summary and quality assessment by the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) tool.
[DOCX File , 52 KB - jmir_v24i12e41042_app6.docx]

References

1. Califf RM. Biomarker definitions and their applications. Exp Biol Med (Maywood) 2018 Feb;243(3):213-221 [FREE Full text] [doi: 10.1177/1535370217750088] [Medline: 29405771]
2. Babrak L, Menetski J, Rebhan M, Nisato G, Zinggeler M, Brasier N, et al. Traditional and Digital Biomarkers: Two Worlds Apart? Digit Biomark 2019 Aug 16;3(2):92-102 [FREE Full text] [doi: 10.1159/000502000] [Medline: 32095769]
3. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019;2(1):40 [FREE Full text] [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
4. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018 Aug;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]
5. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
6. Shin EK, Mahajan R, Akbilgic O, Shaban-Nejad A. Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits. NPJ Digit Med 2018 Oct 2;1(1):50 [FREE Full text] [doi: 10.1038/s41746-018-0056-y] [Medline: 31304329]
7. Guthrie NL, Carpenter J, Edwards KL, Appelbaum KJ, Dey S, Eisenberg DM, et al. Emergence of digital biomarkers to predict and modify treatment efficacy: machine learning study. BMJ Open 2019 Jul 23;9(7):e030710 [FREE Full text] [doi: 10.1136/bmjopen-2019-030710] [Medline: 31337662]
8. BisResearch. Global Digital Biomarkers Market Focus on Key Trends, Growth Potential, Competitive Landscape, Components (Data Collection and Integration), End Users, Application (Sleep and Movement, Neuro, Respiratory and Cardiological Disorders) and Region – Analysis Internet. BisResearch. 2020. URL: https://bisresearch.com/industry-report/digital-biomarkers-market.html [accessed 2022-06-22]
9. Meister S, Deiters W, Becker S. Digital health and digital biomarkers – enabling value chains on health data. Curr Dir Biomed Eng 2016 Sep 1;2(1):577-581. [doi: 10.1515/cdbme-2016-0128]
10. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The Effectiveness of Smartphone Apps for Lifestyle Improvement in Noncommunicable Diseases: Systematic Review and Meta-Analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
11. Lenouvel E, Novak L, Nef T, Klöppel S. Advances in Sensor Monitoring Effectiveness and Applicability: A Systematic Review and Update. Gerontologist 2020 May 15;60(4):e299-e308. [doi: 10.1093/geront/gnz049] [Medline: 31102436]
12. Nagy TD, Haidegger T. A DVRK-based Framework for Surgical Subtask Automation. APH 2019 Sep 09;16(8):61-78. [doi: 10.12700/aph.16.8.2019.8.5]
13. Haidegger T. Probabilistic Method to Improve the Accuracy of Computer-Integrated Surgical Systems. APH 2019 Sep 09;16(8):119-140. [doi: 10.12700/aph.16.8.2019.8.8]
14. Biggs JS, Willcocks A, Burger M, Makeham MA. Digital health benefits evaluation frameworks: building the evidence to support Australia's National Digital Health Strategy. Med J Aust 2019 Apr;210 Suppl 6:S9-S11. [doi: 10.5694/mja2.50034] [Medline: 30927475]
15. Enam A, Torres-Bonilla J, Eriksson H. Evidence-Based Evaluation of eHealth Interventions: Systematic Literature Review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
16. Lau F, Kuziemsky C. Handbook of eHealth Evaluation: An Evidence-based Approach Internet. Victoria, BC, Canada: University of Victoria; 2017.
17. NICE. Evidence Standards Framework for Digital Health Technologies. NICE. 2019. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2022-05-12]
18. Zah V, Burrell A, Zrubka ZA, Asche C. HTA30 Paying for Digital Health Interventions: A Rapid Review of Payers' Evidence Frameworks. Value in Health 2022 Jul;25(7):S509. [doi: 10.1016/j.jval.2022.04.1162]
19. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]

20. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res 2004 Dec 22;4(1):38 [FREE Full text] [doi: 10.1186/1472-6963-4-38] [Medline: 15615589]
21. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
22. 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. Int J Surg 2021 Apr;88:105906. [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
23. Motahari-Nezhad H, Péntek M, Gulácsi L, Zrubka Z. Outcomes of Digital Biomarker-Based Interventions: Protocol for a Systematic Review of Systematic Reviews. JMIR Res Protoc 2021 Nov 24;10(11):e28204 [FREE Full text] [doi: 10.2196/28204] [Medline: 34821568]
24. Coulson S, Woods M. Citizen Sensing: An Action-Orientated Framework for Citizen Science. Front. Commun 2021 Apr 16;6:1-13. [doi: 10.3389/fcomm.2021.629700]
25. National Institutes of Health (NIH). Search Strategy Used to Create the PubMed Systematic Reviews Filter. National Library of Medicine. 2018. URL: https://www.nlm.nih.gov/bsd/pubmed_subsets/sysreviews_strategy.html [accessed 2021-06-15]
26. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/BM.2012.031]
27. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet 2002 Jan 05;359(9300):57-61. [doi: 10.1016/S0140-6736(02)07283-5] [Medline: 11809203]
28. World Health Organization (WHO). ICD-11 International Classification of Diseases 11th Revision The global standard for diagnostic health information Internet. World Health Organization (WHO). Geneva, Switzerland: World Health Organization (WHO); 2020. URL: https://icd.who.int/en [accessed 2022-02-25]
29. World Health Organization (WHO). International Classification of Health Interventions (ICHI) Internet. World Health Organization (WHO). Geneva, Switzerland: World Health Organization (WHO); 2020. URL: https://www.who.int/standards/classifications/international-classification-of-health-interventions [accessed 2022-02-25]
30. World Health Organization (WHO). ICF browser. World Health Organization (WHO). Geneva, Switzerland: World Health Organization (WHO); 2017. URL: https://apps.who.int/classifications/icfbrowser/ [accessed 2022-02-25]
31. Lorenz RC, Matthias K, Pieper D, Wegewitz U, Morche J, Nocon M, et al. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. J Clin Epidemiol 2019 Oct;114:133-140. [doi: 10.1016/j.jclinepi.2019.05.028] [Medline: 31152864]
32. Pieper D, Lorenz RC, Rombey T, Jacobs A, Rissling O, Freitag S, et al. Authors should clearly report how they derived the overall rating when applying AMSTAR 2-a cross-sectional study. J Clin Epidemiol 2021 Jan;129:97-103. [doi: 10.1016/j.jclinepi.2020.09.046] [Medline: 33049325]
33. Schünemann H, Brożek J, Guyatt G, Oxman A. Introduction to GRADE Handbook Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Gradepro. 2013. URL: https://gdt.gradepro.org/app/handbook/handbook.html#h.9rdbelsnu4iy [accessed 2022-02-07]
34. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol 2011 Apr;64(4):407-415. [doi: 10.1016/j.jclinepi.2010.07.017] [Medline: 21247734]
35. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol 2011 Dec;64(12):1294-1302. [doi: 10.1016/j.jclinepi.2011.03.017] [Medline: 21803546]
36. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol 2011 Dec;64(12):1283-1293. [doi: 10.1016/j.jclinepi.2011.01.012] [Medline: 21839614]
37. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol 2011 Dec;64(12):1277-1282. [doi: 10.1016/j.jclinepi.2011.01.011] [Medline: 21802904]
38. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol 2011 Dec;64(12):1303-1310. [doi: 10.1016/j.jclinepi.2011.04.014] [Medline: 21802903]
39. Pollock A, Farmer SE, Brady MC, Langhorne P, Mead GE, Mehrholz J, et al. An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews. J Clin Epidemiol 2016 Feb;70:106-110 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.013] [Medline: 26341023]
40. Zhang Y, Coello PA, Guyatt GH, Yepes-Nuñez JJ, Akl EA, Hazlewood G, et al. GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes or values and preferences-inconsistency, imprecision, and other domains. J Clin Epidemiol 2019 Jul;111:83-93. [doi: 10.1016/j.jclinepi.2018.05.011] [Medline: 29800687]
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences. London, UK: Routledge; 1988:1-567.

42. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews. London, UK: Cochrane; 2019:1-694.
43. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
44. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol 2005 Sep;58(9):894-901. [doi: 10.1016/j.jclinepi.2005.01.006] [Medline: 16085192]
45. Hodkinson A, Kontopantelis E, Adeniji C, van Marwijk H, McMillan B, Bower P, et al. Accelerometer- and pedometer-based physical activity interventions among adults with cardiometabolic conditions: a systematic review and meta-analysis. JAMA Netw Open 2019 Oct 02;2(10):e1912895 Retracted in: JAMA Netw Open. 2020 Dec 1;3(12):e2032700. [doi: 10.1001/jamanetworkopen.2020.32700] [Medline: 31596494]
46. 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]
47. Lynch C, Bird S, Lythgo N, Selva-Raj I. Changing the Physical Activity Behavior of Adults With Fitness Trackers: A Systematic Review and Meta-Analysis. Am J Health Promot 2020 May;34(4):418-430. [doi: 10.1177/0890117119895204] [Medline: 31858812]
48. Hannan AL, Harders MP, Hing W, Climstein M, Coombes JS, Furness J. Impact of wearable physical activity monitoring devices with exercise prescription or advice in the maintenance phase of cardiac rehabilitation: systematic review and meta-analysis. BMC Sports Sci Med Rehabil 2019 Jul 30;11(1):14 [FREE Full text] [doi: 10.1186/s13102-019-0126-8] [Medline: 31384474]
49. Tang MSS, Moore K, McGavigan A, Clark RA, Ganesan AN. Effectiveness of Wearable Trackers on Physical Activity in Healthy Adults: Systematic Review and Meta-Analysis of Randomized Controlled Trials. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e15576 [FREE Full text] [doi: 10.2196/15576] [Medline: 32706685]
50. Freak-Poli RL, Cumpston M, Albarqouni L, Clemes SA, Peeters A. Workplace pedometer interventions for increasing physical activity. Cochrane Database Syst Rev 2020 Jul 21;7:CD009209 [FREE Full text] [doi: 10.1002/14651858.CD009209.pub3] [Medline: 32700325]
51. Halawa A, Enezate T, Flaker G. Device monitoring in heart failure management: outcomes based on a systematic review and meta-analysis. Cardiovasc Diagn Ther 2019 Aug;9(4):386-393 [FREE Full text] [doi: 10.21037/cdt.2019.01.02] [Medline: 31555544]
52. Voruganti DC, Briasoulis A, Chaudhry M, Alvarez P, Cotarlan V, Bhama JK, et al. Cardiac resynchronization therapy and outcomes in patients with left ventricular assist devices: a systematic review and meta-analysis. Heart Fail Rev 2019 Mar;24(2):229-236. [doi: 10.1007/s10741-018-9740-x] [Medline: 30259285]
53. Elkaryoni A, Badarin FA, Khan MS, Ellakany K, Potturi N, Poonia J, et al. Implantable cardioverter-defibrillators and survival in advanced heart failure patients with continuous-flow left ventricular assist devices: a systematic review and meta-analysis. Europace 2019 Sep 01;21(9):1353-1359. [doi: 10.1093/europace/euz125] [Medline: 31086951]
54. Yen H, Chiu H. The effectiveness of wearable technologies as physical activity interventions in weight control: A systematic review and meta-analysis of randomized controlled trials. Obes Rev 2019 Oct;20(10):1485-1493. [doi: 10.1111/obr.12909] [Medline: 31342646]
55. Jang J, Lin H, Chen Y, Hsieh M, Huang Y. Role of Remote Monitoring in Detection of Atrial Arrhythmia, Stroke Reduction, and Use of Anticoagulation Therapy - A Systematic Review and Meta-Analysis. Circ J 2020 Oct 23;84(11):1922-1930. [doi: 10.1253/circj.CJ-20-0633] [Medline: 33012748]
56. Wang S, Su C, Fan H, Hou W, Chen Y. Effects of real-time feedback on cardiopulmonary resuscitation quality on outcomes in adult patients with cardiac arrest: A systematic review and meta-analysis. Resuscitation 2020 Oct;155:82-90. [doi: 10.1016/j.resuscitation.2020.07.024] [Medline: 32755666]
57. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
58. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. Am J Health Promot 2019 Jun;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]
59. Liu JY, Kor PP, Chan CP, Kwan RY, Sze-Ki D. The effectiveness of a wearable activity tracker (WAT)-based intervention to improve physical activity levels in sedentary older adults: A systematic review and meta-analysis. Arch Gerontol Geriatr 2020 Jul 26;91:104211. [doi: 10.1016/j.archger.2020.104211] [Medline: 32739713]
60. Kwan RYC, Salihu D, Lee PH, Tse M, Cheung DSK, Roopsawang I, et al. The effect of e-health interventions promoting physical activity in older people: a systematic review and meta-analysis. Eur Rev Aging Phys Act 2020 Apr 21;17(1):7 [FREE Full text] [doi: 10.1186/s11556-020-00239-5] [Medline: 32336996]
61. Armstrong M, Winnard A, Chynkiamis N, Boyle S, Burtin C, Vogiatzis I. Use of pedometers as a tool to promote daily physical activity levels in patients with COPD: a systematic review and meta-analysis. Eur Respir Rev 2019 Dec 31;28(154):190039 [FREE Full text] [doi: 10.1183/16000617.0039-2019] [Medline: 31722891]

62. Franssen WMA, Franssen GHLM, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020 May 11;17(1):57 [FREE Full text] [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
63. Liu Y, Sun J, Zhu Y, Li Z, Li K, Wang R. Association between CRT(D)/ICD and renal insufficiency: A systematic review and meta-analysis. Semin Dial 2021 Jan;34(1):17-30. [doi: 10.1111/sdi.12937] [Medline: 33296540]
64. Davergne T, Pallot A, Dechartres A, Fautrel B, Gossec L. Use of Wearable Activity Trackers to Improve Physical Activity Behavior in Patients With Rheumatic and Musculoskeletal Diseases: A Systematic Review and Meta-Analysis. Arthritis Care Res (Hoboken) 2019 Jun;71(6):758-767. [doi: 10.1002/acr.23752] [Medline: 30221489]
65. Tsivgoulis G, Katsanos AH, Grory BM, Köhrmann M, Ricci BA, Tsioufis K, et al. Prolonged Cardiac Rhythm Monitoring and Secondary Stroke Prevention in Patients With Cryptogenic Cerebral Ischemia. Stroke 2019 Aug;50(8):2175-2180. [doi: 10.1161/STROKEAHA.119.025169] [Medline: 31216964]
66. Shigemura T, Murata Y, Yamamoto Y, Mizuki N, Toki Y, Wada Y. No definite advantage of a portable accelerometer-based navigation system over conventional technique in total knee arthroplasty: A systematic review and meta-analysis. Orthop Traumatol Surg Res 2019 Jun;105(4):619-626 [FREE Full text] [doi: 10.1016/j.otsr.2019.03.006] [Medline: 31040097]
67. Braakhuis HEM, Berger MAM, Bussmann JBJ. Effectiveness of healthcare interventions using objective feedback on physical activity: A systematic review and meta-analysis. J Rehabil Med 2019 Mar 13;51(3):151-159 [FREE Full text] [doi: 10.2340/16501977-2522] [Medline: 30843082]
68. Gama F, Ferreira J, Carmo J, Costa FM, Carvalho S, Carmo P, et al. Implantable Cardioverter-Defibrillators in Trials of Drug Therapy for Heart Failure: A Systematic Review and Meta-Analysis. J Am Heart Assoc 2020 Apr 21;9(8):e015177 [FREE Full text] [doi: 10.1161/JAHA.119.015177] [Medline: 32290732]
69. Alotaibi S, Hernandez-Montfort J, Ali OE, El-Chilali K, Perez BA. Remote monitoring of implantable cardiac devices in heart failure patients: a systematic review and meta-analysis of randomized controlled trials. Heart Fail Rev 2020 May;25(3):469-479. [doi: 10.1007/s10741-020-09923-1] [Medline: 32002732]
70. Kanitsoraphan C, Rattanawong P, Mekraksakit P, Chongsathidkiet P, Riangwiwat T, Kanjanahattakij N, et al. Baseline fragmented QRS is associated with increased all-cause mortality in heart failure with reduced ejection fraction: A systematic review and meta-analysis. Ann Noninvasive Electrocardiol 2019 Mar 17;24(2):e12597 [FREE Full text] [doi: 10.1111/anec.12597] [Medline: 30329201]
71. Stamatakis E, Gale J, Bauman A, Ekelund U, Hamer M, Ding D. Sitting Time, Physical Activity, and Risk of Mortality in Adults. J Am Coll Cardiol 2019 Apr 30;73(16):2062-2072 [FREE Full text] [doi: 10.1016/j.jacc.2019.02.031] [Medline: 31023430]
72. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of Digital Interventions for Reducing Behavioral Risks of Cardiovascular Disease in Nonclinical Adult Populations: Systematic Review of Reviews. J Med Internet Res 2021 May 14;23(5):e19688 [FREE Full text] [doi: 10.2196/19688] [Medline: 33988126]
73. Villarreal-Zegarra D, Alarcon-Ruiz CA, Melendez-Torres G, Torres-Puente R, Navarro-Flores A, Cavero V, et al. Development of a Framework for the Implementation of Synchronous Digital Mental Health: Realist Synthesis of Systematic Reviews. JMIR Ment Health 2022 Mar 29;9(3):e34760 [FREE Full text] [doi: 10.2196/34760] [Medline: 35348469]
74. Motahari-nezhad H, Miribonjar M, Sadeghdaghighi A. Methodological and Evidence Synthesis Quality Evaluation of Meta- Analyses Assessing the Effect of Antibacterial Envelopes to Reduce CIED-Related Infections. Int J Med Res Heal Sci 2022 Apr;11(2):65-76 [FREE Full text]
75. Finucane AM, O'Donnell H, Lugton J, Gibson-Watt T, Swenson C, Pagliari C. Digital health interventions in palliative care: a systematic meta-review. NPJ Digit Med 2021 Apr 06;4(1):64 [FREE Full text] [doi: 10.1038/s41746-021-00430-7] [Medline: 33824407]
76. Dumit EM, Novillo-Ortiz D, Contreras M, Velandia M, Danovaro-Holliday MC. The use of eHealth with immunizations: An overview of systematic reviews. Vaccine 2018 Dec 18;36(52):7923-7928 [FREE Full text] [doi: 10.1016/j.vaccine.2018.06.076] [Medline: 29983255]
77. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
78. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med 2002 Jun 15;21(11):1575-1600. [doi: 10.1002/sim.1188] [Medline: 12111921]
79. Xin Z, Xue-Ting L, De-Ying K. GRADE in Systematic Reviews of Acupuncture for Stroke Rehabilitation: Recommendations based on High-Quality Evidence. Sci Rep 2015 Nov 12;5:16582 [FREE Full text] [doi: 10.1038/srep16582] [Medline: 26560971]
80. Liang M, Yan L, Mei Z, Luo Y, Hou X, Feng Z. Methodological and reporting quality evaluation of meta-analyses on the Chinese herbal preparation Zheng Qing Feng Tong Ning for the treatment of rheumatoid arthritis. BMC Complement Med Ther 2020 Jun 26;20(1):195 [FREE Full text] [doi: 10.1186/s12906-020-02978-5] [Medline: 32586308]
81. Shi H, Wang D, Chen W, Li Y, Si G, Yang T. Quality of Evidence Supporting the Role of Supplement Curcumin for the Treatment of Ulcerative Colitis: An Overview of Systematic Reviews. Gastroenterol Res Pract 2022;2022:3967935 [FREE Full text] [doi: 10.1155/2022/3967935] [Medline: 35251163]

82. Hindricks G, Padeletti L, Raatikainen P, Arnar D. The current status of cardiac electrophysiology in ESC member countries. The EHRA White Book 2014 2014:1-572.
83. Hussein AA, Wilkoff BL. Cardiac Implantable Electronic Device Therapy in Heart Failure. Circ Res 2019 May 24;124(11):1584-1597. [doi: 10.1161/CIRCRESAHA.118.313571] [Medline: 31120815]
84. Thacker N, Singh P. Regulation (EU) 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/EE. MDR. 2017 Aug. URL: https://www.medical-device-regulation.eu/download-mdr/ [accessed 2022-11-06]
85. Shi L, Lin L. The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. Medicine (Baltimore) 2019 Jun;98(23):e15987 [FREE Full text] [doi: 10.1097/MD.0000000000015987] [Medline: 31169736]
86. Garousi V, Cutting D, Felderer M. Mining user reviews of COVID contact-tracing apps: An exploratory analysis of nine European apps. J Syst Softw 2022 Feb;184:111136 [FREE Full text] [doi: 10.1016/j.jss.2021.111136] [Medline: 34751198]
87. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011 Feb 03;11(1):15 [FREE Full text] [doi: 10.1186/1471-2288-11-15] [Medline: 21291558]
88. Motahari-Nezhad H, Fgaier M, Mahdi Abid M, Péntek M, Gulácsi L, Zrubka Z. Digital Biomarker-Based Studies: Scoping Review of Systematic Reviews. JMIR Mhealth Uhealth 2022 Oct 24;10(10):e35722 [FREE Full text] [doi: 10.2196/35722] [Medline: 36279171]
89. Burrell A, Zrubka Z, Champion A, Zah V, Vinuesa L, Holtorf A, ISPOR Digital Health Special Interest Group Key Project. How Useful Are Digital Health Terms for Outcomes Research? An ISPOR Special Interest Group Report. Value Health 2022 Sep;25(9):1469-1479. [doi: 10.1016/j.jval.2022.04.1730] [Medline: 36049797]

Abbreviations

AMSTAR-2: A Measurement Tool to Assess Systematic Reviews 2
DARE: Database of Abstracts of Reviews of Effectiveness
DBM: digital biomarker
GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
HF: heart failure
ICD: implantable cardiac defibrillator
ICD-11: International Statistical Classification of Diseases and Related Health Problems
ICF: International Classification of Functioning, Disability and Health
ICHI: International Classification of Health Interventions
ISPOR: International Society for Pharmacoeconomics and Outcomes Research
MDR: Medical Devices Regulation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

Edited by R Kukafka, G Eysenbach; submitted 13.07.22; peer-reviewed by N Mir, A Roghanizad; comments to author 17.08.22; revised version received 22.09.22; accepted 07.10.22; published 21.12.22.

Please cite as:
Motahari-Nezhad H, Al-Abdulkarim H, Fgaier M, Abid MM, Péntek M, Gulácsi L, Zrubka Z
Digital Biomarker–Based Interventions: Systematic Review of Systematic Reviews
J Med Internet Res 2022;24(12):e41042
URL: https://www.jmir.org/2022/12/e41042
doi:10.2196/41042
PMID:36542427

Review

Web-Based Service Provision of HIV, Viral Hepatitis, and Sexually Transmitted Infection Prevention, Testing, Linkage, and Treatment for Key Populations: Systematic Review and Meta-analysis

Ping Teresa Yeh[1], MSPH; Caitlin Elizabeth Kennedy[1], MPH, PhD; Ayako Minamitani[1], MPH, MD; Rachel Baggaley[2], MBBS, MSc; Purvi Shah[2,3], MA; Annette Verster[2], MA; Niklas Luhmann[2], MScPH; Maeve Brito de Mello[2], PhD; Virginia Macdonald[2], MPHTM, PhD

[1]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Testing, Prevention and Population Unit, Global HIV, Hepatitis and STI Programmes, World Health Organization, Geneva, Switzerland

[3]Regional Support Team Asia Pacific, Joint United Nations Programme on HIV/AIDS, New Delhi, India

Corresponding Author:
Ping Teresa Yeh, MSPH
Department of International Health
Johns Hopkins Bloomberg School of Public Health
615 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 410 955 3934
Email: teresa.yeh@jhu.edu

Abstract

Background: Despite the growth of web-based interventions for HIV, viral hepatitis (VH), and sexually transmitted infections (STIs) for key populations, the evidence for the effectiveness of these interventions has not been reported.

Objective: This study aimed to inform the World Health Organization guidelines for HIV, VH, and STI prevention, diagnosis, and treatment services for key populations by systematically reviewing the effectiveness, values and preferences, and costs of web-based outreach, web-based case management, and targeted web-based health information for key populations (men who have sex with men, sex workers, people who inject drugs, trans and gender-diverse people, and people in prisons and other closed settings).

Methods: We searched CINAHL, PsycINFO, PubMed, and Embase in May 2021 for peer-reviewed studies; screened abstracts; and extracted data in duplicate. The effectiveness review included randomized controlled trials (RCTs) and observational studies. We assessed the risk of bias using the Cochrane Collaboration tool for RCTs and the Evidence Project and Risk of Bias in Non-randomized Studies of Interventions tools for non-RCTs. Values and preferences and cost data were summarized descriptively.

Results: Of 2711 records identified, we included 13 (0.48%) articles in the effectiveness review (3/13, 23% for web-based outreach; 7/13, 54% for web-based case management; and 3/13, 23% for targeted web-based health information), 15 (0.55%) articles in the values and preferences review, and 1 (0.04%) article in the costs review. Nearly all studies were conducted among men who have sex with men in the United States. These articles provided evidence that web-based approaches are as effective as face-to-face services in terms of reaching new people, use of HIV, VH, and STI prevention services, and linkage to and retention in HIV care. A meta-analysis of 2 RCTs among men who have sex with men in China found increased HIV testing after web-based outreach (relative risk 1.39, 95% CI 1.21-1.60). Among men who have sex with men in the United States, such interventions were considered feasible and acceptable. One cost study among Canadian men who have sex with men found that syphilis testing campaign advertisements had the lowest cost-per-click ratio on *hookup* platforms compared with more traditional social media platforms.

Conclusions: Web-based services for HIV, VH, and STIs may be a feasible and acceptable approach to expanding services to key populations with similar outcomes as standard of care, but more research is needed in low-resource settings, among key populations other than men who have sex with men, and for infections other than HIV (ie, VH and STIs).

(J Med Internet Res 2022;24(12):e40150) doi:10.2196/40150

KEYWORDS

online service delivery; digital health interventions; HIV; viral hepatitis; sexually transmitted infections; key populations; systematic review; mobile phone

Introduction

Background

Access to the internet and social media apps has grown exponentially in recent years, including in low-income settings and among vulnerable populations [1-4]. The potential advantages of providing health-related services on the web include reaching a broader audience; reaching people who are geographically isolated or who may not seek community or clinic-based services; targeting information to different groups and individuals; improving financial and system efficiencies; reducing stigma and discrimination; and offering clients greater anonymity, convenience, and potential self-care [5]. Thus, services are increasingly "going online," including services designed for key populations who are disproportionately affected by HIV, viral hepatitis (VH), and sexually transmitted infections (STIs)—men who have sex with men, sex workers, people who inject drugs, trans and gender-diverse people, and people in prisons and other closed settings.

Web-based health interventions can take a wide variety of forms [6]. For key populations, common strategies include web-based outreach, web-based case management, and targeted web-based health information. Web-based outreach seeks to identify potential key population service users through web-based platforms, such as websites and social media apps, where key populations communicate, learn information, and socialize. Web-based case management can support key populations who have tested positive and need to engage in services to assess risk and adhere to necessary treatment as well as those who have tested negative and need counseling or biomedical prevention options such as preexposure prophylaxis or continued regular self-testing. Providing case management through web-based systems could potentially reduce loss to follow-up and provide behavioral nudges (such as reminders to book an appointment, take a test, or take a medication). Targeted web-based health information uses internet sites and social networking apps to target communication according to user demographics and characteristics. For example, Facebook advertisements can target users of certain ages, social profiles, geographic locations, or other attributes. Population segmentation may allow for more specific targeting of key population audiences to tailored information or linkage to health services.

Objective

Despite the growth of web-based interventions for key populations, particularly accelerated during the COVID-19 pandemic, there has not been a synthesis of the effectiveness of these strategies across key populations. A review of digital health interventions addressing sexual risk, substance use, and common mental health conditions among men who have sex with men [7,8] found such interventions to be acceptable across sociodemographic groups and usually based on individual-level theoretical constructs such as self-efficacy, motivation, behavioral intentions, attitudes, and perceived norms. We identified no similar reviews on digital health interventions with sex workers, people who inject drugs, people in prisons and other closed settings, or trans and gender-diverse populations. To inform the World Health Organization (WHO) guidelines for HIV, VH, and STI service delivery for key populations, we systematically reviewed the effectiveness, values and preferences, and costs of web-based outreach, web-based case management, and targeted web-based health information.

Methods

Overview

This systematic review addressed the following question: *Does providing services on the web improve uptake of HIV, VH, and STI prevention, testing, linkage to treatment, and treatment retention for key populations?* We reviewed the extant literature in 3 related areas, which are components of the evidence-based process used to inform WHO guideline development [9]: (1) effectiveness of the intervention, (2) values and preferences of end users and health workers related to the intervention, and (3) cost information. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [10].

Ethics Approval

Ethics approval was not required for this systematic review as all the data were obtained from published articles.

Effectiveness Review: Eligibility Criteria

We designed the effectiveness review according to the PICO (population, intervention, comparison, and outcomes) format.

Population

The population of interest for the review includes (as described in the Glossary section of the 2022 WHO consolidated guidelines on HIV, VH, and STI prevention, diagnosis, treatment, and care for key populations [11]) men who have sex with men, sex workers, people who inject drugs, trans and gender-diverse people, and people in prisons and other closed settings.

Intervention

Given the range of web-based health interventions, the effectiveness review was split into 3 separate categories of web-based interventions: web-based outreach, web-based case management, and targeted web-based health information. We excluded noninternet phone-based (SMS text messaging or telephone call) outreach and case management approaches, counseling interventions, and general informational and educational videos about HIV, VH, and STI prevention. We excluded studies that used web-based methods to recruit study participants or to deliver the intervention of interest (eg, HIV self-testing) but did not compare service delivery modalities in their outcomes. We also excluded web-based service delivery

interventions that did not specifically target key populations, even if many of their users were members of key populations, if the data were not disaggregated to specific key populations. The following definitions were used:

- *Outreach for HIV, VH, and STI services through web-based platforms* was defined as outreach conducted in any way through the internet (website or app, accessed through any device). Web-based outreach aims to identify and reach key populations who have previously not had contact with health services. Outreach can be active or passive, including reaching out to potential clients for information and linking them to services such as referrals, follow-up reminders, and counseling. These outreach approaches may include web-based peer-to-peer networks, social media influencers, and advertisements. These interventions aim to reach populations using social media platforms and dating apps, who do not necessarily identify themselves as key populations but are equally vulnerable and need to be linked to services across the cascade.
- *Web-based case management* was defined as web-based methods for case managers to conduct risk screening, referral, partner notification, appointment scheduling, and reminders for prevention, testing and treatment services, treatment adherence support, follow-up, counseling, telemedicine, and home delivery of services.
- *Targeted web-based health information* was defined as a web-based awareness generation or demand creation method. These targeted approaches deliver information and education, behavior change communication, or health service advertisements through mechanisms such as Facebook and dating apps, where data can be mined via algorithms to target specific messages to different key population groups (population segmentation and microtargeting). For example, targeted messages that aim to bring about behavioral change by motivating or mobilizing followers to get tested or learn about prevention can be passed through social media influencers.

Comparator

The comparator group (for each of the 3 intervention categories) was standard of care.

Outcomes

The outcomes of interest for each intervention category are presented in Table 1.

Table 1. Outcomes of interest by category of web-based service delivery.

	Web-based outreach	Web-based case management	Targeted web-based health information
Number or proportion of previously unreached people reached	✓		
Use of prevention services (eg, PrEP[a] uptake, PrEP adherence, PEP[b] uptake, PEP adherence, counseling, and condoms)	✓	✓	✓
Uptake of testing services for HIV, VH[c], and STIs[d]	✓	✓	✓
Treatment initiation for HIV, VH, and STIs	✓	✓	✓
Treatment retention or completion for HIV, VH, and STIs		✓	✓
Viral load (eg, HIV and HCV[e]) testing or suppression		✓	✓
Cure (for curable STIs, eg, HCV, syphilis, and gonorrhea)		✓	✓
Mortality		✓	✓

[a]PrEP: preexposure prophylaxis.

[b]PEP: postexposure prophylaxis.

[c]VH: viral hepatitis.

[d]STI: sexually transmitted infection.

[e]HCV: hepatitis C virus.

To be included in the effectiveness review, an article must have (1) had a study design that compared web-based service delivery with standard of care for key populations, including randomized controlled trials (RCTs), non-RCTs, and comparative observational studies (including prospective controlled cohort studies, retrospective controlled cohort studies, cross-sectional studies, controlled before-after studies, and interrupted time series) that compared individuals who received the intervention with those who did not; (2) measured ≥1 of the outcomes of interest for that category of interventions; and (3) been published in a peer-reviewed journal between January 1, 2010, and the search date of May 27, 2021. No restrictions were placed based on location of the intervention or language of the publication.

Search Strategy and Screening

We searched 4 databases (CINAHL, PsycINFO, PubMed, and Embase) for relevant peer-reviewed publications. Search terms covered terms for key populations, infections (HIV, VH, STIs), and web-based service delivery interventions. The full search strategy is presented in Multimedia Appendix 1. This search was complemented by several other methods of identifying articles. First, we ran 2 earlier searches for behavioral interventions for key populations more broadly—one in 2020 as part of a scoping review and one in March 2021 as part of a search only for RCTs. Articles identified through these prior searches were included in the review. Second, we hand searched the references of articles identified for inclusion in the review.

Third, we contacted experts in the field (including members of the WHO Key Population Guideline Development Group) to identify any additional articles that we may have missed.

Titles, abstracts, citation information, and descriptor terms of citations identified through the search strategy were screened for initial inclusion. Full-text articles were obtained of all selected abstracts, and 2 independent reviewers assessed all full-text articles for eligibility to determine final study selection. Differences were resolved through consensus.

Data Management and Analysis

Two reviewers independently abstracted data using standardized forms in Microsoft Excel (Microsoft Corporation). Differences in data abstraction were resolved through consensus and referral to a senior study team member as necessary. We collected the following information from each article: study identification (author, year, title, journal, and language of article), location (country, urban or rural, World Bank income classification, and WHO region), key population description (gender and age), sample size (n), study design (including follow-up periods and loss to follow-up), intervention description (including who delivered intervention, where intervention was provided, and how long or frequent intervention was), comparator description, and study outcomes (analytic approach, outcome measures or definitions, intervention vs comparison group, frequency and percentage or effect sizes with CIs or significance levels, conclusions, and limitations).

For RCTs, risk of bias was assessed using the Cochrane Collaboration's tool for assessing risk of bias [12]. Methodological components of the studies were assessed and classified as high or low risk of bias. For studies that were not randomized trials but were comparative, study rigor was assessed using the Evidence Project 8-item checklist for intervention evaluations [13] and Risk of Bias in Non-randomized Studies of Interventions [14].

Data were analyzed according to the coding categories and outcomes. Where there were multiple studies reporting the same outcome for the same intervention-comparator comparison for the same population, we conducted meta-analysis using Comprehensive Meta-Analysis software (Biostat Inc). All outcomes were stratified and presented by key population. Findings were summarized in Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) evidence profile tables using GRADEPro, prioritizing RCT data over observational data where available.

Values and Preferences Review

The same search terms were used to search and screen for studies to be included in the values and preferences review. Studies were included in this review if they presented primary data examining the values and preferences of potential beneficiaries, communities, providers, and stakeholders for web-based service delivery interventions. These studies could be qualitative or quantitative in nature but had to present primary data collection; think pieces and review articles were not included. Values and preferences literature was summarized qualitatively and was organized by study design and methodology, location, and population.

Cost and Resource Needs

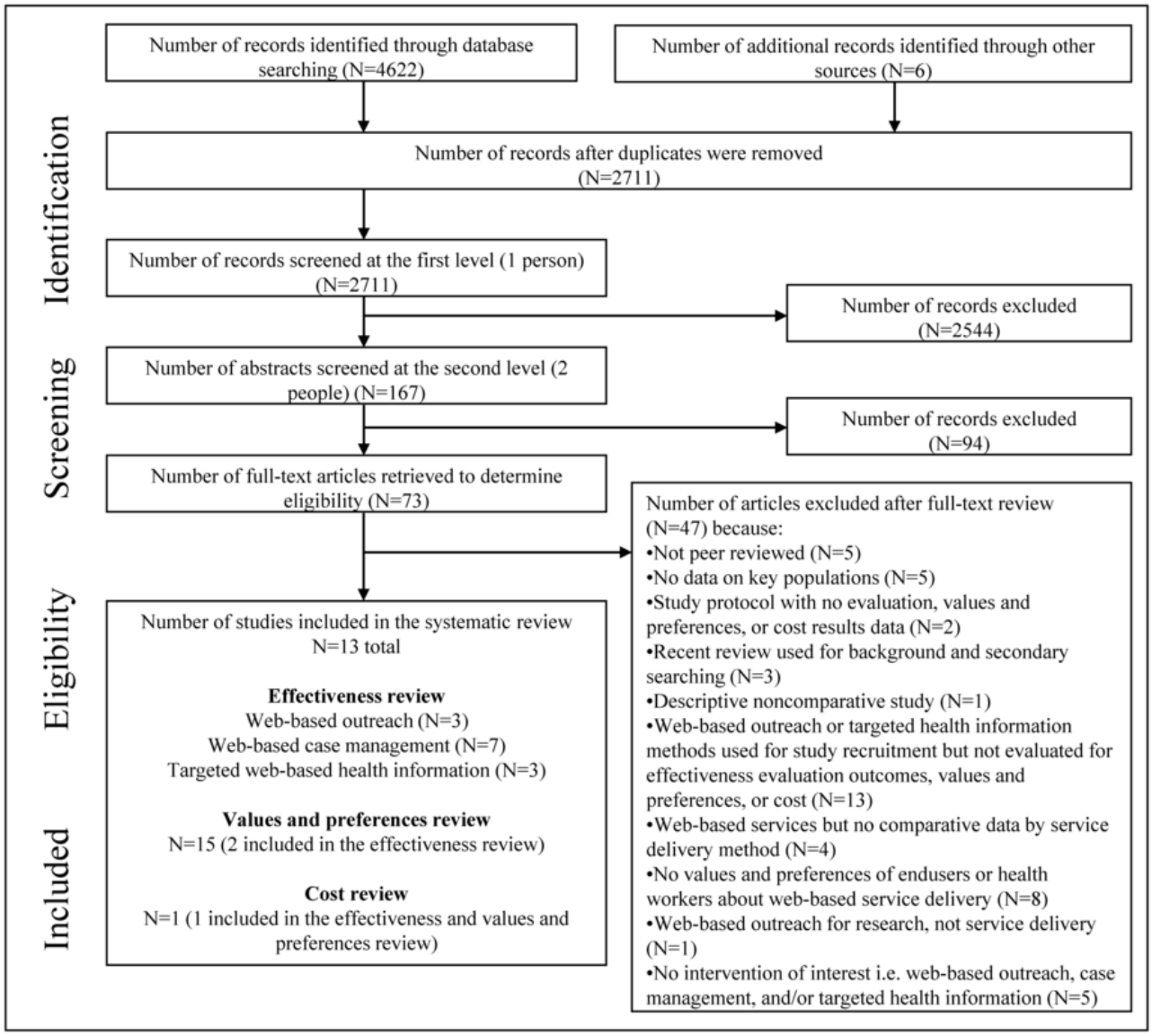
The same search terms were used to search and screen for studies to be included in the cost review. Studies were included in this review if they presented primary data comparing cost, cost-effectiveness, cost utility, or cost-benefit of the intervention and comparison listed in the PICO question or if they presented cost-effectiveness of the intervention as it relates to the PICO outcomes listed in Table 1. We qualitatively summarized the cost literature. We organized the cost literature into 4 categories (health sector costs, other sector costs, patient or family costs, and productivity impacts) and presented it by study design or methodology, location, and population within each category.

Results

Overview

Our database search yielded 4622 records, and we identified another 6 through hand searching and secondary searching (Figure 1). Of the 2711 unique records, 73 (2.69%) were retained for full-text review. Ultimately, we included 3 articles in the effectiveness review on web-based outreach [15-17], 7 articles in the effectiveness review on web-based case management [18-24], 3 articles in the effectiveness review on targeted web-based health information [25-27], 15 articles in the values and preferences review [24,27-40], and 1 article in the costs review [27].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart depicting the search and screening process.



Effectiveness Review

A total of 2 RCTs [16,17] and 1 serial cross-sectional study [15], all among men who have sex with men, were included in the effectiveness review of web-based outreach; 4 RCTs [20,21,23,24] and 3 observational studies [18,19,22] among men who have sex with men, trans and gender-diverse people, and people in prisons and other closed settings were included for web-based case management; and 2 RCTs [25,26] and 1 observational study [27], all among men who have sex with men, were included for targeted web-based health information. Descriptive data for the included articles are presented in Table 2, risk of bias assessments are presented in Multimedia Appendix 2 [15-27], and findings by PICO outcome within each intervention category in GRADE evidence profiles are presented in Multimedia Appendix 3 [15-27].

Table 2. Description of studies included in the effectiveness review.

Study; design	Population; sample size; location	Disease focus	Intervention	Comparison	Outcomes
Web-based outreach					
Tang et al [16], 2018; stepped-wedge cluster RCT[a]	MSM[b]; N=1381; China	HIV and STIs[c]	Integrated web-based HIV testing: an HIV testing intervention was developed through a national image contest, a regional strategy design-a-thon, and local message contests. The final intervention included a multimedia HIV testing campaign, a web-based HIV testing service, and local testing promotion campaigns tailored for MSM.	Conventional HIV testing programs routinely provided by local centers for disease control and community-based organizations	HIV testing and syphilis testing
Zhu et al [17], 2019; RCT	MSM; N=100; China	HIV and STIs	Web-based HIV self-testing via WeChat: intervention participants received 2 oral HIV self-testing kits and 6-month access to WeTest (a private WeChat group that provided app-based messages and referrals to HIV services, ie, brief informational messages on HIV, STIs, and HIV testing; first-person stories about people diagnosed and living with HIV; local data about HIV and STI among MSM; news about national policies related to HIV; stories about general health concerns of MSM; video or textual information about using the oral HIV self-testing kit; and 2-way communication between participant and WeTest team).	Brief video about self-administering the oral HIV self-testing kit (baseline procedures only)	HIV self-testing
Lampkin et al [15], 2016; serial cross-sectional	MSM; N=NR[d]; United States	HIV and STIs	Outreach via Grindr: suburban public health department in San Mateo County, California, used a social networking smartphone app designed for gay and bisexual men as a platform to engage MSM in STI outreach, education, and testing (screening and linkage to care).	Pre-Grindr outreach implementation	Total contacts with MSM by public health department, HIV primary care, and HIV and STI testing
Web-based case management					
Arayasirikul et al [18], 2020; prepost	MSM and TG[e] (HIV+, aged 18-34 years); N=120; United States	HIV	Six-month digital HIV care navigation intervention (connected to personal HIV care navigator via SMS text messaging) to improve engagement in HIV care.	Did not complete intervention	Received HIV primary care, currently taking ART[f], and viral suppression
Brantley et al [19], 2019; cohort	PRIS[g] (HIV+ adults soon to be released); N=238; United States	HIV	Web-based, tailored HIV and STI testing intervention, with baseline assessment and access to a tailored, personalized website.	Web-based provider directory page only	HIV and STI testing
Horvath et al [23], 2020; RCT	MSM (HIV−); N=113; United States	HIV	Mobile app "SUP[h]" for repeat HIV testing (monthly My Health Survey to recommend next test date; prevention "411" directory with HIV and STI information).	No treatment	HIV testing
Horvath et al [24], 2019; RCT	MSM (HIV+ stimulant users); N=90; United States	HIV	Mobile app "APP+" to improve ART adherence (IMB[i] HIV and ART content, choose-your-own-adventure story, and medication self-monitoring).	No treatment	ART adherence

Study; design	Population; sample size; location	Disease focus	Intervention	Comparison	Outcomes
Kuo et al [20], 2019; RCT	PRIS (HIV+ adults soon to be or recently released); N=110; United States	HIV	"CARE+Corrections" program to enhance HIV care engagement (computerized motivational interview or individual risk reduction plan) and SMS text messaging after release.	Opioid overdose prevention video and HIV providers or resources printout	Engagement in HIV care
Songtaweesin et al [21], 2020; RCT	MSM and TG (HIV−, aged 15-19 years); N=200; Thailand	HIV	Mobile app with youth - friendly services plus a mobile PrEP[j] app (self-assessment of HIV risk, rewards, and reminders for PrEP and clinic appointments).	Mobile app with youth-friendly services only	PrEP adherence
Young et al [22], 2014; cohort	PRIS (HIV+); N=1201; United States	HIV	Telemedicine (managed by a university-based multidisciplinary subspecialty team via a telemedicine clinic).	On-site management by correctional facility physicians	Viral suppression
Targeted web-based health information					
Bauermeister et al [26], 2015; RCT	MSM; N=130; United States	HIV and STIs (hepatitis A virus, hepatitis B virus, human papilloma virus, chlamydia, gonorrhea, and syphilis)	The tailored website included content customized based on prior testing experiences and motivations, barriers and resources to testing, and important values (gathered during baseline assessment). These personalized messages were included in web-based content, for example: STI facts; personal motivations, values, and strengths assessment regarding STI testing; exploration of barriers (eg, financial costs, social norms, and prioritization) to the participant's desire to get tested for HIV and STIs; and a listing of providers.	Access to the web-based provider directory (no personalized or tailored content)	Hepatitis A virus, hepatitis B virus, and human papilloma virus vaccination and HIV and STI testing
Young et al [25], 2013; RCT	MSM; N=112; United States	HIV	Project Harnessing Online Peer Education: social networking or peer leaders on Facebook to deliver information on HIV prevention or discuss HIV-related topics both individually and as a group via chat, wall posts, and personal messages over 12 weeks.	Peer leaders on Facebook to deliver information on general health	HIV testing

Study; design	Population; sample size; location	Disease focus	Intervention	Comparison	Outcomes
Ross et al [27], 2016; serial cross-sectional	MSM; N=NR; Canada	STIs (syphilis)	The Winnipeg Regional Health Authority developed a campaign in 2014 highlighting the syphilis outbreak and the importance of seeking testing. Over 1 month, advertisements appeared on 4 web-platforms: Grindr, Facebook, Squirt, and the Gay Ad Network. When clicked, the advertisements would direct the user to an information website.	Precampaign	Syphilis testing

[a]RCT: randomized controlled trial.

[b]MSM: men who have sex with men.

[c]STI: sexually transmitted infection.

[d]NR: not reported.

[e]TG: trans and gender-diverse people.

[f]ART: antiretroviral therapy.

[g]PRIS: people in prisons and other closed settings.

[h]SUP: Status Update Project.

[i]IMB: Information-Motivation-Behavioral.

[j]PrEP: preexposure prophylaxis.

Web-Based Outreach

One stepped-wedge RCT among 1381 men who have sex with men in China compared integrated web-based HIV testing (multimedia HIV testing campaign, web-based HIV testing service, and local testing promotion campaigns tailored for men who have sex with men) with conventional HIV testing routinely provided by local centers for disease control and community-based organizations [16]. Another RCT among 100 men who have sex with men in China compared web-based HIV self-testing via WeChat promotion or referrals with watching a brief video about self-administering the oral HIV self-testing kit [17]. A serial cross-sectional study among men who have sex with men in the United States compared using Grindr for STI outreach, education, and screening or linkage to care with standard of care [15].

Two studies reported the impact of web-based outreach on condom use. A stepped-wedge cluster RCT among men who have sex with men in China found little or no difference in self-reported condom use, comparing clusters who received the web-based HIV testing intervention with conventional HIV testing programs routinely provided by local centers of disease control and community-based organizations (relative risk [RR] 1.00, 95% CI 0.86-1.17); with little risk of bias, this RCT provided high-certainty evidence of no effect [16]. Another RCT of men who have sex with men in China provided low-certainty evidence that web-based outreach may make little or no difference to self-reported condom use, regardless of partner type (RR 0.90, 95% CI 0.39-2.06) [17].

The same 2 RCTs also reported on infection testing among men who have sex with men in China. A meta-analysis provided moderate-certainty evidence that the interventions increased HIV testing (RR 1.39, 95% CI 1.21-1.60) [16,17]. One RCT provided moderate-certainty evidence of probably no difference in syphilis testing (RR 0.92, 95% CI 0.70-1.21) [16].

One serial cross-sectional study among men who have sex with men in the United States provided moderate-certainty evidence that the intervention probably reached more previously unreached people [15]. When only traditional outreach methods were used (October 2011 to March 2012), the local public health department had contact with 60 men who have sex with men. After implementing outreach via Grindr (October 2013 to March 2014), the department had contact with 816 men who have sex with men. There was no denominator to calculate the rates and possible confounding from other factors, creating a potential risk of bias.

No studies provided comparative outcomes on other measures of the use of prevention services or on treatment initiation.

Web-Based Case Management

A total of 3 RCTs [20,23,24] and 3 comparative observational studies [18,19,22] examining web-based case management were conducted in the United States, and an additional RCT was conducted in Thailand [21]. Furthermore, 4 studies were conducted among men who have sex with men [18,21,23,24] (2 also included transgender women [18,21]), and 3 studies were conducted among people in prisons and other closed settings [19,20,22] (2 of which examined postrelease linkage or engagement in care [19,20]). All interventions were designed for web-based case management of primarily HIV services, whether in terms of engagement in preexposure prophylaxis services [21], HIV and STI testing [19,23], or linkage to care, antiretroviral therapy adherence, and HIV care engagement [18,20,22,24].

One RCT among men who have sex with men and transgender women in Thailand provided low-certainty evidence that web-based case management probably had no effect on the use

of prevention services. This study compared preexposure prophylaxis adherence among those who used a web-based case management app with those who did not (RR 1.12, 95% CI 0.78-1.59) [21].

One RCT among men who have sex with men in the United States showed no difference in the uptake of repeat HIV testing with web-based case management (RR 1.24, 95% CI 0.78-1.95) [23] but with very low certainty.

A cohort study among prisoners living with HIV in the United States found that there may be no difference in linkage to care after release from prison, comparing those who used a web-based tailored or personalized website with those who only had access to a web-based provider directory web page (RR 1.09, 95% CI 0.92-1.29; low-certainty evidence) [19]. Another cohort study among men who have sex with men and transgender women in the United States found a modest increase in the proportion of those who received primary HIV care in the last 6 months, comparing those who completed a 6-month digital HIV care navigation intervention with those who did not (RR 1.20, 95% CI 1.01-1.42; low-certainty evidence) [18].

One RCT in the United States among 90 men who have sex with men, who are living with HIV, and who use stimulants provided moderate-certainty evidence of probably higher overall antiretroviral therapy adherence in the intervention than in the control arm at 4 months (intervention: self-reported percentage ART adherence in the past 30 days 89% [95% CI 83.4-94.6] vs control: 77.2% [95% CI 66.7-87.7]; difference: 11.8% [95% CI 0.34-23.2]; *P*=.04), but 2 months later at 6 months, the improvements in adherence had probably dissipated (intervention: 85.3% [95% CI 80.0-90.6] vs control: 89.0% [95% CI 83.2-94.9]; difference: −3.7% [95% CI −11.4 to 4.0]; *P*=.34) [24].

One RCT among 110 people living with HIV who were recently incarcerated in the United States provided low-certainty evidence that web-based outreach may make little or no difference in engagement in HIV care, measured by having seen an HIV care provider in the community at least once in the past 24 weeks (RR 0.98, 95% CI 0.85-1.12) [20]. A cohort study among 120 men who have sex with men and transgender women living with HIV in the United States provided low-certainty evidence that web-based case management may make little to no difference in the proportion of people in each group self-reporting currently taking antiretroviral therapy (RR 1.19, 95% CI 0.97-1.45) [18].

The RCT among recently incarcerated persons living with HIV in the United States also provided low-certainty evidence that receiving a computerized motivational interview and individual risk reduction plan prerelease plus SMS text messaging about care navigation after release may make little or no difference to HIV viral suppression (laboratory-assessed viral load <200 copies/mL; RR 0.97, 95% CI 0.69-1.36) [20]. Similarly, a cohort study among men who have sex with men and transgender women living with HIV in the United States found that using web-based case management was associated with little or no difference in viral suppression (self-reported viral load <200 copies/mL; RR 1.05, 95% CI 0.79-1.40) [18], but another cohort study among 1201 incarcerated persons in the United States provided moderate-certainty evidence of probable improvement in the proportion of participants who had laboratory-assessed HIV virologic suppression at any of their first 6 care visits (RR 1.53, 95% CI 1.43-1.64) [22].

No studies measured our other outcomes of cure or mortality.

Targeted Web-Based Health Information

None of the included articles precisely fit our topic definition for "targeted web-based health information," but 3 articles were close enough to give a general idea of its effectiveness; the observational study may have been the closest to the intervention we desired to evaluate. This serial cross-sectional study among men who have sex with men in Winnipeg, Canada, compared syphilis testing rates before and after a social media campaign highlighting the syphilis outbreak and the importance of seeking testing, where advertisements on 4 web-based platforms (Grindr, Facebook, Squirt, and the Gay Ad Network) brought users to an information website connecting them with testing [27]. One RCT among 130 men who have sex with men in the United States compared participants who had access to a website customized to the user based on their personal prior testing experiences and motivations, barriers, and resources to testing with those who received access to a web-based provider directory with no tailored content [26]. Another RCT among 112 men who have sex with men in the United States compared having social networks or peer leaders on Facebook delivering HIV information in group settings and individually (via chat, wall posts, and personal messages) with peer leaders on Facebook delivering general health information [25].

One RCT among men who have sex with men in the United States provided low-certainty evidence that targeted web-based health information may make little or no difference in HIV testing (following up for HIV test result after requesting an HIV self-test kit and returning the kit; RR 3.56, 95% CI 0.32-39.65), although more participants in the intervention arm requested the kit than those in the control arm [25]. A serial cross-sectional study among men who have sex with men in Canada provided low-certainty evidence of little or no difference in the number of people who ordered a syphilis test before, during, or after a syphilis testing web-based advertisement campaign (RR 1.00, 95% CI 0.94-1.07) [27]. Another RCT among men who have sex with men in the United States found no impact of targeted web-based health information on HIV and STI testing (RR 1.46, 95% CI 0.72-2.94), although the sample size was small and the certainty of evidence was very low [26].

One RCT among men who have sex with men in the United States provided very low-certainty evidence on whether targeted web-based health information improves uptake of prevention services: no vaccinations for STIs were conducted in either arm, but sample sizes were very small (68 in the intervention arm and 36 in the control arm) [26].

No studies measured other outcomes of interest: treatment for HIV, VH, and STIs, treatment retention or completion for HIV, VH, and STIs, viral load testing or suppression, cure, or mortality.

Values and Preferences Review

Three cross-sectional studies examined the values and preferences around web-based outreach for key populations. These were conducted among men who have sex with men in Spain [36], men who have sex with men in China [40], and people who self-identified as a sexual or gender minority in East Africa [39]. Another 7 studies—all among men who have sex with men in the United States—examined values and preferences around web-based case management, sometimes in the context of assessing pilot programs [24,30-35]. Three studies in Canada, England, and the United States explored values and preferences around targeted web-based health information, all among men who have sex with men [27-29]. Specific data from each included article in the values and preferences review are presented in Multimedia Appendix 4 [24,27-40].

Regarding web-based outreach, in Spain, men who have sex with men thought it was acceptable to receive unsolicited messages about rapid HIV, syphilis, and hepatitis C testing on social media or *hookup* apps [36]. In Uganda, Tanzania, Rwanda, South Sudan, and Kenya, almost half of the survey respondents who self-identified as a sexual or gender minority were *very likely* to engage in a sexual health program if outreach was conducted on the web, over SMS text message, or over email [39]. In China, men who have sex with men reported high interest and willingness to use a "men who have sex with men-friendly physician finder function" within gay mobile apps [40].

Regarding web-based case management, men who have sex with men in China expressed interest in features and functions related to sexual health that could be embedded into existing smartphone apps or developed as stand-alone apps [35]. Men who have sex with men in the United States were strongly in favor of a smartphone app developed for web-based case management [31]. Positive features of such apps included their ease of use (eg, easy to navigate, fast, and convenient), the ability to set reminders or alarms to take medication at a certain time each day, trackers for adherence, and communication with providers, which helped users feel supported in their care process. However, several studies mentioned concerns about ensuring confidentiality in the web-based environment [24,30,31].

When asked about targeted web-based health information specifically, men who have sex with men expressed a diverse range of acceptability, ranging from indifference to frustration to enabling care seeking to influencing risk behaviors, and most were comfortable interacting with health services on the web, including through platforms that are not typically for health services, such as geosocial networking sites, for example, Facebook, or dating apps, for example, Grindr [27-29]. One study found that targeted web-based health information provided through *hookup sites* may garner more interest or be more acceptable than standard social media [27].

In 2 qualitative studies among frontline outreach workers, managers, or public health volunteers who worked with men who have sex with men in Canada [37,38], participants said that web-based technologies have reshaped the *gay or queer community*, have changed norms for social or sexual interactions, and can help reach out to hard-to-reach people. These studies found that web-based outreach generally allowed for more nonintrusive and anonymous communication (beneficial for clients), and that quick feedback helped them to be responsive to user needs. Respondents also noted some barriers to web-based outreach, such as quality of service, collaborations between outreach service agencies and companies that own apps and websites, budgetary and staff or volunteer capacity constraints, and data security and safety. On the basis of their experience, service providers described 4 ethical dilemmas as outreach moved to web-based platforms: (1) managing personal or professional boundaries with clients, (2) disclosing personal or identifiable information to clients, (3) maintaining client confidentiality and anonymity, and (4) security and data storage measures of web-based information.

Cost Review

Only 1 study was identified for the cost review [27]. This study among men who have sex with men in Canada found that syphilis testing campaign advertisements had the lowest cost-per-click ratio on the *hookup* platforms Grindr and Squirt, compared with more traditional social media platforms such as Facebook and the Gay Ad Network. No studies measured cost-effectiveness.

Discussion

Principal Findings

Overall, our effectiveness review found comparable outcomes when using web-based service delivery methods compared with standard of care, indicating that web-based approaches may be at least as effective as face-to-face services in terms of reaching new people, use of HIV and STI prevention services, and linkage to and retention in HIV care. However, the certainty of evidence for many outcomes in the effectiveness review was generally moderate or low, and for several PICO outcomes, either no statistically significant effect was reported or no studies measured the outcome of interest. These findings are broadly similar to those of other systematic reviews of digital health interventions with diverse health topics and populations [41-43]. When considering the values and preferences of end users and health workers, we generally found that web-based services for key populations were feasible and acceptable, similar to findings from other more general systematic reviews [44-46]. One cost study found that targeted messaging using web-based dating platforms may reach more end users than traditional web-based social media platforms. Programmatic descriptions from FHI 360's "Going online budgeting guide" designed to help accelerate the impact of HIV programs [47] show a wide range of costs for programs, depending on scope of work, including but not limited to country or regional costs, connectivity level, program intensity or scale, vendors, in-person trips or training needs, and equipment needs.

For some people who are unable to attend face-to-face services, web-based services may offer their only means of accessing information, support, referral, and case management. A study among youth found that web-based sexual health information is most valuable to youth who lack alternatives, and youth were more likely to take follow-up action if they had sought

information for reasons related to privacy or having *no one to ask*, especially among gender minority youth [48]. However, some people have limited access to the internet, low reading ability, or low digital literacy; therefore, web-based services may not meet the needs of the most vulnerable. Potential harms relating to data security and confidentiality are of particular concern for key populations who may engage in criminalized and stigmatized activities and experience discrimination, arrest, or harassment if confidentiality is breached. These could also be issues for people who share devices, younger key populations, or groups such as sex workers who may change devices frequently. The ethics surrounding data mining are of special concern when targeting health information using social media platforms. However, capacity building of outreach staff and health workers regarding data security and confidentiality could mitigate this challenge. Other limitations to web-based services may include loss of in-person rapport and relationships when counseling and financial costs for end users (such as internet or airtime or for-profit apps).

Web-based services should be an additional, complementary approach within larger HIV, STI, and VH programs for key populations. Web-based services could be an important additional way to reach more people, reduce costs, reduce waiting times (by reducing clinic attendance), and allow more time for more complex case management in health facilities. Systematic planning on how web-based services can complement other services and target audiences should be conducted, with key populations playing a central role in design, implementation, and monitoring. Content development should be informed by appropriate and accurate health content and information aligned with recommendation practices (eg, from health program guidelines or evidence-based normative practices) and country policies for platform and language use [49]. In addition, maintenance of safety and security measures of web-based services will be required [50].

This review had several strengths. We conducted a comprehensive search across multiple databases as well as a hand search and secondary search. We assessed the methodological quality of the included articles and examined not only the effectiveness of web-based services for key populations but also end users' and health workers' values and preferences and costing. However, the available published research on web-based services for key populations was conducted mainly among men who have sex with men in high-income countries, and web-based service delivery options were usually limited to HIV prevention and treatment services. Among the included studies, we were only able to meta-analyze 1 outcome that provided comparable measurement methods for similar populations. Further research is needed on web-based services for other key populations in low- and middle-income settings and their effect on outcomes related to STI and VH. More research is needed on the cost-effectiveness of different types of web-based services for key population groups. Although many government- and community-based programs are already using web-based services for key populations, suggesting that it is a feasible intervention in many settings [51], these could not be included in our effectiveness review because no explicit comparison has been made between their web-based service delivery and in-person modalities. Finally, we excluded SMS text messaging interventions and other digital health interventions that were not primarily conducted on the web. Although we did this purposefully to narrow this review's focus, we are not able to comment on the effectiveness, values and preferences, or costs related to these other approaches.

Using the evidence from this review and discussion among the experts in the WHO Key Population Guideline Development Group, the new consolidated WHO guideline for key populations includes a new conditional recommendation based on low-certainty evidence: "Online delivery of HIV, viral hepatitis, and STI services to key populations may be offered as an additional option while ensuring that data security and confidentiality are protected" [11].

Following this review's search cutoff date and the WHO guideline development group meeting that led to the above recommendation, several study protocols [52,53] have been published for forthcoming RCTs, suggesting a growing evidence base for this topic.

Conclusions

Web-based services for HIV, VH, and STIs are becoming increasingly common, especially in high-income settings, and are generally accepted by end users and health workers, suggesting their feasibility as an additional option for service delivery for key populations. Although largely acceptable, there will be people who do not wish to access web-based services and prefer face-to-face communication with health workers. This review of the extant peer-reviewed literature suggests that web-based service delivery may be a feasible and acceptable approach to expanding services to key populations with similar outcomes as standard of care; however, more research is needed on how to improve the effectiveness of these web-based service delivery interventions. Further research is also needed in low-resource settings, among key populations who are not men who have sex with men, and with infections other than HIV (ie, VH and STIs).

Acknowledgments

The authors would like to thank Johns Hopkins graduate students Xuhao Yang and Jingjia Cynthia Li and Lake Erie College of Osteopathic Medicine medical student Shawn Kaura for their assistance in reference searching, reference screening, and data abstraction. The authors would also like to thank Joyce Yehjin Jang and Cortney McLellan for their help in submitting this manuscript, Cadi Irvine for her comments on the systematic review protocol and manuscript, and Lisa Maher for providing additional study information. The World Health Organization receives funding from The Bill and Melinda Gates Foundation and the US President's Emergency Plan for AIDS Relief and United States Agency for International Development for the work on

key populations. The funding sponsors played no role in the design and conduct of the study or in the review and approval of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for the systematic review of web-based service delivery for key populations.
[PDF File (Adobe PDF File), 123 KB - jmir_v24i12e40150_app1.pdf]

Multimedia Appendix 2
Risk of bias assessments for articles included in the effectiveness review.
[PDF File (Adobe PDF File), 157 KB - jmir_v24i12e40150_app2.pdf]

Multimedia Appendix 3
Grading of Recommendations, Assessment, Development, and Evaluations tables presenting the summary of evidence used for the effectiveness review.
[PDF File (Adobe PDF File), 275 KB - jmir_v24i12e40150_app3.pdf]

Multimedia Appendix 4
Description of and key findings from the studies included in the values and preferences review.
[PDF File (Adobe PDF File), 231 KB - jmir_v24i12e40150_app4.pdf]

References

1. Cheng TC, Fu H, Xu D, Yip W. Technology platforms are revolutionizing health care service delivery in China. NEJM Catalyst 2022 Feb 23. [doi: 10.1056/CAT.21.0414]
2. Mobasheri MH, King D, Johnston M, Gautama S, Purkayastha S, Darzi A. The ownership and clinical use of smartphones by doctors and nurses in the UK: a multicentre survey study. BMJ Innov 2015 Oct 07;1(4):174-181. [doi: 10.1136/bmjinnov-2015-000062]
3. Ngini CU, Furnell SM, Ghita BV. Accessing the global accessibility of the Internet. Internet Res 2002;12(4):329-338. [doi: 10.1108/10662240210438399]
4. Gong W, Li ZG, Stump RL. Global internet use and access: cultural considerations. Asia Pac J Market Log 2007 Jan 16;19(1):57-74. [doi: 10.1108/13555850710720902]
5. Baraitser P, Syred J, Spencer-Hughes V, Howroyd C, Free C, Holdsworth G. How online sexual health services could work; generating theory to support development. BMC Health Serv Res 2015 Dec 05;15:540 [FREE Full text] [doi: 10.1186/s12913-015-1200-x] [Medline: 26637295]
6. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. Geneva, Switzerland: World Health Organization; 2018. URL: https://apps.who.int/iris/handle/10665/260480 [accessed 2022-10-31]
7. Meiksin R, Melendez-Torres GJ, Falconer J, Witzel TC, Weatherburn P, Bonell C. eHealth interventions to address sexual health, substance use, and mental health among men who have sex with men: systematic review and synthesis of process evaluations. J Med Internet Res 2021 Apr 23;23(4):e22477 [FREE Full text] [doi: 10.2196/22477] [Medline: 33890855]
8. Meiksin R, Melendez-Torres GJ, Falconer J, Witzel TC, Weatherburn P, Bonell C. Theories of change for e-health interventions targeting HIV/STIs and sexual risk, substance use and mental ill health amongst men who have sex with men: systematic review and synthesis. Syst Rev 2021 Jan 11;10(1):21 [FREE Full text] [doi: 10.1186/s13643-020-01523-2] [Medline: 33423693]
9. WHO handbook for guideline development. 2nd Edition. World Health Organization. Geneva, Switzerland: World Health Organization; 2014. URL: https://www.who.int/publications/i/item/9789241548960 [accessed 2022-10-31]
10. 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. PLoS Med 2021 Mar;18(3):e1003583 [FREE Full text] [doi: 10.1371/journal.pmed.1003583] [Medline: 33780438]
11. Consolidated guidelines on HIV, viral hepatitis and STI prevention, diagnosis, treatment and care for key populations. World Health Organization. Geneva, Switzerland: World Health Organization; 2022 Jul 29. URL: https://www.who.int/publications/i/item/9789240052390 [accessed 2022-10-31]
12. Sterne JA, 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 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]

13. Kennedy CE, Fonner VA, Armstrong KA, Denison JA, Yeh PT, O'Reilly KR, et al. The Evidence Project risk of bias tool: assessing study rigor for both randomized and non-randomized intervention studies. Syst Rev 2019 Jan 03;8(1):3 [FREE Full text] [doi: 10.1186/s13643-018-0925-0] [Medline: 30606262]
14. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
15. Lampkin D, Crawley A, Lopez TP, Mejia CM, Yuen W, Levy V. Reaching suburban men who have sex with men for STD and HIV services through online social networking outreach: a public health approach. J Acquir Immune Defic Syndr 2016 May 01;72(1):73-78. [doi: 10.1097/QAI.0000000000000930] [Medline: 27097365]
16. Tang W, Wei C, Cao B, Wu D, Li KT, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: a closed cohort stepped wedge cluster randomized controlled trial. PLoS Med 2018 Aug;15(8):e1002645 [FREE Full text] [doi: 10.1371/journal.pmed.1002645] [Medline: 30153265]
17. Zhu X, Zhang W, Operario D, Zhao Y, Shi A, Zhang Z, et al. Effects of a mobile health intervention to promote HIV self-testing with MSM in China: a randomized controlled trial. AIDS Behav 2019 Nov;23(11):3129-3139 [FREE Full text] [doi: 10.1007/s10461-019-02452-5] [Medline: 30852728]
18. Arayasirikul S, Turner C, Trujillo D, Le V, Wilson EC. Efficacy and impact of digital HIV care navigation in young people living with HIV in San Francisco, California: prospective study. JMIR Mhealth Uhealth 2020 May 08;8(5):e18597 [FREE Full text] [doi: 10.2196/18597] [Medline: 32383680]
19. Brantley AD, Page KM, Zack B, Friedrich KR, Wendell D, Robinson WT, et al. Making the connection: using videoconferencing to increase linkage to care for incarcerated persons living with HIV post-release. AIDS Behav 2019 Jan;23(Suppl 1):32-40. [doi: 10.1007/s10461-018-2115-4] [Medline: 29680934]
20. Kuo I, Liu T, Patrick R, Trezza C, Bazerman L, Uhrig Castonguay BJ, et al. Use of an mHealth intervention to improve engagement in HIV community-based care among persons recently released from a correctional facility in Washington, DC: a pilot study. AIDS Behav 2019 Apr;23(4):1016-1031 [FREE Full text] [doi: 10.1007/s10461-018-02389-1] [Medline: 30627850]
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. J Int AIDS Soc 2020 Sep;23 Suppl 5(Suppl 5):e25564 [FREE Full text] [doi: 10.1002/jia2.25564] [Medline: 32869511]
22. Young JD, Patel M, Badowski M, Mackesy-Amiti ME, Vaughn P, Shicker L, et al. Improved virologic suppression with HIV subspecialty care in a large prison system using telemedicine: an observational study with historical controls. Clin Infect Dis 2014 Jul 01;59(1):123-126 [FREE Full text] [doi: 10.1093/cid/ciu222] [Medline: 24723283]
23. Horvath KJ, Lammert S, Danh T, Mitchell JW. The feasibility, acceptability and preliminary impact of mobile application to increase repeat HIV testing among sexual minority men. AIDS Behav 2020 Jun;24(6):1835-1850. [doi: 10.1007/s10461-019-02755-7] [Medline: 31823111]
24. Horvath KJ, Lammert S, MacLehose RF, Danh T, Baker JV, Carrico AW. A pilot study of a mobile app to support HIV antiretroviral therapy adherence among men who have sex with men who use stimulants. AIDS Behav 2019 Nov;23(11):3184-3198 [FREE Full text] [doi: 10.1007/s10461-019-02597-3] [Medline: 31309348]
25. Young SD, Cumberland WG, Lee SJ, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. Ann Intern Med 2013 Sep 03;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]
26. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, Harper GW, Horvath K, Weiss G, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874 [FREE Full text] [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
27. Ross C, Shaw S, Marshall S, Stephen S, Bailey K, Cole R, et al. Impact of a social media campaign targeting men who have sex with men during an outbreak of syphilis in Winnipeg, Canada. Can Commun Dis Rep 2016 Feb 04;42(2):45-49 [FREE Full text] [doi: 10.14745/ccdr.v42i02a04] [Medline: 29770003]
28. Contesse MG, Fredericksen RJ, Wohlfeiler D, Hecht J, Kachur R, Strona FV, et al. Acceptability of using geosocial networking applications for HIV/sexually transmitted disease partner notification and sexual health services. Sex Transm Dis 2020 Jan;47(1):41-47. [doi: 10.1097/OLQ.0000000000001089] [Medline: 31688728]
29. Kesten JM, Dias K, Burns F, Crook P, Howarth A, Mercer CH, et al. Acceptability and potential impact of delivering sexual health promotion information through social media and dating apps to MSM in England: a qualitative study. BMC Public Health 2019 Sep 06;19(1):1236 [FREE Full text] [doi: 10.1186/s12889-019-7558-7] [Medline: 31492120]
30. Balán IC, Lopez-Rios J, Nayak S, Lentz C, Arumugam S, Kutner B, et al. SMARTtest: a smartphone app to facilitate HIV and syphilis self- and partner-testing, interpretation of results, and linkage to care. AIDS Behav 2020 May;24(5):1560-1573 [FREE Full text] [doi: 10.1007/s10461-019-02718-y] [Medline: 31679075]
31. Fontenot HB, Rosenberger JG, McNair KT, Mayer KH, Zimet G. Perspectives and preferences for a mobile health tool designed to facilitate HPV vaccination among young men who have sex with men. Hum Vaccin Immunother 2019;15(7-8):1815-1823 [FREE Full text] [doi: 10.1080/21645515.2019.1568156] [Medline: 30625049]

32. Fuchs JD, Stojanovski K, Vittinghoff E, McMahan VM, Hosek SG, Amico KR, et al. A mobile health strategy to support adherence to antiretroviral preexposure prophylaxis. AIDS Patient Care STDS 2018 Mar;32(3):104-111 [FREE Full text] [doi: 10.1089/apc.2017.0255] [Medline: 29565183]
33. 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 Behav 2021 Apr;25(4):1001-1012 [FREE Full text] [doi: 10.1007/s10461-020-03054-2] [Medline: 33044687]
34. Refugio ON, Kimble MM, Silva CL, Lykens JE, Bannister C, Klausner JD. Brief report: PrEPTECH: a telehealth-based initiation program for HIV pre-exposure prophylaxis in young men of color who have sex with men. A pilot study of feasibility. J Acquir Immune Defic Syndr 2019 Jan 01;80(1):40-45 [FREE Full text] [doi: 10.1097/QAI.0000000000001873] [Medline: 30272632]
35. Ventuneac A, John SA, Whitfield TH, Mustanski B, Parsons JT. Preferences for sexual health smartphone app features among gay and bisexual men. AIDS Behav 2018 Oct;22(10):3384-3394 [FREE Full text] [doi: 10.1007/s10461-018-2171-9] [Medline: 29948335]
36. Alarcón Gutiérrez M, Fernández Quevedo M, Martín Valle S, Jacques-Aviñó C, Díez David E, Caylà JA, et al. Acceptability and effectiveness of using mobile applications to promote HIV and other STI testing among men who have sex with men in Barcelona, Spain. Sex Transm Infect 2018 Sep;94(6):443-448. [doi: 10.1136/sextrans-2017-053348] [Medline: 29626174]
37. Brennan DJ, Souleymanov R, Lachowsky N, Betancourt G, Pugh D, McEwen O. Providing online-based sexual health outreach to gay, bisexual, and queer men in Ontario, Canada: qualitative interviews with multisectoral frontline service providers and managers. AIDS Patient Care STDS 2018 Jul;32(7):282-287. [doi: 10.1089/apc.2018.0027] [Medline: 29897787]
38. Fantus S, Souleymanov R, Lachowsky NJ, Brennan DJ. The emergence of ethical issues in the provision of online sexual health outreach for gay, bisexual, two-spirit and other men who have sex with men: perspectives of online outreach workers. BMC Med Ethics 2017 Nov 03;18(1):59 [FREE Full text] [doi: 10.1186/s12910-017-0216-7] [Medline: 29100520]
39. Ybarra M, Nyemara N, Mugisha F, Garofalo R. Opportunities to harness technology to deliver HIV prevention / healthy sexuality programming to sexual and gender minority adults living in East Africa. AIDS Behav 2021 Apr;25(4):1120-1128. [doi: 10.1007/s10461-020-03059-x] [Medline: 33123824]
40. Cao B, Zhao P, Bien C, Pan S, Tang W, Watson J, et al. Linking young men who have sex with men (YMSM) to STI physicians: a nationwide cross-sectional survey in China. BMC Infect Dis 2018 May 18;18(1):228 [FREE Full text] [doi: 10.1186/s12879-018-3145-2] [Medline: 29776395]
41. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: a systematic review. Digit Health 2018 May 10;4:2055207618770861 [FREE Full text] [doi: 10.1177/2055207618770861] [Medline: 29942631]
42. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MD, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
43. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
44. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
45. Cox A, Lucas G, Marcu A, Piano M, Grosvenor W, Mold F, et al. Cancer survivors' experience with telehealth: a systematic review and thematic synthesis. J Med Internet Res 2017 Jan 09;19(1):e11 [FREE Full text] [doi: 10.2196/jmir.6575] [Medline: 28069561]
46. Brunton L, Bower P, Sanders C. The contradictions of telehealth user experience in chronic obstructive pulmonary disease (COPD): a qualitative meta-synthesis. PLoS One 2015 Oct 14;10(10):e0139561 [FREE Full text] [doi: 10.1371/journal.pone.0139561] [Medline: 26465333]
47. Eveslage B. Virtual HIV interventions: a budgeting and programming aid, version 2. FHI360. Durham, NC, USA: FHI360; 2022 Feb. URL: https://www.fhi360.org/sites/default/files/media/documents/resource-going-online-budgeting-guide.pdf [accessed 2022-10-31]
48. Mitchell KJ, Ybarra ML, Korchmaros JD, Kosciw JG. Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. Health Educ Res 2014 Feb;29(1):147-157 [FREE Full text] [doi: 10.1093/her/cyt071] [Medline: 23861481]
49. Recommendations on digital interventions for health system strengthening. World Health Organization. Geneva, Switzerland: World Health Organization; 2019 Jun 6. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2022-10-31]
50. Going online to accelerate the impact of HIV programs. FHI360. Washington, DC, USA: FHI360; 2019. URL: https://www.fhi360.org/sites/default/files/media/documents/resource-linkages-vision-going-online.pdf [accessed 2022-10-31]

51. Global strategy on digital health 2020-2025. World Health Organization. Geneva, Switzerland: World Health Organization; 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2022-10-31]
52. Sha Y, Li C, Xiong Y, Hazra A, Lio J, Jiang I, et al. Co-creation using crowdsourcing to promote PrEP adherence in China: study protocol for a stepped-wedge randomized controlled trial. BMC Public Health 2022 Sep 07;22(1):1697 [FREE Full text] [doi: 10.1186/s12889-022-14117-5] [Medline: 36071401]
53. Yan X, Li Y, Su H, Xing Y, Zhang B, Lu Z, et al. Protect MSM from HIV and other sexually transmitted diseases by providing mobile health services of partner notification: protocol for a pragmatic stepped wedge cluster randomized controlled trial. BMC Public Health 2020 Jul 14;20(1):1107 [FREE Full text] [doi: 10.1186/s12889-020-09162-x] [Medline: 32664934]

Abbreviations

GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
PICO: population, intervention, comparison, and outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: relative risk
STI: sexually transmitted infection
VH: viral hepatitis
WHO: World Health Organization

Edited by G Eysenbach; submitted 08.06.22; peer-reviewed by A Yeung, J Stekler; comments to author 19.09.22; revised version received 30.09.22; accepted 20.10.22; published 22.12.22.

Please cite as:
Yeh PT, Kennedy CE, Minamitani A, Baggaley R, Shah P, Verster A, Luhmann N, de Mello MB, Macdonald V
Web-Based Service Provision of HIV, Viral Hepatitis, and Sexually Transmitted Infection Prevention, Testing, Linkage, and Treatment for Key Populations: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(12):e40150
URL: https://www.jmir.org/2022/12/e40150
doi:10.2196/40150
PMID:36548036

Review

Patient-Centered Digital Health Records and Their Effects on Health Outcomes: Systematic Review

Martijn R Brands[1], MD; Samantha C Gouw[1], MD, PhD; Molly Beestrum[2], MLiS; Robert M Cronin[3], MS, MD; Karin Fijnvandraat[1,4], MD, PhD; Sherif M Badawy[5,6], MBBCh, MS, MD

[1]Department of Pediatric Hematology, Emma Children's Hospital, Amsterdam Reproduction & Development, Public Health, Amsterdam UMC location University of Amsterdam, Amsterdam, Netherlands

[2]Galter Health Sciences Library at Northwestern University, Chicago, IL, United States

[3]Department of Medicine, The Ohio State University, Columbus, OH, United States

[4]Department of Molecular Cellular Hemostasis, Sanquin Research and Landsteiner Laboratory, Amsterdam, Netherlands

[5]Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[6]Division of Hematology, Oncology, and Stem Cell Transplant, Ann and Robert H Lurie Children's Hospital of Chicago, Chicago, IL, United States

Corresponding Author:
Sherif M Badawy, MBBCh, MS, MD
Department of Pediatrics
Feinberg School of Medicine, Northwestern University
420 E Superior St
Chicago, IL, 60611
United States
Phone: 1 312 227 4789
Email: sherif.badawy@northwestern.edu

Abstract

Background: eHealth tools such as patient portals and personal health records, also known as patient-centered digital health records, can engage and empower individuals with chronic health conditions. Patients who are highly engaged in their care have improved disease knowledge, self-management skills, and clinical outcomes.

Objective: We aimed to systematically review the effects of patient-centered digital health records on clinical and patient-reported outcomes, health care utilization, and satisfaction among patients with chronic conditions and to assess the feasibility and acceptability of their use.

Methods: We searched MEDLINE, Cochrane, CINAHL, Embase, and PsycINFO databases between January 2000 and December 2021. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed. Eligible studies were those evaluating digital health records intended for nonhospitalized adult or pediatric patients with a chronic condition. Patients with a high disease burden were a subgroup of interest. Primary outcomes included clinical and patient-reported health outcomes and health care utilization. Secondary outcomes included satisfaction, feasibility, and acceptability. Joanna Briggs Institute critical appraisal tools were used for quality assessment. Two reviewers screened titles, abstracts, and full texts. Associations between health record use and outcomes were categorized as *beneficial*, *neutral or clinically nonrelevant*, or *undesired*.

Results: Of the 7716 unique publications examined, 81 (1%) met the eligibility criteria, with a total of 1,639,556 participants across all studies. The most commonly studied diseases included diabetes mellitus (37/81, 46%), cardiopulmonary conditions (21/81, 26%), and hematology-oncology conditions (14/81, 17%). One-third (24/81, 30%) of the studies were randomized controlled trials. Of the 81 studies that met the eligibility criteria, 16 (20%) were of high methodological quality. Reported outcomes varied across studies. The benefits of patient-centered digital health records were most frequently reported in the category health care utilization on the "use of recommended care services" (10/13, 77%), on the patient-reported outcomes "disease knowledge" (7/10, 70%), "patient engagement" (13/28, 56%), "treatment adherence" (10/18, 56%), and "self-management and self-efficacy" (10/19, 53%), and on the clinical outcome "laboratory parameters," including HbA_{1c} and low-density lipoprotein (LDL; 16/33, 48%). Beneficial effects on "health-related quality of life" were seen in only 27% (4/15) of studies. Patient satisfaction (28/30, 93%), feasibility (15/19, 97%), and acceptability (23/26, 88%) were positively evaluated. More beneficial effects were reported for digital health records that predominantly focus on active features. Beneficial effects were less frequently observed among patients with a high disease burden and among high-quality studies. No unfavorable effects were observed.

Conclusions: The use of patient-centered digital health records in nonhospitalized individuals with chronic health conditions is potentially associated with considerable beneficial effects on health care utilization, treatment adherence, and self-management or self-efficacy. However, for firm conclusions, more studies of high methodological quality are required.

Trial Registration: PROSPERO (International Prospective Register of Systematic Reviews) CRD42020213285; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=213285

(J Med Internet Res 2022;24(12):e43086) doi:10.2196/43086

KEYWORDS

telemedicine; health records; personal; electronic health records; outcome assessment; health care

Introduction

Background

The prevalence and disease burden of chronic health conditions is on the rise. The World Health Organization predicts that by 2030, chronic noncommunicable health conditions will account for >50% of the total disease burden [1,2]. In particular, cardiovascular conditions, cancer, respiratory conditions, and diabetes have the highest morbidity and mortality [1]. Currently, 60% of the US population has at least 1 chronic condition and 42% of the population has multiple chronic conditions [3]. This results in a high individual disease burden owing to the large impact on social participation and required patient self-management skills. Self-management refers to a person's ability to manage the clinical, psychosocial, and societal aspects of their illness and its care [4]. In contrast, self-efficacy is a person's belief that he or she can successfully execute this behavior [4]. Apart from a high individual disease burden, the prevalence of chronic conditions imposes a high macroeconomic burden [5]. Furthermore, an increasing shortage of health care providers is expected, among others in the United States [6] and Europe [7,8]. In combination with the increased pressure put on health systems by unexpected events such as the COVID-19 pandemic, this shortage threatens the delivery of essential health services [9]. To preserve the access to care for all patients, new technologies are increasingly being developed and adopted, including patient-centered digital health records.

Such patient-centered digital health records can significantly help engage and empower patients with a chronic health condition [10-13]. Patient-centered digital health records enable patients to take on a more active role in their care by allowing them to view parts of their medical records, such as medication lists, laboratory and imaging results, allergies, and correspondence. Other common features include secure messaging, requesting prescription refills, video consultation, paying bills, and managing appointments. Examples of patient-centered digital health records include patient portals and personal health records (PHRs). Patient-centered digital health records differ in the volume and detail of the provided medical data, functionalities, and level of patient control, as shown in Textbox 1. Highly engaged patients are reported to have increased disease knowledge, better self-management, more self-efficacy, and improved clinical outcomes [14-16]. The effects of using patient-centered digital health records may be most substantial for patients with chronic conditions. Many self-management skills are required, and their potential gains are the highest. Not only patients but the entire health care system might benefit from an increased adoption of patient-centered digital health records.

Textbox 1. Proposed taxonomy of patient-centered digital health records [10,17-21].

- Electronic health record (EHR): a digital version of a health care provider's paper chart, used by health care professionals alone. Patients cannot access data in an EHR. An EHR might contain data from one health care institution or from multiple institutions. Its scope can range from regional, to national, or international.
- Patient portal: the patient-facing interface of an EHR that enables people to view sections of their medical record. This might include access to test results, medication lists, or therapeutic instructions. Health care providers or health care offices determine what health information is accessible for patients. Patient portals often have additional features such as patient-professional messaging, requesting prescription refills, scheduling appointments, or communicating patient-reported outcomes. By definition, patient portals are "tethered," in which "tethered" refers to a patient portal's connection to an EHR. Occasionally, a patient portal is referred to as a tethered personal health record (PHR).
- PHR: a PHR is similar to a patient portal and can have similar features. However, the main difference is that contents are managed and maintained by individuals, not health care providers. People can access, manage, and share their health information, and that of others for whom they are authorized, such as parents or caretakers. Health information from different health care institutions may reside in a single patient-managed PHR. In general, PHRs are not tethered unless otherwise specified. Few tethered PHRs currently exist but are increasingly being developed [22].
- Patient-centered digital health records: an umbrella term referring to patient portals, tethered PHRs, and part of the untethered PHRs. Patient-centered digital health records enable a 2-way exchange of health information between patients and the health care system and provide patients with the ability to view, download, or transmit their health information on the web. This health information is updated at regular intervals. In addition, it enables communication between patients and the health care system, either by adding or editing health information, exchanging patient-reported outcomes, or by using communication tools such as messaging. Additional functionalities are often present.
- "Electronic medical record" is an outdated term [21]. It can be considered a professional-centered EHR with limited functionalities.

Currently, huge investments of time and resources are made in patient-centered digital health records. However, limited insight exists in how the use of patient-centered digital health records by patients with a broad range of chronic conditions affects clinical and patient-reported outcomes and health care utilization. Moreover, we lack an overview of their effects on patient satisfaction, and the feasibility and acceptability of their use by people with chronic conditions. Previous systematic reviews focused on one health condition [23], focused on one type of digital health record [24-27], investigated a select set of health outcomes [24,26,28], or are now obsolete in this rapidly changing technological landscape [23,25,27].

Objectives

Therefore, in this systematic review, we summarized the available evidence on patient-centered digital health records. Our primary objective was to assess how patient-centered digital health records for nonhospitalized patients with chronic conditions affect clinical and patient-reported health outcomes and health care utilization. Our secondary objective was to evaluate patient satisfaction with and feasibility and acceptability of using patient-centered digital health records. Results of this systematic review may help guide future development and implementation.

Methods

The protocol for this study was registered in the International PROSPERO (International Prospective Register of Systematic Reviews) Register of Systematic Reviews (CRD42020213285) [29]. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed [30].

Literature Search

A medical librarian (MB) conducted the original literature search using the following databases: MEDLINE, Cochrane Library, CINAHL, Embase, and PsycINFO. All original studies published between January 1, 2000, and December 1, 2020, were assessed. A search update in MEDLINE was performed for all studies published between December 1, 2020, and December 31, 2021. Multimedia Appendix 1 presents the full search strategy. Articles published before 2000 were excluded because of the rapidly changing field of digital health technology [30].

Eligibility Criteria

Patient-centered digital health records were defined as mobile health (mHealth) or eHealth technologies that enable a 2-way exchange of health information between patients and the health care system, such as patient portals, PHRs, or mHealth apps with a health record functionality. A patient-centered digital health record provides patients with the ability to view, download, or transmit their health information on the web. This health information was updated at regular intervals. In addition, a patient-centered digital health record allows for communication between patients and the health care system, either by adding or editing health information, exchanging patient-reported outcomes, or by using communication tools such as messaging. Several other functionalities are common, but were not considered essential; for example, appointment scheduling, requesting prescription refill, viewing educational material, using decision support tools, and using connected wearables. Exclusion criteria were nondigital health records, digital health records intended for hospitalized patients, and digital health records that are not accessible to patients, such as the clinician-facing components of the electronic health record (EHR).

Studies

Studies investigating patient-centered digital health records intended for nonhospitalized patients with a chronic health condition were included. Only studies published in English were included. Eligible studies included randomized controlled trials (RCTs), quasi-experimental studies, nonexperimental observational studies (including cohort and cross-sectional studies), and pilot or feasibility studies. Of mixed methods studies, only nonqualitative parts were used for data extraction. Studies that only described health care providers' experiences were excluded.

Participants

Studies on patients with a chronic health condition of all age groups were considered. Chronic conditions included all diseases with a moderate to high disease burden and moderate to high impact on daily life. Consequently, these conditions demand considerable self-management skills from patients to manage the clinical, psychosocial, and societal aspects of chronic condition and its care. The selection of chronic conditions included in our search strategy was based on the Charlson Comorbidity Index, other literature, and clinical expertise [31,32]. Diseases included cancer, arthritis, HIV, AIDS, asthma, chronic obstructive pulmonary disease, chronic heart conditions, hematologic disease, chronic kidney disease, celiac disease, inflammatory bowel disease, cystic fibrosis, diabetes mellitus, and multiple sclerosis (MS).

Outcomes

Studies were required to report at least one primary or secondary outcome. Primary outcomes were clinical outcomes (including disease events and complications, vital parameters, and laboratory parameters), patient-reported outcomes (including self-management and self-efficacy, patient engagement, health-related quality of life (HRQoL), stress and anxiety, and treatment adherence), and health care utilization (including the number of emergency department [ED] visits and hospitalizations, the use of preventive or recommended care services by patients, and regular workload for health care professionals). Secondary outcomes included technology-related outcomes (including patient satisfaction, feasibility, and acceptability). Definitions and examples of these 13 outcomes are presented in Table 1.

Table 1. Definitions and examples of all health outcomes included in this systematic review.

Included study outcomes	Definitions and examples
Clinical outcomes	
Disease events and complications	• For example, asthma exacerbation, chronic kidney disease progression, and death
Vital parameters	• For example, blood pressure, BMI, weight, and respiratory parameters
Laboratory parameters	• For example, HbA_{1c}[a], LDL[b], cholesterol, eGFR[c], HIV viral load, and CD4+ T-cell count
Patient-reported outcomes	
Self-management and self-efficacy	• Self-management is a person's ability to manage the clinical, psychosocial, and societal aspects of illness and its care. • Self-efficacy is the belief that a person can successfully execute this behavior (eg, measured by the validated Diabetes Empowerment Scale) [4]

Patient engagement	• Patient engagement comprises 3 suboutcomes:
	  • Patient activation: patients believe that their own role in managing their care is important, patients' confidence and knowledge to take action, how much they take action, and if patients are capable of staying on course under stress (eg, measured by the Patient Activation Measure PAM13) [33]
	  • Patient involvement: patients' involvement and participation in treatment decisions, and patients' involvement in sharing information, preparing and conducting a medical consultation, and accepting instructions from doctors and nurses [34] (eg, measured by the number of patients that is in possession of an Asthma Action Plan)
	  • Disease knowledge: patients' knowledge of a disease and its related care activities (eg, measured by the Brief Diabetes Knowledge Test) [35]

Health-related quality of life	• All aspects of one's quality of life that are health-related, including physical functioning, social functioning, and mental health (eg, measured by the 36-Item Short Form Survey SF-36) [36] • A reduction in anxiety or stress was considered a suboutcome (eg, measured by the parenting stress index) [37]
Treatment adherence	• The extent to which a person's behavior (taking medication, following a diet, or the execution of lifestyle changes) corresponds with health care providers' recommendations [38] (eg, adherence to HIV medication)
Health care utilization: >all types of encounters between patients and health care providers, including ED[d] visits, hospitalizations, outpatient clinic appointments, and telephone calls	
ED visits and hospitalizations	• Reductions in undesirable events (eg, reductions in emergency department visits and hospitalizations)
Recommended care services	• Increased use of recommended care services by people with uncontrolled disease, and the improved use of preventive care services (eg, follow-up outpatient clinic visits among people with uncontrolled HIV, eye examinations in people with diabetes)
Regular workload	• A decrease in regular workload for health care professionals (eg, patients use email instead of interruptive telephone calls as a first method of contact)
Technology-related outcomes	
Patient satisfaction	• Patient satisfaction with accessing and using patient-centered digital health records • Patient satisfaction with the effects of using patient-centered digital health records (eg, sense of control, perceived quality of care)
Feasibility	• Adherence to patient-centered digital health records and user retention rates, for which no universal cut-off values are available
Acceptability	• The perceived usability of patient-centered digital health records and how these affect behavior, as well as identified facilitators and barriers

[a]HbA_{1c}: glycated hemoglobin.

[b]LDL: low-density lipoprotein.

[c]eGFR: estimated glomerular filtration rate.

[d]ED: emergency department.

Data Extraction

Two independent reviewers (MB and SB) assessed titles, abstracts, and full texts for eligibility. Disagreements were resolved by discussion, if necessary, with a third reviewer (SG).

A modified, electronic version of the standardized Cochrane data extraction form [39] was used to extract the following data items: first author's name; publication year; study design; disease or diseases studied; study aim; country and setting; participants' age and sex; sample size; inclusion and exclusion criteria; follow-up duration; description, features, and purpose of the patient-centered digital health record and (if applicable) of the comparator; size and description of the control group (if applicable); device used; description of health outcomes and results; and main study findings.

Quality Appraisal

For quality appraisal, Joanna Briggs Institute (JBI) critical appraisal tools for RCTs, cross-sectional studies, cohort studies, and quasi-experimental studies were used [40]. JBI tools were modified to better suit the assessment of digital health record studies. Several items were added, including adequate patient-centered digital health record descriptions and selection bias measures, as presented in Multimedia Appendix 2. As the JBI tools differed in the number of items, all scores were converted to a 15-point scale. Articles with a score of ³12 were considered of "high quality," between 8.5 and 11.9 of "medium quality," and <8.5 of "low quality."

Data Synthesis

Associations between patient-centered digital health record use and health outcomes were categorized in 3 groups: "beneficial," "neutral or clinically nonrelevant," or "undesired." Categorizations were determined by our interpretation of study findings, based on meaningful clinical effects and statistical significance (P<.05), and could therefore differ from the authors' conclusions. Statistical significance was considered relevant only if the effect size were clinically significant. If available, minimal clinically important differences were used to assess effect sizes. The summarization of effects was based on the vote-counting method, as no meta-analysis could be performed. The findings were summarized for all conditions, grouped by disease category (diabetes mellitus, cardiopulmonary diseases, hematology-oncology diseases, and other diseases), and grouped according to outcome type (clinical outcomes, patient-reported outcomes, health care utilization, and technology-related outcomes).

Subgroup Analyses

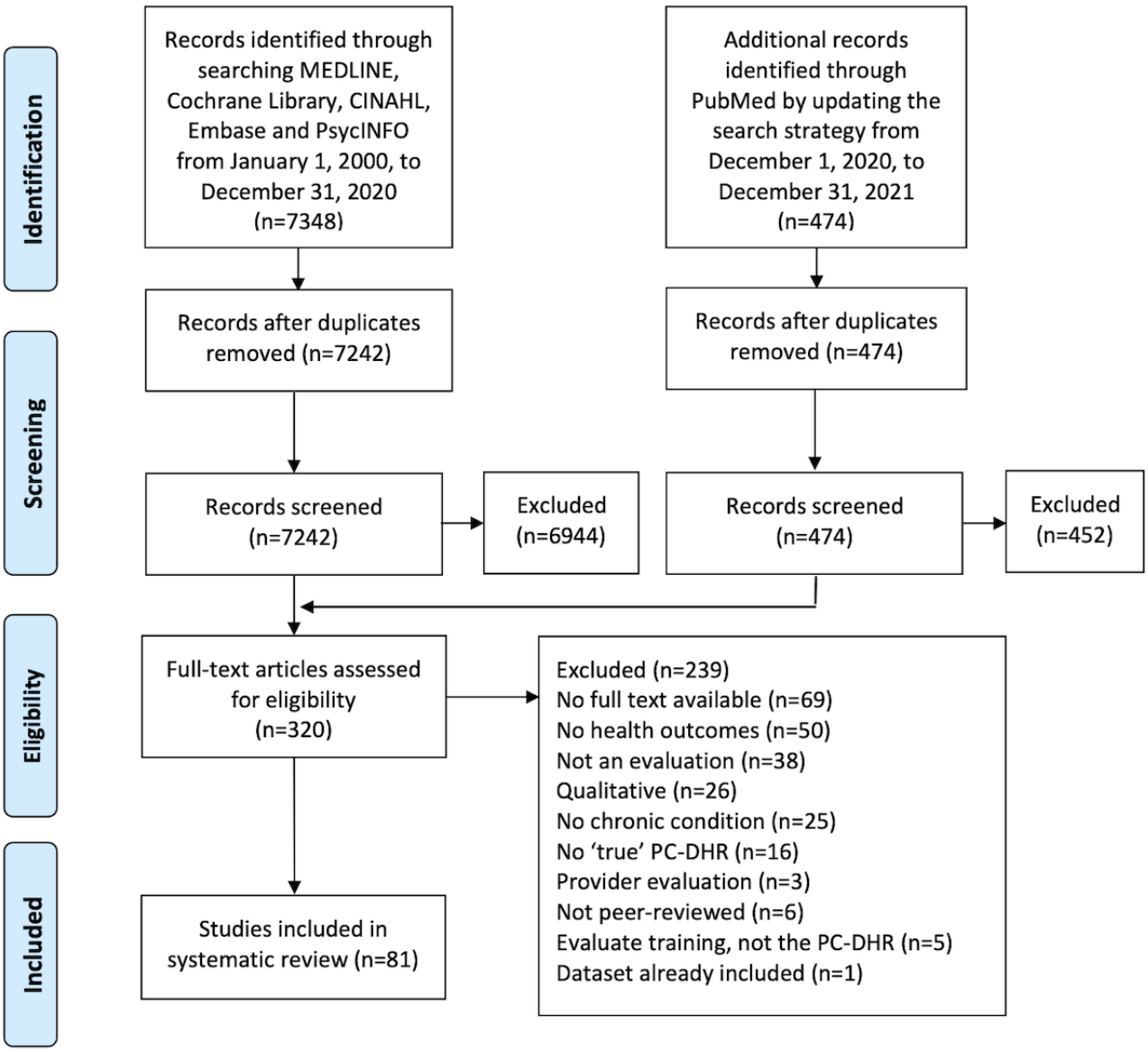
Several subgroup analyses were performed. The first subgroup included conditions with a high disease burden. These included conditions with either impaired social participation or that require a high level of self-management skills. Impaired social participation was defined as being unable to participate in work or school or engage with friends and family as desired because of the condition or its treatment. High self-management skills are defined as recurrent actions demanded from patients to prevent or treat the disease or its consequences, including high disease-related knowledge needed to actively engage in decision-making. This subgroup was determined based on clinical expertise of the study team. Second, we assessed 2 subgroups: patient-centered digital health records that predominantly offered passive features and those that predominantly offered active features. Passive features are those through which the patient receives information but does not actively add information. Active features are those in which the patient performs an action and actively engages with the digital health record. The third subgroup of interest included studies with high methodological quality. A sensitivity analysis was performed to investigate whether our results were influenced by poor quality studies. Finally, the subgroups of interest were studies that included older participants (mean age >55 years), a high number of female participants (>45%), or a racially diverse population (<50% White participants).

Results

Overview

The search yielded 7716 unique publications. After screening the titles and abstracts, 320 full-text articles were retrieved. A total of 81 articles met the inclusion criteria. No non-English articles that met the inclusion criteria were identified. Figure 1 shows the study PRISMA flowchart. In total, 1,639,556 participants were included in the studies of this systematic review. Most (74/81, 91%) studies included only adult participants. Of the total 1,369,913 participants, 99% (n=1,629,660) were adults. Nine studies included children or their parents, with a total number of 9297 children and 599 parents. Sample sizes of studies varied from 10 to 267,208 participants. Furthermore, 46% (747,370/1,639,556) of the participants were female. Of the 81 included studies, health literacy was reported by 7 (9%) studies and insurance status by 15 (20%) studies. Race distribution was reported by 74% (60/81) of studies, of which 47 (78%) studies included a population of which more than half were White and 26 (43%) studies of which >75% were White.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. PC-DHR: patient-centered digital health record.



Study Characteristics

Study characteristics are presented in Tables 2-5 (36 studies are listed in Table 2; 11 studies are listed in Table 3, 14 studies are listed in Table 4, and 20 studies are listed in Table 5). Most investigated conditions were type 1 or 2 diabetes mellitus (37/81, 46%), cardiovascular conditions (14/81, 17%), and malignancies (11/81, 14%). Studies were mostly conducted in the following countries: United States (58/81, 72%), the Netherlands (7/81, 9%), Canada (5/81, 6%), and United Kingdom (3/81, 4%). In addition, 30% (24/81) of the studies were RCTs, 27% (22/81) were cross-sectional studies, 20% (16/81) were retrospective observational cohort studies, and 23% (18/81) were quasi-experimental studies, including pretest-posttest and feasibility studies. One study was a secondary data analysis of the intervention group in an RCT. Of the 55 studies that reported follow-up durations, 6 (7%) studies had a follow-up of less than a month, 25 (31%) studies between 1 and 6 months, 14 (17%) studied between 7 and 12 months, and 10 (12%) studies of >12 months.

Explanations of the patient-centered digital health records investigated in each study are presented in Tables 6-9. Patient-centered digital health records range from a pilot patient portal enabling patients to view a limited set of their medical data to comprehensive PHRs, offering extensive data access and enabling appointment scheduling and prescription refill requests. A minority (12/81, 15%) of studies specifically evaluated ≥1 digital health record features such as secure messaging or a medication adherence module. In addition, 15% (12/81) of studies used a hybrid approach to assess a combination of a digital health record with a connected device, or with training, coaching, or face-to-face visits.

Table 2. Study characteristics of studies investigating diabetes mellitus (of 37 studies investigating diabetes mellitus, 36 are listed in Table 2).[a]

Author, year	Country, setting	Study population, disease, controlled?	Burden[b]	Study design	Sample size	Age (years)[c], mean (SD)	Gender[c] (female), n (%)	Race[c] (White), n (%)
Bailey et al [41], 2019	United States, 2 academic hospitals	Adults with DM[d], on high-risk medication	−	Pilot or feasibility	100	56 (11)	57 (57)	48 (48)
Boogerd et al [42], 2017	Netherlands, 7 medical centers	Parents of children <13 years with DM type 1	+	Pilot or feasibility	I[e]=54, C[f]=51	9.1 (2.7): Children	30 (56)	NR[g]
Byczkowski et al [43], 2014	United States, 1 academic hospital	Parents of children with DM (or CF[h] or JIA[i])	±	Cross-sectional	I=126, C=89	11 (NR)	69 (54.8)	115 (91.3)
Chung et al [44], 2017	United States, outpatient care organization	Adults with DM	−	Cohort	I=12,485, C=2831	56 (12)	5493 (44)	5119 (41)
Conway et al [45], 2019	United Kingdom, Scotland's health system	Patients with DM	−	Cross-sectional	1095	58 (12)	405 (36.99)	873 (78.73)
Devkota et al [46], 2016	United States, 6 PCPs[j]	Patients with DM type 2	−	Cohort	I=409, C=1101	58 (12)[k]	235 (57.5)	250 (61.1)
Dixon et al [47], 2016	United States, 3 community centers	Adults with DM type 2	−	Pilot or feasibility	96	53 (11)	56 (58)	47 (49)
Graetz et al [48], 2018	United States, integrated health system	Adults with DM	−	Cross-sectional	267,208	NR	127,458 (47.7)	116,770 (43.7)
Graetz et al [49], 2020	United States, integrated health system	Adults with DM with at least 1 oral drug	−	Cross-sectional	111,463	64 (13)	51,545 (46.24)	45,205 (40.56)
Grant et al [50], 2008	United States, 11 PCPs	Adults with DM using medication	−	RCT[l]	I=126, C=118	59 (10)	54 (42.9)	117 (92.9)
Lau et al [51], 2014	Canada, 1 academic hospital	Adults with DM	−	Cohort	I=50, C=107	55 (14)	22 (44)	NR
Lyles et al [52], 2016	United States, integrated health system	Adults with DM type 2 using statins	−	Cohort	I=8705, C=9055	61 (11)[k]	4013 (46.1)	3134 (36)[k]
Martinez et al [53], 2021	United States, 4 medical centers	Adults with DM type 2 using medication	−	Pilot or feasibility	60	58 (13)	33 (55)	41 (68)
McCarrier et al [54], 2009	United States, 1 diabetes clinic	Adults <50 years with uncontrolled DM type 1	+	RCT	I=41, C=36	57 (8)	15 (37)	39 (95)
Osborn et al [55], 2013	United States, 1 academic hospital	Adults with DM type 2 using medication	−	Cross-sectional	I=62, C=13	57 (8)	39 (63)	46 (74)
Price-Haywood and Luo [56], 2017	United States, integrated health system	Adults with DM or HT[m]	−	Cohort	I=10,497, C=90,522	NR	6205 (59.11)	8055 (76.74)
Price-Haywood et al [57], 2018	United States, integrated health system	Adults with DM or HT	−	Cohort	I=11,138, C=89,880	58 (13)	6,204 (55.7)	NR
Quinn et al [58], 2018	United States, 26 PCPs	Adults <65 years with DM type 2	−	RCT	I=82, C=25	54 (8)	39 (48)	51 (62)
Reed et al [59], 2015	United States, integrated health system	Adults with DM, HT, CAD[n], asthma, or CHF[o]	±	Cross-sectional	1041	NR	587 (56.4)	618 (59.4)
Reed et al [60], 2019	United States, integrated health system	Adults with DM+HT, CAD, asthma, or CHF	±	Cross-sectional	165,477	NR	79,594 (48.1)	NR (60.9)
Reed et al [61], 2019	United States, integrated health system	Adults with DM, asthma, HT, CAD, CHF or CV event risk	±	Cross-sectional	I=1392, C=407	NR	719 (51.7)	816 (58.6)

Author, year	Country, setting	Study population, disease, controlled?	Burden[b]	Study design	Sample size	Age (years)[c], mean (SD)	Gender[c] (female), n (%)	Race[c] (White), n (%)
Riippa et al [62], 2014	Finland, 10 PCPs	Adults with DM, HT or HC[p]	–	RCT	I=80, C=57	61 (9)	45 (56)	NR
Riippa et al [63], 2015	Finland, 10 PCPs	Adults with DM, HT or HC	–	RCT	I=80, C=57	61 (9)	45 (56)	NR
Robinson et al [64], 2020	United States, 1 veteran hospital	Veterans with uncontrolled DM type 2	–	Cross-sectional	I=446, C=754	66 (8)	28 (6.3)	384 (86.1)
Ronda et al [65], 2014	Netherlands, 62 PCPs+1 hospital	Adults with DM	–	Cross-sectional	I=413, C=758	64 (12)	154 (37.3)	383 (93.6)
Ronda et al [66], 2015	Netherlands, 62 PCPs+1 hospital	Adults with DM	–	Cross-sectional	I=413, C=219	59 (13)	154 (37.3)	383 (93.6)
Sabo et al [67], 2021	United States, 21 practices	Adults with DM type 2	–	Cohort	I=189, C=148	61 (13)	75 (40.9)	113 (72.9)
Sarkar et al [68], 2014	United States, integrated health system	Adults with DM	–	Cohort	I=8705, C=9055	61 (11)[k]	4013 (46.1)	5072 (58.27)
Seo et al [69], 2020	South Korea, 1 academic hospital	Patients with DM	–	Cohort	I=133, C=7320	54 (10)	23 (17.3)	NR
Sharit et al [70], 2018	United States, 1 veterans center	Overweight veterans with prediabetes	–	Pilot or feasibility	38	58 (8)	9 (24)	8 (21)[k]
Shimada et al [71], 2016	United States, Veteran registry	Veterans with uncontrolled DM, HT or LDL[q]	–	Cohort	I=50,482, C=61,204	61 (10)	2060 (4.08)	35,761 (70.84)
Tenforde et al [72], 2012	United States, 1 community hospital	Adults <75 years with DM	–	Cohort	I=4036, C=6710	59 (10)	1857 (46)[k]	3,390 (84)[k]
van Vugt et al [73], 2016	Netherlands, 52 PCPs	Patients with DM type 2	–	RCT	I=66, C=66	68 (10)	54 (41)	91 (69)
Vo et al [74], 2019	United States, integrated health system	Adults <80 years with DM type 2	–	RCT	I=673, C=603	61 (10)	296 (44)	394 (58.5)
Wald et al [75], 2009	United States, 230 PCPs	Patients with DM type 2	–	RCT	126	59 (NR)	53 (42.1)	117 (92.9)
Zocchi et al [76], 2021	United States, nationwide	Patients with DM type 2, partly uncontrolled	–	Cohort	95,043	63 (10)	4,339 (4.57)	68,954 (72.55)

[a]All studies are listed in Tables 2-5 and are reported in the disease category of the condition that is most prominently investigated. The study by Druss et al [77] is therefore listed in Table 5.

[b]If conditions are considered to have a high disease burden or demand high self-management skills, a positive sign is shown. Otherwise, a sign is indicated. A ± sign indicates that multiple diseases have been studied, and only some of the diseases were considered to have a high disease burden.

[c]If available, age (years), gender, and race were reported by digital health record users (“the intervention group”).

[d]DM: diabetes mellitus.

[e]I: intervention.

[f]C: control.

[g]NR: not reported.

[h]CF: cystic fibrosis.

[i]JIA: juvenile idiopathic arthritis.

[j]PCP: primary care practice.

[k]Presented numbers were estimated based on the data provided in the original articles.

[l]RCT: randomized controlled trial.

[m]HT: hypertension.

[n]CAD: coronary artery disease.

[o]CHF: congestive heart failure.

[p]HC: hypercholesterolemia.

[q]LDL: low-density lipoprotein.

Table 3. Study characteristics of studies investigating cardiopulmonary diseases (of 21 studies investigating cardiopulmonary diseases, 11 are listed in Table 3).[a]

Author, year	Country, setting	Study population, disease, controlled?	Burden[b]	Study design	Sample size	Age (years)[c], mean (SD)	Gender[c] (female), n (%)	Race[c] (White), n (%)
Aberger et al [78], 2014	United States, renal transplant clinic	Postrenal transplant patients with HT[d]	+	Pilot or feasibility	66	54 (NR[e])	34 (52)[f]	48 (72)[f]
Ahmed et al [79], 2016	Canada, 2 academic hospitals	Adults with asthma using medication	+	RCT[g]	I[h]=49, C[i]=51	NR	32 (68)	NR
Apter et al [80], 2019	United States, multicenter hospitals	Adults with asthma using prednisone	+	RCT	I=151, C=150	49 (13)	270 (89.7)	4 (1.3)
Fiks et al [81], 2015	United States, 3 PCPs[j]	Children aged 6-12 years with asthma, partly uncontrolled	+	RCT	I=30, C=30	8.3 (1.9)	26 (87) among parents	13 (43)
Fiks et al [82], 2016	United States, 20 PCPs	Children aged 6-12 years with asthma, partly uncontrolled	+	Pilot or feasibility	I=237, C=8896	NR	101 (42.8)	144 (61.5)
Kogut et al [83], 2014	United States, 1 community hospital	Adults aged >49 years with cardiopulmonary disorders	±	Pilot or feasibility	30	NR	14 (47)	NR
Kim et al [84], 2019	South Korea, 1 academic hospital	Patients with obstructive sleep apnea	−	RCT	I=30, C=13	43 (10)[f]	NR (15)	NR
Lau et al [85], 2015	Australia, nationwide	Adults with asthma	+	RCT	I=154, C=176	40 (14)	124 (80.5)	NR
Manard et al [86], 2016	United States, PCP registry	Adults with uncontrolled HT	−	Cohort	I=400, C=1171	61 (12)	262 (65.5)	72
Toscos et al [87], 2020	United States, 1 community hospital	Patients with nonvalvular AF[k] with OAC[l]	+	RCT	I=76, C=77	71 (9)	60 (37.5)	153 (99.4)
Wagner et al [88], 2012	United States, 24 PCPs	Patients with hypertension, partly uncontrolled	−	RCT	I=193, C=250	55 (12)	145 (75.1)	96 (50.5)

[a]All studies are listed in Tables 2-5 and are reported in the disease category of the condition that is most prominently investigated. The studies by Price-Haywood and Luo [56], Price-Haywood et al [57], Reed et al [59], Reed et al [60], Reed et al [61], Riippa et al [62], Riippa et al [63], Shimada et al [71] are listed in Table 2. The study by Martinez Nicolás et al [89] is listed in Table 4. The study by Druss et al [77] is therefore listed in Table 5.

[b]If conditions are considered to have a high disease burden or demand high self-management skills, a positive sign is shown. Otherwise, a sign is indicated. A ± sign indicates that multiple diseases have been studied, and only some of the diseases were considered to have a high disease burden.

[c]If available, age (years), gender, and race were reported by digital health record users ("the intervention group").

[d]HT: hypertension.

[e]NR: not reported.

[f]Presented numbers were estimated based on the data provided in the original articles.

[g]RCT: randomized controlled trial.

[h]I: intervention.

[i]C: control.

[j]PCP: primary care practice.

[k]AF: atrial fibrillation.

[l]OAC: oral anticoagulant drug.

Table 4. Study characteristics of studies investigating hematological and oncological diseases (n=14).

Author, year	Country, setting	Study population, disease, controlled?	Burden[a]	Study design	Sample size	Age (years)[b], mean (SD)	Gender[b] (female), n (%)	Race[c] (White), n (%)
Cahill et al [90], 2014	United States, cancer center	Adults with glioma	+	Cross-sectional	186	44 (13)	87 (46.8)	149 (86.1)
Chiche et al [91], 2012	France, 1 community hospital	Adults with ITP[c]	±	RCT[d]	I[e]=28, C[f]=15	48 (15)[g]	21 (75)	NR[h]
Collins et al [92], 2003	United Kingdom, hemophilia centers	Patients with hemophilia >11 years	+	Pilot or feasibility	10	NR	NR	NR
Coquet et al [93], 2020	United States, cancer center	Patients with cancer+chemotherapy	+	Cohort	I=3223, C=3223	59 (15)	1,554 (49.78)	1,804 (49.68)
Groen et al [94], 2017	Netherlands, cancer center	Patients with lung cancer	+	Pilot or feasibility	37	60 (8)	16 (47)	37 (100)
Hall et al [95],2014	United States, Cancer Center	Patients with resection for CRC[i] or EC[j]	+	Pilot or feasibility	49	59 (12)[g]	37 (76)	48 (98)
Hong et al [96], 2016	United States, academic pediatric hospital	Children aged 13-17 years with cancer or a blood disorder+parents	+	Cross-sectional	46	15 (1.2)[g]	10 (63) among children	NR
Kidwell et al [97], 2019	United States, multicenter hospitals	Patients aged 13-24 years with sickle cell disease	+	Pilot or feasibility	44	19 (NR)	24 (55)	0 (0)
Martinez Nicolás et al [89], 2019	Spain, 4 community hospitals	Patients with COPD[k], CHF[l], or hematologic malignancy	+	Pilot or feasibility	577,121	42 (23)	319,725[g] (55)	NR
O'Hea et al [98], 2021	United States, cancer centers	Adult women with nonmetastatic breast cancer ending treatment	+	RCT	I=100, C=100	61 (11)	100 (100)	85 (85)
Pai et al [99], 2013	Canada, cancer center	Adult men with prostate cancer	+	Cross-sectional	17	64 (7)[g]	0 (0)	16 (95)
Tarver et al [100], 2019	United States, academic hospital	Patients with colorectal cancer	+	Cross-sectional	22	58 (10)	10 (45)	NR
Wiljer et al [101], 2010	Canada, breast cancer registry	Patients with breast cancer	+	Pilot or feasibility	311	NR	303 (99.7)	NR
Williamson et al [102], 2017	United States, pediatric cancer center	Pediatric cancer survivors	+	Cohort	56	NR	27 (48)	49 (88)

[a]If conditions are considered to have a high disease burden or demand high self-management skills, a positive sign is shown. Otherwise, a sign is indicated. A ± sign indicates that multiple diseases have been studied, and only some of the diseases were considered to have a high disease burden.

[b]If available, age (years), gender, and race were reported by digital health record users ("the intervention group").

[c]ITP: idiopathic thrombocytopenic purpura.

[d]RCT: randomized controlled trial.

[e]I: intervention.

[f]C: control.

[g]Presented numbers were estimated based on the data provided in the original articles.

[h]NR: not reported.

[i]CRC: colorectal cancer.

[j]EC: endometrial cancer.

[k]COPD: chronic obstructive pulmonary disease.

[l]CHF: congestive heart failure.

Table 5. Study characteristics of studies investigating other diseases (of 21 studies investigating other diseases, 20 are listed in Table 5). Diseases include kidney disease (n=3, 15%), mental health disorders (n=3, 15%), multiple sclerosis (n=2, 10%), inflammatory bowel disease (n=2, 10%), rheumatologic conditions (n=2, 10%), and others (n=8, 40%).[a]

Author, year	Country, setting	Study population, disease, controlled?	Burden[b]	Study design	Sample size	Age (years)[c], mean (SD)	Gender[c] (female), n (%)	Race[c] (White), n (%)
Anand et al [103], 2017	Thailand, HIV clinic	MSM[d] and transgender women with HIV, partly uncontrolled	+	RCT[e]	186	30 (10)[f]	7 (4)	0 (0)
Bidmead and Marshall [104], 2016	United Kingdom, 1 community hospital	Patients with IBD[g]	+	Cross-sectional	60	NR[h]	NR	NR
Crouch et al [105], 2015	United States, 1 HIV clinic	Veterans with HIV, partly uncontrolled	+	Cross-sectional	I[i]=20, C[j]=20	43 (11)	1 (5)	19 (95)
Druss et al [106], 2014	United States, 1 mental health center	Patients with a mental disorder+chronic condition	+	RCT	I=85, C=85	49 (7)	42 (49)	13 (15)
Druss et al [77], 2020	United States, 2 mental health centers	Patients with a mental disorder+DM[k], HT[l], or HC[m]	+	RCT	I=156, C=155	51 (6.5)	95 (61)	29 (19)
Jhamb et al [107], 2015	United States, 4 nephrology clinics	Adults visiting nephrology clinics, partly uncontrolled	+	Cross-sectional	1098	58 (16)	549 (50)	952 (86.7)
Kahn et al [108], 2010	United States, HIV clinic	Patients with HIV or AIDS	+	Pilot or feasibility	136	NR	15 (11)[f]	106 (78)[f]
Keith McInnes et al [109], 2013	United States, 8 Veteran hospitals	Veterans with HIV, partly uncontrolled	+	Cross-sectional	1871	NR	51 (2.73)	342 (18.28)
Keith McInnes et al [110], 2017	United States, Veterans care system	Veterans with HIV+detectable viral load, partly uncontrolled	+	Cohort	3374	NR	128 (3.79)	1130 (33.49)
Kiberd et al [111], 2018	Canada, dialysis clinic	Adult with home dialysis	+	Pilot or feasibility	41	57 (2)	13 (48)	NR
Lee et al [112], 2017	South Korea, 1 surgery department	Patients with cleft lip or cleft palate surgery	−	Pilot or feasibility	50	36 (NR)	33 (66)	NR
Miller et al [113], 2011	United States, MS[n] clinic	Patients with MS	+	RCT	I=104, C=102	48 (9)	73 (71.6)	80 (78.4)
Navaneethanet al [114], 2017	United States, multiple health centers	Adults with chronic kidney disease, partly uncontrolled	+	RCT	I=152, C=57	68 (NR)[f]	79 (52)	117 (77)
Plimpton [115], 2020	United States, HIV clinic	Women with HIV, partly uncontrolled	+	Pilot or feasibility	22	41 (11)	22 (100)	7 (32)
Reich et al [116], 2019	United States, 1 community hospital	Adults with IBD[o]	+	RCT	I=64, C=63	42 (16)	28 (46)	48 (77)
Scott Nielsen et al [117], 2012	United States, 1 academic center	Adults with MS	+	Cross-sectional	I=120, C=120	45 (11)	90 (75)	115 (95.8)
Son and Nahm [118], 2019	United States, online senior community	Patients >49 years with 1 or more chronic conditions	±	Secondary data analysis	272	70 (9)	191 (70.2)	213 (78.3)
Tom et al [119], 2012	United States, integrated health system	Parents of children age <6 years with 1 or more chronic conditions	±	Cross-sectional	I=166, C=90	3 (1)	66 (39.8)	113 (68.1)
van den Heuvel et al [120], 2018	Netherlands, 3 hospitals	Adults with bipolar disorder	+	Cross-sectional	39	45 (11)	44 (67)	NR

Author, year	Country, setting	Study population, disease, controlled?	Burden[b]	Study design	Sample size	Age (years)[c], mean (SD)	Gender[c] (female), n (%)	Race[c] (White), n (%)
van der Vaart et al [121], 2014	Netherlands, 1 hospital	Patients with rheumatoid arthritis	+	Cross-sectional	214	62 (13)	140 (65.4)	NR

[a]All studies are listed in Tables 2-5 and are reported in the disease category of the condition that is most prominently investigated. The study by Byczkowski et al [43] is therefore listed in Table 2.

[b]If conditions are considered to have a high disease burden or demand high self-management skills, a positive sign is shown. Otherwise, a sign is indicated. A ± sign indicates that multiple diseases have been studied, and only some of the diseases were considered to have a high disease burden.

[c]If available, age (years), gender, and race were reported by digital health record users (“the intervention group”).

[d]MSM: men who have sex with men.

[e]RCT: randomized controlled trial.

[f]Presented numbers were estimated based on the data provided in the original articles.

[g]IBD: inflammatory bowel disease.

[h]NR: not reported.

[i]I: intervention.

[j]C: control.

[k]DM: diabetes mellitus.

[l]HT: hypertension.

[m]HC: hypercholesterolemia.

[n]MS: multiple sclerosis.

[o]IBD: inflammatory bowel disease.

Table 6. Patient-centered digital health record descriptions for disease category diabetes mellitus (of 37 studies investigating diabetes mellitus, 36 are listed in Table 6).[a]

Author, year	Name	Type	What is evaluated?[b]	Passive features	Active features	Focus[c]
Bailey et al [41], 2019	Electronic Medication Complete Communication	PP[d]	Adherence module alone	View health information (medical summary), read after-visit summary, read educational material	Report medication concerns, monitor medication use	Active
Boogerd et al [42], 2017	Sugarspace	PP	PP	View treatment goals, read educational material	Parent-professional communication, peer support	Active
Byczkowski et al [43], 2014	In-house developed	PP	PP	View health information (including laboratory results, medication), view appointments, read disease-specific information	Messaging, upload documents, receive reminders	Passive
Chung et al [44], 2017	Not reported	PP	Messaging	View health information	Messaging	Active
Conway et al [45], 2019	My Diabetes My Way	Tethered PHR[e]	PHR	View health information from primary and secondary care (including clinical parameters, medication, and correspondence), read educational material	Report self-measurements	Passive
Devkota et al [46], 2016	MyChart	PP	PP	View health information (including laboratory results, diagnoses, medication, vital signs), read educational material	Messaging, request prescription refills, schedule appointments, pay bills	Passive
Dixon et al [47], 2016	CareWeb	PP	Medication module alone	View health information (including measurements, medication)	Report barriers to medication adherence	Passive
Graetz et al [48], 2018 and Graetz et al [49], 2020	"Kaiser Permanente portal"	PP	PP	View health information (including laboratory results)	Messaging, schedule appointments, request prescription refills, pay bills	Active
Grant et al [50], 2008	Not reported	PP	PP	View health information (including medication, laboratory results)	Edit medication lists, messaging, report adherence barriers or adverse effects	Active
Lau et al [121], 2014	BCDiabetes	PP	PP	View health information (including laboratory results), view care plan, read educational material	Messaging, use a journal	Passive
Lyles et al [52], 2016	"Kaiser Permanente portal"	PP	Medication module alone	View health information (including medical history, laboratory results, and visit summaries)	Messaging, schedule appointments, request prescription refills	Active
Martinez et al [53], 2021	My Diabetes Care, part of My Health at Vanderbilt	PP	Diabetes module	View health information (including laboratory results and vaccinations), visualize information, read educational material	Messaging, peer support, decision support tools	Active
McCarrier et al [54], 2009	Living with Diabetes Intervention	PP	PP+case manager	View health information (including correspondence, action plans, and laboratory results), read diabetes-related information	Upload blood glucose readings, use a journal	Active
Osborn et al [55], 2013	My Health At Vanderbilt	PP	PP	View health information (including vital signs, laboratory results, and medication), read educational information	Messaging, manage appointments, use health screening tools, pay bills	Passive
Price-Haywood and Luo [56], 2017 and Price-Haywood et al [57], 2018	MyOchsner	PP	PP	View health information (including an after-visit summary, allergies, and laboratory results)	Messaging, request prescription refills, schedule appointments	Passive
Quinn et al [58], 2018	Not reported	PP	PP	View self-reported health information (including medication and measurements), read educational material	Messaging, report self-measurements and medication changes, receive automated feedback	Active

Author, year	Name	Type	What is evaluated?[b]	Passive features	Active features	Focus[c]
Reed et al [59], 2015	"Kaiser Permanente portal"	PP	Messaging alone	View health information (including laboratory results and correspondence)	Messaging, request prescription refills, schedule appointments	Active
Reed et al [60], 2019 (1) and Reed et al [61], 2019	"Kaiser Permanente portal"	PP	PP	View health information from primary care and secondary care (including laboratory results and visit summaries)	Messaging, request prescription refills, schedule visits	Passive
Riippa et al [62], 2014 and Riippa et al [63], 2015	Not reported	PP	PP	View health information (including diagnoses, laboratory results, vaccinations, and medication), view care plan, read educational material	Messaging	Passive
Robinson et al [64], 2020	My HealtheVet	PP	Messaging alone	View health information (including medication and correspondence), view appointments	Messaging, request prescription refills, receive reminders, upload notes and measurements, use a journal	Passive
Ronda et al [65], 2014 and Ronda et al [66], 2015	Digitaal logboek	PP	PP	View diabetes-specific health information (including laboratory results, diagnoses, and medication), view treatment goals, view appointments	Messaging, upload self-measurements	Passive
Sabo et al [67], 2021	Diabetes Engagement and Activation Platform	PP	PP	View health information (including medication and self-reported glucose measurements)	Report diet, physical activity, blood glucose measurements, complications, mental health and goals, receive alerts	Active
Sarkar et al [68], 2014	"Kaiser Permanente portal"	PP	PP	View health information (including medical history, laboratory results, and visit summaries), view appointments	Messaging, request prescription refills	Passive
Seo et al [69], 2020	My Chart in My Hand	Tethered PHR	PHR+sugar function	View health information (including laboratory results, medication, allergies, diagnoses)	Edit information, schedule appointment; sugar function: log treatment, food intake, and exercise	Active
Sharit et al [70], 2018	My HealtheVet	PP	Track Health module+wearable	View health information (including medication and correspondence), view appointments	Messaging, request prescription refills, receive reminders; track Health module: record diet and activity, upload data from connected accelerometer	Active
Shimada et al [71], 2016	My HealtheVet	PP	Messaging, prescription refills	View health information (including medication and correspondence), view appointments	Messaging, request prescription refills, receive reminders, upload notes and self-measurements, use a journal	Active
Tenforde et al [72], 2012	MyChart	PP	PP	View health information (including diagnoses and laboratory results), read diabetes educational material	Messaging, view glucometer readings, receive reminders	Passive
van Vugt et al [73], 2016	e-Vita	Tethered PHR	PHR+personal coach	View health information (measurements), read diabetes education	Messaging, self-management support program for personal goal setting and evaluation	Active
Vo et al [74], 2019	"Kaiser Permanente portal"	PP	PP+PreVisit Prioritization messaging	View health information (including medical history, laboratory results, and visit summaries), view appointments	PreVisit Prioritization messaging to report priorities before a clinic visit, request prescription refills	Active

Author, year	Name	Type	What is evaluated?[b]	Passive features	Active features	Focus[c]
Wald et al [75], 2009	Patient Gateway	Tethered PHR	PHR	View health information (including medication, allergies, and laboratory results)	Suggest corrections, report care concerns, ask for referrals, create care plans before visits	Active
Zocchi et al [76], 2021	My HealtheVet	PP	PP	View health information (including medication, laboratory results, imaging, and correspondence)	Messaging, requesting prescription refills, download health information	Active

[a]All studies are listed once in Tables 2-5 and are reported in the disease category of the condition that is most prominently investigated. We have included only the functionalities that the authors have reported in their articles. We have applied the taxonomy as presented in Textbox 1 on the information provided by the authors. Therefore, our classification of patient-centered digital health records might not correspond with the term used by the authors.

[b]In this column, we indicated whether authors evaluated the complete patient-centered digital health record, or only part of it.

[c]By definition, patient-centered digital health records have both passive and active features. In this column, we indicate whether patient-centered digital health records predominantly offer passive or active features. In passive features, patients receive information but do not actively add it. In terms of active features, patients perform an action and actively engage with the portal.

[d]PP: patient portal.

[e]PHR: personal health record.

Table 7. Patient-centered digital health record descriptions for disease category cardiopulmonary diseases (of 21 studies investigating cardiopulmonary diseases, 11 are listed in Table 7).[a]

Author, year	Name	Type	What is evaluated?[b]	Passive features	Active features	Focus[c]
Aberger et al [78], 2014	Good Health Gateway	PP[d]	PP+BP[e] cuff	View BP measurements, view treatment goals	Communicate self-reported adherence, receive automated and tailored feedback	Active
Ahmed et al [79], 2016	My Asthma Portal	PP	PP	View health information (including medication and diagnoses), read general and tailored asthma information	Monitor and receive feedback on self-management practices	Passive
Apter et al [80], 2019	MyChart	PP	PP	View health information (including laboratory results, vaccinations, and medication), view appointments	Messaging, request prescription refills, schedule appointments	Passive
Fiks et al [81], 2015 and Fiks et al [82], 2016	MyAsthma	PP	PP	View care plan, read educational material	Report symptoms, treatment adherence, concerns and side effects	Active
Kim et al [84], 2019	MyHealthKeeper	Tethered PHR[f]	PHR+activity tracker	View previously uploaded self-reported data	Upload self-reported data (eg, diet, sleep, weight, BP, step count), connect with wearables, receive feedback from health care providers	Active
Kogut et al [83], 2014	ER-Card	Untethered PHR	PHR+home visits by pharmacists	View patient-reported medication list	Pharmacists view and review patient-reported medication lists, and discuss potential concerns in home visits	Active
Lau et al [85], 2015	Healthy.me	Untethered PHR	PP+extra feature	View Asthma Action Plan, read educational content	Schedule appointments, peer support, self-report medication, use a journal	Passive
Manard et al [86], 2016	Not reported	PP	PP+BP cuff	View health information (including laboratory results, vital signs, and diagnoses)	Messaging, request prescription refills, upload measurements from connected BP cuff	Passive
Toscos et al [87], 2020	MyChart	PP	PP+smart pill bottle	View health information (including laboratory results, vaccinations, and medication), view appointments	Messaging, request prescription refills, schedule appointments Smart Pill Bottle: a device that sends notifications when a user opens or fails to open the lid, based on the dose schedule	Active
Wagner et al, 2012 [88]	MyHealthLink	Tethered PHR	PHR	View health information (including diagnoses, medication, and allergies), read educational material	Messaging, goal setting, upload self-measurements (including BP)	Active

[a]All studies are listed once in Tables 2-5 and are reported in the disease category of the condition that is most prominently investigated. We have included only the functionalities that the authors have reported in their articles. We have applied the taxonomy as presented in Textbox 1 on the information provided by the authors. Therefore, our classification of patient-centered digital health records might not correspond with the term used by the authors.

[b]In this column, we indicated whether authors evaluated the complete patient-centered digital health record, or only part of it.

[c]By definition, patient-centered digital health records have both passive and active features. In this column, we indicate whether patient-centered digital health records predominantly offer passive or active features. In passive features, patients receive information but do not actively add it. In terms of active features, patients perform an action and actively engage with the portal.

[d]PP: patient portal.

[e]BP: blood pressure.

[f]PHR: personal health record.

Table 8. Patient-centered digital health record descriptions for disease category hematological and oncological diseases (n=14).[a]

Author, year	Name	Type	What is evaluated?[b]	Passive features	Active features	Focus[c]
Cahill et al [90], 2014	MyMDAnderson	Tethered PHR[d]	PHR	View health information (including correspondence, operative reports, laboratory results, and imaging), read education material	Messaging, request prescription refills, schedule appointments	Passive
Chiche et al [91], 2012	Sanoia	PP[e]	PP+ITP[f] features	View health information (including allergies, vaccinations, medication, and test results), ITP-specific educational material, read emergency protocols	Messaging	Passive
Collins et al [92], 2003	Advoy	PP	PP	View health information (treatment regimen), read educational material	Registration of symptoms and medication use, automated alerts are sent to professionals	Active
Coquet et al [93], 2020	MyHealth portal	PP	Email use	View health information (including laboratory results)	Messaging, schedule appointments, request prescription refills, pay bills	Active
Groen et al [94], 2017	MyAVL	PP	PP	View health information (including laboratory results, lung function, and correspondence), view appointments, read personalized information	Upload patient-reported outcomes, receive tailored physical activity advice	Active
Hall et al [95], 2014	MyFoxChase	PP	Genetic screening	View health information (including laboratory results), view appointments, read educational material	Messaging, receive alerts if genetic screening results are available	Passive
Hong et al [96], 2016	MyChart	PP	PP	View health information (including laboratory results, medication, allergies)	Messaging, schedule appointments, request prescription refills, use a journal	Passive
Kidwell et al [97], 2019	MyChart	PP	PP	View health information (including laboratory results, medication, diagnoses, and allergies), view appointments, read information about sickle cell disease	Messaging	Passive
Martinez Nicolás et al [89], 2019	Not reported	PP	PP	View health information (including laboratory results, imaging, and medication)	Messaging, teleconsulting, schedule appointments, upload glucose measurements	Active
O'Hea et al [98], 2021	Polaris Oncology Survivorship Transition	PP	PP	View health information (including diagnoses, operative reports, and medication), view appointments, read educational material	Request a referral	Passive
Pai et al [99], 2013	PROVIDER	Tethered PHR	PHR	View health information (including laboratory results, medication, pathology, imaging, and correspondence), read educational material	Messaging, use decision support tools, fill in questionnaires	Passive
Tarver et al [100], 2019	OpenMRS	Tethered PHR	PHR+extra feature	View health information (including treatment history, diagnoses, and care plan), view a treatment summary, read educational material	Messaging, peer support	Passive
Wiljer et al [101], 2010	InfoWell	Tethered PHR	PHR	View health information (including medication, laboratory results, imaging, and pathology), view appointments	Patients can organize and upload care information	Passive
Williamson et al [102], 2017	SurvivorLink	Untethered PHR	PHR	Read educational material	Upload health documents and share these with professionals	Active

[a]All studies are listed once in Tables 2-5 and are reported in the disease category of the condition that is most prominently investigated. We have included

only the functionalities that the authors have reported in their articles. We have applied the taxonomy as presented in Textbox 1 on the information provided by the authors. Therefore, our classification of patient-centered digital health records might not correspond with the term used by the authors.

[b]In this column, we indicated whether authors evaluated the complete patient-centered digital health record, or only part of it.

[c]By definition, patient-centered digital health records have both passive and active features. In this column, we indicate whether patient-centered digital health records predominantly offer passive or active features. In passive features, patients receive information but do not actively add it. In terms of active features, patients perform an action and actively engage with the portal.

[d]PHR: personal health record.

[e]PP: patient portal.

[f]ITP: idiopathic thrombocytopenic purpura.

Table 9. Patient-centered digital health record descriptions for disease category other diseases (of 21 studies investigating other diseases, 20 are listed in Table 9).[a]

Author, year	Name	Type	What is evaluated?[b]	Passive features	Active features	Focus[c]
Anand et al [103], 2017	Adam's Love	PP[d]	PP	View health information (HIV test results), receive appointment reminders	Schedule HIV test appointments, use e-counseling, receive appointment reminders	Active
Bidmead et al [104], 2016	Patients Know Best	Tethered PHR[e]	PHR	View health information (including medication, laboratory results, and correspondence), read educational material	Communication with health care providers, upload and share health information	Active
Crouch et al [105], 2015	My HealtheVet	PP	PP	View health information (including laboratory results and correspondence)	Messaging, request prescription refills	Passive
Druss et al [106], 2014	My-HealthRecord	PP	PP+training	View health information (including diagnoses, measurements, laboratory results, medication, and allergies), view treatment goals	Prompts remind patients of routine preventive service	Passive
Druss et al [77], 2020	Not reported	PP	PP+training	View health information (including medication, allergies, measurements, and laboratory results)	Formulate long-term goals, that are translated into action plans with progress tracking	Active
Jhamb et al [107], 2015	Not reported	PP	PP	View health information (including diagnoses, allergies, immunizations, and laboratory results)	Messaging, schedule appointments, request prescription refills	Passive
Kahn et al [108], 2010	MyHERO	PP	PP	View health information (including diagnoses, medication, laboratory results, and allergies), view appointments, read information on interpreting test results	Upload notes and self-measurements	Passive
Keith McInnes et al [109], 2013 and Keith McInnes et al [110], 2017	My HealtheVet	PP	PP	View health information (including medication and correspondence), view appointments	Messaging, request prescription refills, receive reminders, upload notes and self-measurements, use a journal	Passive
Kiberd et al [111], 2018	RelayHealth	PP	PP	View health information (including test results and medication)	Messaging	Active
Lee et al [112], 2017	CoPHR	PP	PP	View health information (including diagnoses, laboratory results, medication, allergies, vital signs, and correspondence), view appointments, view treatment plan, read educational information	Manage and edit appointments and health information	Passive
Miller et al [113], 2011	Mellen Center Care Online	Untethered PHR	PHR	Review previously entered symptoms and HRQoL[f]	Messaging, report symptoms and HRQoL and evaluate changes, preparation for appointments	Active
Navaneethan et al [114], 2017	MyChart	PP	PP+part of users received training	View health information (including medication and laboratory results), read educational material	Messaging, schedule appointments, request prescription refills	Passive
Plimpton [115] 2020	Not reported	PP	PP	View health information	Messaging	Passive
Reich et al [116], 2019	MyChart	PP	PP	View health information (including laboratory results, diagnoses, medication, and vital signs)	Messaging	Passive
Scott Nielsen et al [117], 2012	PatientSite10	PP	PP	View health information (including laboratory results, and imaging), read educational material	Messaging, schedule appointments, request prescription refills, upload self-measurements, pay bills	Active

Author, year	Name	Type	What is evaluated?[b]	Passive features	Active features	Focus[c]
Son and Nahm [118], 2019	MyChart	PP	PP+training	View health information (including medication and laboratory results), read educational material	Messaging, schedule appointments, request prescription refills	Passive
Tom et al [119], 2012	My-GroupHealth	PP	PP	View health information (including diagnoses, medication, and test results), read after-visit summaries, proxy access	Messaging, schedule appointments	Passive
van den Heuvel et al [120], 2018	"PHR-BD"	Tethered PHR	Tethered PHR+mood chart	View health information (including diagnoses, laboratory results, medication, and correspondence), read educational material	Messaging, report symptoms in a mood chart, view personal crisis plan	Active
van der Vaart et al [121], 2014	Not reported	PP	PP	View health information (including diagnoses, medication, and laboratory results), read educational material	Report and monitor HRQoL outcomes	Active

[a]All studies are listed once in Tables 2-5 and are reported in the disease category of the condition that is most prominently investigated. We have included only the functionalities that the authors have reported in their articles. We have applied the taxonomy as presented in Textbox 1 on the information provided by the authors. Therefore, our classification of patient-centered digital health records might not correspond with the term used by the authors.

[b]In this column, we indicated whether authors evaluated the complete patient-centered digital health record, or only part of it.

[c]By definition, patient-centered digital health records have both passive and active features. In this column, we indicate whether patient-centered digital health records predominantly offer passive or active features. In passive features, patients receive information but do not actively add it. In terms of active features, patients perform an action and actively engage with the portal.

[d]PP: patient portal.

[e]PHR: personal health record.

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

Outcomes

An overview of reported associations for each health outcome is shown in Figure 2. The proportions of beneficial effects reported per health outcome are presented in Multimedia Appendices 3 and 4. For high-quality studies, proportions are presented in Multimedia Appendix 3. An overview of study conclusions and associated outcomes is presented in Tables 10-13. Studies were grouped according to disease group.

Figure 2. Health outcomes associated with patient-centered digital health record use. Associations refer to meaningful clinical effects or statistical significance. If studies report multiple health outcome within 1 category, each health outcome is included separately. *The proportion of health outcomes for which beneficial effects were reported. ED: emergency department.

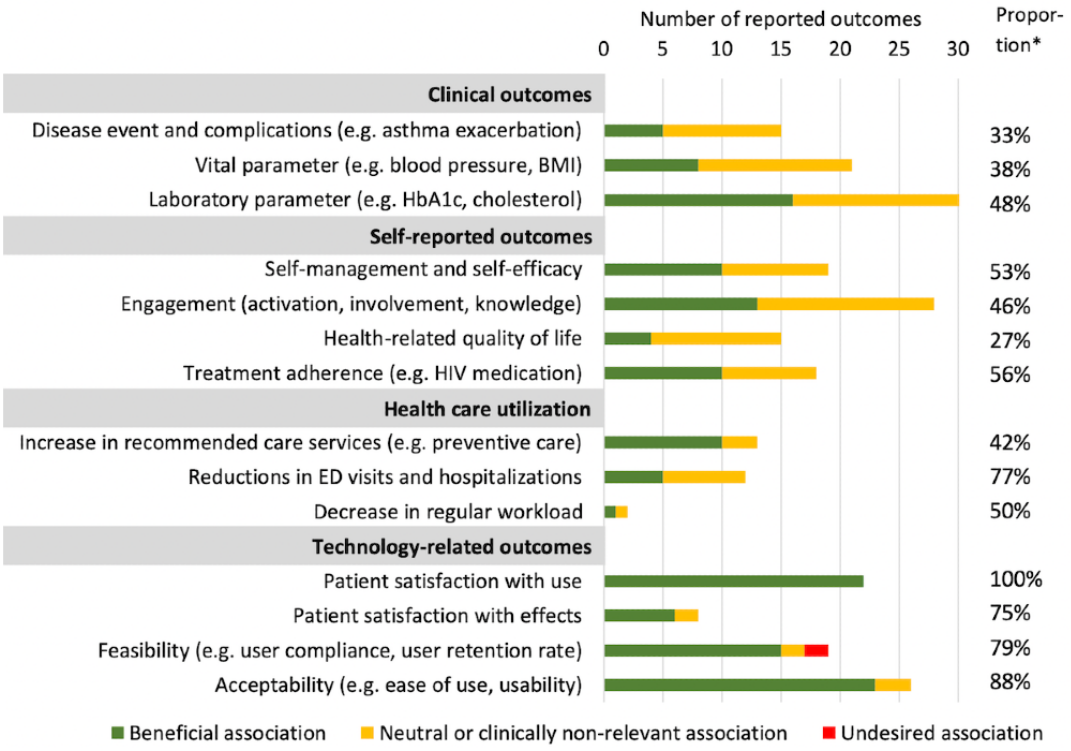


Table 10. Conclusions and health outcomes: all studies investigating diabetes (n=37), of which 8 (22%) are of high methodological quality.[a]

Author, year	Participants	Comparison	Main conclusion	Study design	Clinical	Patient reported	Care utilization	Technology	Quality[b]
Boogerd et al [42], 2017	Parents of children with DM[c] type 1	PP[d] users versus PP nonusers	Patient portal use is not associated with less parental stress. The more stress, the more parents use the portal.	QE[e]	☒	☒	—[f]	☒	☒
Lau et al [51], 2014	Patients with DM	Pretest PP nonuse versus posttest PP use	Patient portal use is associated with improved glycemic control.	Cohort	☒	—	—	—	☒
Lyles et al [52], 2016	Adults with DM type 2 using statins, registered for PP	Prescription refill use versus no refill use	Requesting prescription refills is associated with improved statin adherence.	Cohort	—	☒	—	—	☒
McCarrier et al [54], 2009	Adults aged <50 years with uncontrolled DM type 1	Nurse-aided PP users versus PP nonusers	Patient portal use results in improved self-efficacy, but not in improved glycemic control.	RCT[g]	☒	☒	—	☒	☒
Price-Haywood and Luo [56], 2017	Adults with DM (or HT[h])	PP users versus PP nonusers	Patient portal use is associated with more primary care visits and telephone encounters, but not with less hospitalizations or ED[i] visits.	Cohort	☒	—	☒	—	☒
Sarkar et al [68], 2014	Adults with DM, registered for PP	Recurrent prescription refill use versus occasional refill use versus no refill use	Recurrent use of prescription refills is associated with improvements in adherence and lipid control.	Cohort	☒	—	—	—	☒
Shimada et al [71], 2016	Veterans with uncontrolled DM, registered for PP	Messaging and prescription refills users versus PP users who use neither	Messaging or requesting prescription refills is associated with improved glycemic control.	Cohort	☒	—	—	—	☒
van Vugt et al [73], 2016	Patients with DM type 2, registered for PHR[j]	PHR+personal coach versus PHR use alone	PHR use does not result in improved glycemic control, self-care, distress, nor well-being, regardless of personal coaching.	RCT	☒	☒	—	☒	☒
Dixon et al [47], 2016	Adults with DM type 2	Pretest PP nonusers versus posttest PP users	Patient portal use is associated with improved adherence, but not with changes in clinical outcomes nor care utilization.	QE	☒	☒	☒	—	☒
Druss et al [77], 2020	Patients with a mental disorder+DM, HT or HC[k]	PP users versus PP nonusers	Patient portal use does not result in clinically relevant improvements in perceived quality of care, patient activation nor HRQoL[l].	RCT	☒	☒	☒	—	☒
Graetz et al [49], 2020	Adults with DM with at least 1 oral drug	PP users versus PP nonusers	Patient portal use is associated with small, likely irrelevant improvements in glycemic control and medication adherence.	Cross	☒	☒	—	—	☒
Grant et al [50], 2008	Adults with DM using medication	Tethered PP use versus untethered PP use	Using a tethered patient portal results in increased patient participation, but not improved glycemic control.	RCT	☒	☒	—	—	☒
Reed et al [60], 2019	Adults with DM+HT, asthma, CAD[m], or CHF[n]	PP users versus PP nonusers	Patient portal use is associated with more outpatient office visits, and with reduced ED visits and preventable hospitalizations.	Cross	☒	—	☒	—	☒
Riippa et al [62], 2014	Adults with DM, HT, or HC	PP users versus PP nonusers	Patient portal use does not result in clinically relevant improvements in patient activation, except among adults with low baseline activation.	RCT	—	☒	—	—	☒

Author, year	Participants	Comparison	Main conclusion	Study design	Clinical	Patient reported	Care utilization	Technology	Quality[b]
Riippa et al [63], 2015	Adults with DM, HT, or HC	PP users versus PP nonusers	Patient portal use does not result in clinically relevant improvement in patient activation nor HRQoL.	RCT	—	☒	☒	☒	☒
Robinsonet al [64], 2020	Veterans with uncontrolled DM type 2, registered for PP	Responders on team-initiated messages versus nonresponders	Responding on messages is associated with improved self-management and self-efficacy.	Cross	—	☒	—	—	☒
Ronda et al [65], 2014	Adults with DM	Recurrent PP users versus PP nonusers	Recurrent patient portal use is associated with better self-efficacy and knowledge.	Cross	—	☒	—	☒	☒
Ronda et al [66], 2015	Adults with DM, registered for PP	Persistent users versus early quitters	Recurrent users believe the patient portal increases disease knowledge, and they find it useful.	Cross	—	☒	—	☒	☒
Sabo et al [67], 2021	Adults with DM type 2, registered for PP	PP users versus PP nonusers	Patient portal use has minor, clinically irrelevant effects on BMI, and no effects on glycemic control nor blood pressure.	RCT	☒	—	—	—	☒
Seo et al [69], 2020	Patients with DM, registered for PHR	Continuous users versus noncontinuous users	Continuous use of a tethered PHR is associated with slightly improved glycemic control. Clinical implications are doubtful.	Cohort	☒	—	—	—	☒
Sharit et al [70], 2018	Overweight veterans with prediabetes	Pretest PP nonuse versus posttest PP use	Using an accelerometer-connected patient portal is associated with improvements in physical activity and blood pressure.	QE	☒	☒	—	☒	☒
Tenforde et al [72], 2012	Adults aged <75 years with DM	PP users versus PP nonusers	Patient portal use is associated with slightly improved diabetes control, lipid profile, and blood pressure. Clinical implications are doubtful.	Cohort	☒	—	☒	—	☒
Vo et al [74], 2019	Adults aged <80 years with DM type 2, registered for PP	Previsit message use versus no previsit message use	Sending previsit prioritization messages does not result in improved glycemic control, but does result in improved perceived shared-decision-making.	RCT	☒	☒	—	—	☒
Zocchi et al [76], 2021	Patients with DM type 2, registered for PP	PP users	Among existing patient portal users with uncontrolled DM or high LDL[o], increased use is associated with improved control.	Cohort	☒	—	—	—	☒
Bailey et al [41], 2019	Adults with DM, on high-risk medication	PP users	Patients are satisfied with the patient portal.	QE	—	—	—	☒	☒
Byczkowski et al [43], 2014	Parents of children with DM (or CF[p] or JIA[q])	PP users	Patients consider the patient portal to be useful in managing and understand their child's disease.	Cross	—	☒	—	☒	☒
Chung et al [44], 2017	Adults with DM, registered for PP	Message users versus message nonusers	Using secure messaging is associated with better glycemic control.	Cohort	☒	☒	☒	—	☒
Conway et al [45], 2019	Patients with DM, registered for PP	PP users	Patients believe the tethered diabetes PHR might improve their diabetes self-care.	Cross	—	☒	—	☒	☒
Devkota et al [46], 2016	Patients with DM type 2	PP users who read and write emails versus PP nonusers	Reading and writing emails is associated with improved glycemic control.	Cohort	☒	☒	—	—	☒

Author, year	Participants	Comparison	Main conclusion	Study design	Clinical	Patient reported	Care utilization	Technology	Quality[b]
Graetz et al [48], 2018	Adults with DM	PP users versus PP nonusers	Patient portal use is associated with improved adherence to medication and preventive care utilization.	Cross	—	☒	☒	—	☒
Martinez et al [53], 2021	Adults with DM type 2 using medication, registered for PP	Pretest PP nonuse versus posttest PP use	Patient portal use results in clinically not relevant improvements in patient activation and self-efficacy. This is related to the very short follow-up period of the study.	QE	—	☒	—	☒	☒
Osborn et al [55], 2013	Adults with DM type 2 using medication	PP users versus PP nonusers	Patient portal use is not associated with improved glycemic control, as compared with nonusers. However, among users, more frequent use is associated with improved glycemic control.	Cross	☒	—	—	—	☒
Price-Haywood et al [57], 2018	Adults with DM (or HT)	PP users versus PP nonusers	Messaging is associated with improved glycemic control.	Cohort	☒	—	—	—	☒
Quinn et al [58], 2018	Adults aged <65 years with DM type 2	PP+extra module users versus PP users	Messaging is associated with better glycemic control. Note: glycemic parameters were predicted and not represent measurements.	RCT	☒				☒
Reed et al [59], 2015	Adults with DM, HT, asthma, CAD, or CHF, registered for PP	PP users	One-third of patients report that messaging in a patient portal results in less health care visits and improved overall health.	Cross	—	☒	☒	—	☒
Reed et al [61], 2019	Adults with DM, asthma, HT, CAD, CHF, or CV[r] event risk	PP users versus PP nonusers	One-third of patients report that using the patient portal improves overall health.	Cross	—	☒	—	☒	☒
Wald et al [75], 2009	Patients with DM type 2	PHR users who created a previsit plan	Users who create a previsit care plan feel better prepared for visits.	RCT	—	☒	—	☒	☒

[a]Studies are listed multiple times in Tables 10-13. Per disease category, the relevant subconclusion and health outcomes are described. Associations with health outcomes are color-coded as green for beneficial, yellow for neutral or clinically nonrelevant, or red for undesired. The half green and half yellow symbol implies that one study investigated multiple outcomes in one category and reported beneficial associations for some outcomes and neutral associations for others.

[b]Quality appraisal—green: high quality; yellow: medium quality; red: low quality.

[c]DM: diabetes mellitus.

[d]PP: patient portal.

[e]QE: quasi-experimental, including pretest-posttest studies and feasibility studies.

[f]The study did not assess any health outcome in a certain category.

[g]RCT: randomized controlled trial.

[h]HT: hypertension.

[i]ED: emergency department.

[j]PHR: personal health record.

[k]HC: hypercholesteremia.

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

[m]CAD: coronary artery disease.

[n]CHF: congestive heart failure.

[o]LDL: low-density lipoprotein.

[p]CF: cystic fibrosis.

[q]JIA: juvenile idiopathic arthritis.

[r]CV: cardiovascular.

Table 11. Conclusions and health outcomes: studies investigating cardiopulmonary diseases (n=21), of which 6 (29%) are of high methodological quality.[a]

Author, year	Participants	Comparison	Conclusion	Study design	Clinical	Patient reported	Care utilization	Technology	Quality[b]
Ahmed et al [79], 2016	Adults with asthma using medication	PP[c] users versus PP nonusers	Patient portal use does not result in durable improvements in HRQoL[d] nor asthma control.	RCT[e]	☒	☒	☒	☒	☒
Fiks et al [81], 2015	Children aged 6-12 years with asthma	PP users versus PP nonusers	Patient portal use results in improved asthma control.	RCT	☒	☒	☒	☒	☒
Lau et al [85], 2015	Adults with asthma	PHR[f] users versus PHR nonusers	PHR use does not increase the use of asthma action plans, and does not affect asthma control, health care utilization nor work or school participation.	RCT	☒	☒	☒	—[g]	☒
Manard et al [86], 2016	Adults with uncontrolled HT[h]	PP users versus PP nonusers	Using a patient portal linked with a blood pressure cuff is not associated with improved blood pressure control.	Cohort	☒	—	—	—	☒
Price-Haywood and Luo [56], 2017	Adults with HT (or DM[i])	PP users versus PP nonusers	Patient portal use is associated with more primary care visits and telephone encounters, but not hospitalizations or ED[j] visits. Effects on blood pressure control are not clinically relevant.	Cohort	☒	—	☒	—	☒
Shimada et al [71], 2016	Veterans with uncontrolled HC[k] or HT, registered for PP	Users of both messaging and prescription refills versus nonusers	Messaging or requesting prescription refills are both associated with improved lipid control. Requesting prescription refills is associated with improved blood pressure control.	Cohort	☒	—	—	—	☒
Apter et al [80], 2019	Adults with asthma using prednisone	PP use+training versus PP use+assistance via home visits	Patient portal use results in minor improvements in asthma control and HRQoL. Conducting home visits results in more improvements in these outcomes.	RCT	☒	☒	☒	—	☒
Druss et al [77], 2020	Patients with a mental disorder+DM[i], HT[j], or HC[k]	PP users versus PP nonusers	Patient portal use does not result in clinically relevant improvements in perceived quality of care, patient activation, nor HRQoL.	RCT	☒	☒	☒	—	☒
Fiks et al [82], 2016	Children aged 6-12 years with asthma	PP users versus PP nonusers	Patient portal use is associated with improved treatment adherence. Among patients with uncontrolled asthma, its use is associated with more care visits. Adoption is low.	QE[l]	—	☒	☒	☒	☒
Martinez Nicolás et al [89], 2019	Patients with COPD[m] or CHF[n]	Pretest PP nonuse versus posttest PP use	Patient portal use is associated with less hospitalizations, readmissions, and ED visits among patients with CHF and COPD.	QE	☒	—	☒	—	☒
Reed et al [60], 2019	Adults with DM+HT, asthma, CAD[m], or CHF[n]	PP users versus PP nonusers	Patient portal use is associated with more outpatient office visits, and with reduced ED visits and preventable hospitalizations.	Cross	☒	—	☒	—	☒
Riippa et al [62], 2014	Adults with DM, HT, or HC	PP users versus PP nonusers	Patient portal use does not result in clinically relevant improvements in patient activation, except for patients with low baseline activation.	RCT	—	☒	—	—	☒
Riippa et al [63], 2015	Adults with DM, HT, or HC	Patient portal versus usual care	Patient portal use does not result in clinically relevant improvement in patient activation nor HRQoL.	RCT	—	☒	—	☒	☒

Author, year	Participants	Comparison	Conclusion	Study design	Clinical	Patient reported	Care utilization	Technology	Quality[b]
Toscos et al [87], 2020	Patients with nonvalvular AF[o] with an oral anticoagulant drug	PP users versus PP nonusers	Using a patient portal connected to a Smart Pill Bottle does not result in improved drug adherence.	RCT	—	☒	—	—	☒
Wagner et al [88], 2012	Patients with HT	PHR users versus PHR nonusers	Using a tethered PHR does not result in clinically relevant improvements in blood pressure control, patient activation nor health care utilization. Adoption is low.	RCT	☒	☒	☒	☒	☒
Aberger et al [78], 2014	Postrenal transplant patients with HT	PP users	Using a patient portal–linked blood pressure monitoring system is associated with improved blood pressure control.	QE	☒	—	—	—	☒
Kim et al [84], 2019	Patients with obstructive sleep apnea	PHR+activity tracker versus PHR alone versus nonusers	Using a tethered PHR results in more weight loss, regardless of its connection to an activity tracker. No sleep-related outcome improvements are seen.	RCT	☒	—	—	☒	☒
Kogut et al [83], 2014	Adults aged >49 years with cardiopulmonary disorders	PHR users versus PHR nonusers	Pharmacists reviewing patient-reported medication lists in a PHR might identify more medication-related problems.	QE	☒	—	—	—	☒
Price-Haywood et al [57], 2018	Adults with HT or DM	PP users versus PP nonusers	Messaging is not associated with improved blood pressure control.	Cohort	☒	—	—	—	☒
Reed et al [59], 2015	Adults with DM, HT, asthma, CAD[p], or CHF, registered for PP	PP users	One-third of patients report that messaging in a patient portal results in less health care visits and improved overall health.	Cross-sectional	—	☒	☒	—	☒
Reed et al [61], 2019	Adults with DM, asthma, HT, CAD, CHF, or CV[q] event risk	PP users versus PP nonusers	A third of patients reports that using the patient portal improves overall health.	Cross-sectional	—	☒	—	☒	☒

[a]Studies are listed multiple times in Tables 10-13. Per disease category, the relevant subconclusion and health outcomes are described.

[b]For color coding of quality appraisal and health outcomes, see Table 10.

[c]PP: patient portal.

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

[e]RCT: randomized controlled trial.

[f]PHR: personal health record.

[g]The study did not assess any health outcome in a certain category.

[h]HT: hypertension.

[i]DM: diabetes mellitus.

[j]ED: emergency department.

[k]HC: hypercholesteremia.

[l]QE: quasi-experimental, including pilot or feasibility studies.

[m]COPD: chronic obstructive pulmonary disease.

[n]CHF: Congestive heart failure.

[o]AF: atrial fibrillation.

[p]CAD: coronary artery disease.

[q]CV: cardiovascular.

Table 12. Conclusions and health outcomes: studies investigating hematological and oncological diseases (n=14), of which 2 are of high methodological quality (14%).[a]

Author, year	Participants	Comparison	Conclusion	Study design	Clinical	Patient reported	Care utilization	Technology	Quality[b]
Cahill et al [90], 2014	Adults with a brain tumor	PHR[c] users versus PHR nonusers	Using a tethered PHR is associated with improvements in patient uncertainty.	Cross-sectional	—	☒	—[d]	—	☒
Coquet et al [93], 2020	Patients with cancer+chemotherapy, registered for PP[e]	Email users versus email nonusers	Sending emails is associated with improved 2-year survival, less missed appointments, and less hospitalizations.	Cohort	☒	—	☒	—	☒
Chiche et al [91], 2012	Adults with ITP[f]	PP users versus PP nonusers	Patient portal use does not result in improved HRQoL[g]. The portal is acceptable and feasible.	RCT[h]	—	☒	—	☒	☒
Groen et al [94], 2017	Patients with lung cancer	PP users	Patient portal use does not affect HRQoL nor patient engagement. It is feasible and acceptable.	QE[i]	—	☒	—	☒	☒
Hall et al [95], 2014	Patients with cancer resection	PP users	Disclosing results of genetic cancer screening in a patient portal might be feasible and acceptable, and is not associated with more anxiety. Yet, few abnormal results were observed.	QE	—	☒	—	☒	☒
Kidwell et al [97], 2019	Patients aged 13-24 years with sickle cell disease	PP users	Patient portal use is not associated with improved medical decision-making by patients. It is acceptable and easy to use.	QE	—	☒	—	☒	☒
Martinez Nicolás et al [89], 2019	Patients with hematologic malignancy	Pretest PP nonuse versus posttest PP use	Patient portal use is not associated with less hospitalizations, readmissions, nor ED[j] department visits.	QE	☒	—	☒	—	☒
Williamson et al [102], 2017	Pediatric cancer survivors	PHR users versus PHR registrants	Patient portal use is not associated with less missed appointments.	Cohort	—	—	☒	☒	☒
Collins et al [92], 2003	Patients with hemophilia >11 years	Users	An electronic treatment log is considered feasible and easy to use.	QE	—	—	—	☒	☒
Hong et al [96], 2016	Children aged 13-17 years with cancer or a blood disorder+parents	PP users	A small cohort considers a patient portal to be feasible and useful.	Cross-sectional	—	☒	—	☒	☒
O'Hea et al [98], 2021	Women with breast cancer	PP users versus PP nonusers	Patient portal use does not result in improved HRQoL nor disease knowledge.	RCT	—	☒	—	—	☒
Pai et al [99], 2013	Men with prostate cancer	PHR users	Patients are satisfied with a tethered PHR and find it increases disease knowledge.	Cross-sectional	—	☒	—	☒	☒
Tarver et al [100], 2019	Patients with colorectal cancer	Tethered PHR users	Patients are satisfied with an integrated care plan and find it useful.	Cohort	—	—	—	☒	☒
Wiljer et al [101], 2010	Patients with breast cancer	Pretest PHR nonusers versus posttest PHR users	PHR use is not associated with improved self-efficacy, nor with a clinically relevant decrease in anxiety. Satisfaction is high.	QE	—	☒	—	☒	☒

[a]Studies are listed multiple times in Tables 10-13. Per disease category, the relevant subconclusion and health outcomes are described.

[b]For color coding of quality appraisal# and health outcomes, see Table 10.

[c]PHR: personal health record.

[d]The study did not assess any health outcome in a certain category.

[e]PP: patient portal.

[f]ITP: idiopathic thrombocytopenic purpura.
[g]HRQoL: health-related quality of life.
[h]RCT: randomized controlled trial.
[i]QE: quasi-experimental, including pilot or feasibility studies.
[j]ED: emergency department.

Table 13. Conclusions and health outcomes: studies investigating other diseases (n=21), of which 2 (10%) are of high methodological quality.[a]

Author, year	Participants	Comparison	Conclusion	Study design	Clinical	Patient reported	Care utilization	Technology	Quality[b]
Miller et al [113], 2011	Patients with multiple sclerosis	PHR[c] use versus PHR that only enables messaging	Using an untethered PHR results in slightly improved HRQoL[d], but not in improved self-efficacy, disease control nor health care utilization.	RCT[e]	—	☒	☒	—[f]	☒
Navaneethan et al [114], 2017	Adults with chronic kidney disease	PP[g] users+coach versus PP users versus PP nonusers	Patient portal use, regardless of added training, does not result in improved kidney function, nor altered health care utilization.	RCT	☒	—	☒	—	☒
Anand et al [103], 2017	MSM[h] and transgender women with HIV	PP users	The patient portal is feasible and acceptable.	RCT	—	—	—	☒	☒
Druss et al [106], 2014	Patients with a mental disorder+chronic condition	PP users versus PP nonusers	Patient portal use results in increased use of preventive health services and medical visits, but not in improved HRQoL.	RCT	☒	☒	☒	—	☒
Druss et al [77], 2020	Patients with a mental disorder+DM[i], HT[j], or HC[k]	PP users versus PP nonusers	Patient portal use does not result in clinically relevant improvements in perceived quality of care, patient activation, nor HRQoL.	RCT	☒	☒	☒	—	☒
Jhamb et al [107], 2015	Adults visiting nephrology clinics	PP users versus PP nonusers	Patient portal use might be associated with improved blood pressure control, although its clinical relevance is unclear.	Cross-sectional	☒	—	—	☒	☒
Keith McInnes et al [109], 2013	Veterans with HIV	PP users versus PP nonusers	Patient portal use is associated with improved adherence to HIV medication.	Cross-sectional	—	☒	—	☒	☒
Keith McInnes et al [110], 2017	Veterans with HIV+detectable viral load, registered for PP	Messaging or prescription refill users versus nonusers	Requesting prescription refills is associated with improved HIV control, but messaging is not.	Cohort	☒	—	—	—	☒
Kiberd et al [111], 2018	Adult with home dialysis	Pretest PP nonuse versus posttest PP use	Patient portal use is not associated with improvements in HRQoL nor perceived quality of care. Both were already high at baseline.	QE[l]	—	☒	—	☒	☒
Lee et al [112], 2017	Patients with cleft lip or cleft palate surgery	PP users versus PP tailored for lip or cleft palate surgery	Using a tailored, disease-specific patient portal is associated with increased disease knowledge.	QE	—	☒	—	☒	☒
Reich et al [116], 2019	Patients with inflammatory bowel disease	PP users versus PP nonusers	Patient portal use does not result in improved HRQoL, but results in a higher vaccination rate. Patient satisfaction is high.	RCT	—	☒	☒	☒	☒
Scott Nielsen et al [117], 2012	Patients with multiple sclerosis	PP users versus PP nonusers	Messaging in a patient portal is associated with more clinic visits, but not with less ED[m] visits nor hospitalizations.	Cross-sectional	—	—	☒	☒	☒
Tom et al [119], 2012	Parents of children age <6 years with 1 ore more chronic condition(s)	PP users versus PP nonusers	Patient portal use is not associated with improved access to care, nor perceived quality of care. It is considered feasible.	Cross-sectional	—	☒	—	☒	☒
van den Heuvel et al [120], 2018	Adults with bipolar disorder	Pretest PHR nonusers versus posttest PHR users	PHR use is not associated with improved HRQoL, patient empowerment, symptom reduction, nor disease burden.	Cross-sectional	—	☒	—	☒	☒

Author, year	Participants	Comparison	Conclusion	Study design	Clinical	Patient reported	Care utilization	Technology	Quality[b]
van der Vaart et al [121], 2014	Patients with rheumatoid arthritis	Pretest PP nonusers versus posttest PP users	Patient portal use is not associated with improved patient empowerment. It is considered useful and understandable.	Cross-sectional	—	☒	—	☒	☒
Bidmead et al [104], 2016	Patients with inflammatory bowel disease	PHR users	PHR use is not associated with improved self-management.	Cross-sectional	☒	☒	☒	☒	☒
Byczkowski et al [43], 2014	Parents of children with CF[o] or JIA[p] (or DM)	PP users	Patients consider the patient portal to be useful in managing and understand their child's disease.	Cross	—	☒	—	☒	☒
Crouch et al [105], 2015	Veterans with HIV	PP users versus PP nonusers	Patient portal use is associated with improved patient activation, disease knowledge, HIV load, but not with improved CD4-count nor treatment adherence	Cross-sectional	—	☒	—	☒	☒
Kahn et al [108], 2010	Patients with HIV or aids	PP users	Patients are satisfied with the patient portal and consider it to be helpful in managing their problems.	QE	—	☒	—	☒	☒
Plimpton [115], 2020	Women with HIV	Pretest PP nonuse versus posttest PP use	Patient portal use is associated with an increase in planned visits, but not with a decrease in missed visits. A trend toward improved viral load is seen.	QE	☒	—	☒	—	☒
Son et al [118], 2019	Patients aged >49 years with 1 or more chronic condition(s)	PP users	Patients consider a patient portal to be helpful in increasing self-management.	Cohort	—	☒	—	☒	☒

[a]Studies are listed multiple times in Tables 10-13. Per disease category, the relevant subconclusion and health outcomes are described.

[b]For color coding of quality appraisal and health outcomes, see Table 10.

[c]PHR: personal health record.

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

[e]RCT: randomized controlled trial.

[f]The study did not assess any health outcome in a certain category.

[g]PP: patient portal.

[h]MSM: men who have sex with men.

[i]DM: diabetes mellitus.

[j]HT: hypertension.

[k]HC: hypercholesteremia.

[l]QE: quasi-experimental, including pilot or feasibility studies.

[m]ED: emergency department.

[o]CF: cystic fibrosis.

[p]JIA: juvenile idiopathic arthritis.

Clinical Outcomes

In 44 studies investigating a total of 69 clinical outcomes, a beneficial association with digital health record use was reported for 42% (29/69) of the outcomes. Hospitalizations and exacerbations were the most frequently studied disease events and complications, with beneficial effects reported in half of the studies (2/4 and 2/4, respectively). Blood pressure was the most frequently studied vital parameter, with beneficial effects reported in 36% (5/14) of the studies. HbA_{1c} and cholesterol levels were the most frequently studied laboratory parameters, with beneficial effects reported in 53% (10/19) and 57% (4/7) of the studies, respectively. No clinical outcomes were unfavorably affected by patient-centered digital health record use. In comparison with the total population, higher proportions of beneficial effects were reported for diabetes mellitus and cardiopulmonary diseases. When focusing on 14 high-quality studies, beneficial effects were observed less frequently, in only 30% (7/23) of the clinical outcomes.

Studies that assessed vital parameters generally reported few other health outcomes. However, among the studies that assessed disease events and complications, and laboratory parameters, beneficial effects were often associated with improved treatment adherence [52,68,71,81]. We hypothesize that this might be related to the removal of logistical barriers for patients in

obtaining web-based prescription refills, as opposed to having to call health care providers or send them an email. Of the 6 high-quality studies that investigated treatment adherence, 2 studies assessed patient-centered digital health records that enabled patients to request prescription refills and found beneficial effects on adherence [52,68].

Patient-Reported Outcomes

Overall, in 53 studies investigating a total of 86 patient-reported outcomes, a beneficial association with digital health record use was reported for 45% (39/86) of the outcomes. Of the 18 studies investigating 19 self-management or self-efficacy outcomes, beneficial effects were reported in 53% (9/19). Of these 9 studies, 56% (5/9) used validated questionnaires. For patient engagement outcomes, large differences in the proportions of beneficial effects were observed: from 11% (1/9) for patient activation, to 56% (5/9) for patient involvement, and 70% (7/10) for disease knowledge. However, only in measuring patient activation, validated questionnaires were principally used (8/9, 88% of studies). For HRQoL, beneficial effects were reported in 27% (4/15) of the studies, of which half used validated HRQoL questionnaires. No patient-reported outcomes were unfavorably affected by patient-centered digital health record use. In comparison to the total population, higher proportions of beneficial effects were reported for diabetes mellitus, especially for patient engagement and treatment adherence. Lowest proportions were reported for cardiopulmonary diseases, especially for patient engagement. When focusing on 10 high-quality studies, a lower proportion (7/19, 37%) of beneficial effects was observed.

We observed that improvements in patient engagement were especially facilitated by strengthening patient-professional communication; for example, through secure messaging [71,81,93]. In addition, both self-efficacy and HRQoL primarily seemed to be reinforced through the use of 2 functionalities: patient-professional communication [54,90,113] and information on disease progression [90,113].

Health Care Utilization

For 24 studies investigating a total of 27 health care utilization outcomes, a beneficial association with digital health record use was observed for 59% (16/27) of the outcomes. The highest proportion (10/13, 77%) of beneficial effects was reported for an increased use of recommended care services. Of these 13 studies, 5 (38%) focused on recommended care services for people with uncontrolled disease, 4 (31%) on the use of preventive care services, and 4 (31%) on medical follow-up rates. In 25% (3/12) of the studies that assessed reductions in ED visits and hospitalizations, these were accompanied by an increased use of other care services, including outpatient clinic appointments and secure messaging. Compared with the total population, highest proportions of beneficial effects were reported for diabetes mellitus and hematological and oncological diseases. When focusing on 7 high-quality studies, lower proportions (3/9, 33%) of beneficial effects were observed.

Technology-Related Outcomes

For 39 studies investigating a total of 75 technology-related outcomes, a beneficial association with digital health record use was observed for 88% (66/75) of the outcomes. All (22/22, 100%) studies reported high patient satisfaction with accessing and using digital health records. Furthermore, 75% (6/8) of the studies reported high patient satisfaction with the effects of using digital health records. High feasibility was reported by 79% (15/19) of the studies, and high acceptability by 88% (23/26) of the studies. Highest feasibility was reported for digital health records intended for people with hematological and oncological diseases. Lowest feasibility and acceptability were reported for digital health records intended for people with cardiopulmonary diseases. When focusing on 6 high-quality studies, proportions of studies that found beneficial effects were similar.

High Disease Burden or Self-management

A subgroup of 47 studies that investigated patients with a high disease burden or high self-management was assessed. The following conditions were included: malignancies (11 studies), asthma (9 studies), HIV infection and AIDS (6 studies), hematologic conditions (5 studies), chronic kidney disease (3 studies), chronic heart failure (4 studies), mental disorders (3 studies), multiple sclerosis (2 studies), inflammatory bowel disease (2 studies), rheumatologic conditions (2 studies), insulin-dependent diabetes mellitus (2 studies), atrial fibrillation (1 study), cystic fibrosis (1 study), and posttransplant patients (1 study). In general, the digital health records assessed in this subgroup were more often tailored to specific patient populations through the addition of specialized functionalities or connected wearables.

In comparison with studies investigating patients with no high disease burden, studies investigating patients with a high disease burden reported considerably higher proportions of beneficial effects for vital parameters, patient engagement, reductions in ED visits and hospitalizations, and for all technology-related outcomes. Considerably lower proportions of beneficial effects were reported for laboratory parameters, health-related quality of life, treatment adherence, and increased use of recommended care services. For the 9 high methodological quality studies on high disease burden or self-management, the proportions of studies that found beneficial effects were roughly similar.

Focus on Passive Versus Active Features

Of the 81 studies, 41 (51%) of the studied patient-centered digital health records focused on passive features and 40 (49%) focused on active features. In comparison with digital health records with an active focus, more beneficial effects were observed among digital health records with a passive focus for laboratory parameters (9/16, 56% vs 7/17, 41%), self-management and self-efficacy (7/11, 64% vs 3/8, 38%), patient engagement (9/15, 60% vs 4/13, 31%), and for an increased use of recommended care services (5/6, 83% vs 5/7, 71%). Compared with digital health records with a passive focus, more beneficial effects were observed among digital health records with an active focus on disease events or complications (4/10, 40% vs 1/5, 20%) and reductions in ED visits and hospitalizations (4/6, 67% vs 1/6, 17%). However, when focusing on high-quality studies, higher proportions of beneficial effects were seen for digital health records with an active focus on all clinical outcomes, patient-reported outcomes,

reductions in ED visits and hospitalizations, patient satisfaction, and acceptability.

Quality Appraisal

Of the 81 included studies, 27 (33%) studies were graded as low quality, 38 (47%) as medium quality, and 16 (20%) as high quality (Tables 10-13). Studies investigating cardiopulmonary conditions were of the highest quality, with 29% (6/21) of the studies graded as high quality. Of the 24 included RCTs, 7 (29%) were of high quality. Only 38% (9/24) of the RCTs concealed allocation to treatment groups, and 67% (16/24) used intention-to-treat analyses. Of the 57 studies with other designs, 9 (16%) were graded as high quality. Overall, 15% (12/81) of studies reported power calculations.

Among the 65 studies that were graded as medium or low quality, only 35% (23/65) used reliable or validated tools for the measurement of all their outcomes and 48% (31/65) for part of their outcomes. Of these 65 studies, 10 (15%) studies took adequate measures to limit selection bias and 17 (26%) studies used a control group or randomized participants.

When focusing on the 16 high-quality studies, 3 functionalities appeared to be the most effective: secure messaging to lower barriers in patient-professional interaction, prescription refill functions to improve medication adherence, and information provision on disease progression. In addition, in 16 high-quality studies, the proportions of beneficial effects were similar for a subgroup of studies that included older participants (mean age >55 years), which included a high number of female participants (>45%), or included a racially diverse population (<50% White participants), as compared with the total population.

Discussion

Principal Findings

In this systematic review, we evaluated evidence on the effects of the use of patient-centered digital health records in nonhospitalized patients with chronic health conditions on clinical and patient-reported outcomes, health care utilization, and technology-related outcomes. Beneficial effects were most frequently reported for the use of recommended care services (10/13, 77%) and for 4 patient-reported outcomes: disease knowledge (7/10, 70%), patient involvement (5/9, 56%), treatment adherence (10/18, 56%), and self-management and self-efficacy (10/19, 53%). Regarding clinical outcomes, beneficial effects were reported in 42% (29/69) of the studies. Beneficial effects were least frequently reported for disease events and complications (5/15, 33%) and health-related quality of life (4/15, 27%). For digital health records that predominantly focused on active features, higher proportions of beneficial effects on nearly all health outcomes were observed among the high-quality studies.

In this study, we observed that patient-centered digital health record use may be associated with an increased use of recommended care services. Beneficial effects on ED visits and hospitalizations were mainly observed when accompanied by an increased rate of follow-up appointments or secure messaging [60,89,93]. This might imply that reducing ED visits and hospitalizations is primarily achieved by facilitating patient-professional communication.

Beneficial effects were most often reported for patients with diabetes or cardiopulmonary disorders. We suggest 2 explanations. First, the focus of digital health records has been directed toward patients with diabetes and asthma for some time because of the sheer number of people with these conditions. This could have resulted in higher-quality patient-centered digital health records and patients who were more accustomed to their use. Second, the relative improvements in health outcomes might be smaller among patients with a condition with a high disease burden because of a higher baseline level of self-management skills and disease knowledge.

The proportions of beneficial effects varied considerably between health outcomes, which may be explained by 2 reasons. First, outcomes with a higher proportion of beneficial effects were more often the primary study outcomes than the secondary outcomes. Digital health records were more frequently tailored for these outcomes, yielding higher beneficial effects. Second, outcome assessment was generally less robust for outcomes with a higher proportion of beneficial effects, such as self-management and patient engagement, which might have resulted in more false-positive effects.

Comparison With Earlier Evidence

Our results are more positive than those of the previous systematic reviews. This might be because of the increasing acceptance of digital health records, their improving quality, the increasing body of literature, or variations in digital health record definitions used. Two previous reviews found mixed effects on the use of portals on health outcomes and health care utilization [27] and reported positive effects on qualitatively assessed self-management in only one-third of the studies [25]. A recent systematic review that focused on portals intended for hospitalized patients found mixed results for patient engagement [26]. A systematic review that included only qualitative studies found that portal use was associated with positive effects on self-efficacy, treatment adherence, and disease knowledge [28]. In a review on eHealth interventions that aim to promote medication use, a weak association between digital health record use and health-related quality of life was observed [10]. This implies that digital health record engagement is not yet sufficient to affect patients' overall health-related quality of life.

Strengths and Limitations

This systematic review has several strengths. Our search strategy was comprehensive, to account for the lack of consensus in digital health record terminology. In addition, a wide variety of health outcomes were considered relevant to determine the impact of digital health record use. However, several limitations of this study must be considered. First, comparisons between studies were difficult because of the variety in evaluated functionalities. A similar diversity was observed among the reported follow-up durations, participants' ages, study sample sizes, and outcomes. Second, because it was not possible to perform a meta-analysis owing to the heterogeneity in reported (disease-specific) outcome measurements and effects, we used the vote-counting method. Therefore, we could not report the

effect estimates and indicated directions of effects [122]. Third, owing to a lack of agreement on feasibility and acceptability thresholds, much is left to the authors' discretion. Fourth, JBI critical appraisal tools rank every item equally despite being not equally important. Finally, publication bias could have resulted in overestimation of the positive effects of patient-centered digital health records. More studies with positive results have been published. In addition, many of the included studies assessed more "mature" patient-centered digital health records, which could have overestimated the effects.

We observed that high patient satisfaction rates did not fully reflect in other health outcomes. This can be partly attributed to acquiescence bias and satisficing [123]. Moreover, satisfaction was often reduced to a narrow ease-of-use questionnaire, instead of satisfaction with the contribution to overall disease management. Finally, several studies only included recurrent users in their analyses, which could falsely increase feasibility. Moreover, these recurrent users likely experienced positive effects of using digital health records, which would have resulted in an overestimation of effects in randomized studies with no intention-to-treat analysis and in all nonrandomized studies.

The voluntary adoption of patient-centered digital health records by patients might reflect an intrinsic, preexisting motivation for self-management and care engagement bias, which may overestimate their effects. Patient-centered digital health record use could even be considered a surrogate measure for engagement [109,124,125]. Thus, it might be best to consider digital health records as vehicles for empowerment, strengthening existing self-management capabilities [126,127].

The effects of using patient-centered digital health records on health outcomes are not always direct but often depend on intermediate steps. For example, requesting prescription refills might depend on the actions performed by (slow-responding) physicians, nurses, or pharmacies. Thus, if using a digital health record would have no observable effects on health outcomes, this could also be a result of these intermediate steps or unforeseen processes and may not be attributable to the use of the patient-centered digital health record.

The proportion of beneficial effects reported in high-quality studies was lower as compared with all included studies for clinical outcomes (30% vs 42%), patient-reported outcomes (37% vs 45%), and health care utilization (33% vs 59%). Nevertheless, the proportions are clinically relevant and promising considering this newly emerging field. The observed differences might be related to 4 factors. First, the selection of motivated, well-educated, digitally minded participants might have overestimated the results in most low- and moderate-quality studies. Second, most studies did not measure ongoing user activity, and assumed that registered users became recurrent users. Third, nearly all low- and moderate-quality studies reported high dropout rates, which could overestimate acceptance rates. Finally, the lack of consensus on digital health record terminology hindered the interpretation of findings. We would advocate the use of uniform definitions, such as those presented in Textbox 1 [10,17-20].

Future Research

Future studies should adopt additional measures to adhere to a uniform taxonomy, use log data, and limit selection bias. The exclusion of less-engaged people could further expand the digital divide between patients who are digitally proficient and those who are not, resulting in an increasingly unequal distribution of care services. We suggest that researchers include a diverse population based on age, gender, disease burden, race, education level, and health literacy [128]. Finally, further research should focus on determining which functionalities are mostly responsible for the effects on the outcomes.

Conclusions

The use of patient-centered digital health records in chronic conditions is potentially associated with beneficial effects on several patient-reported outcomes and recommended care services in a considerable number of studied digital health records. The rates of the effects were approximately similar for different patient groups. Feasibility and acceptability were high. Our findings support further implementation of patient-centered digital health records in clinical practice. Yet, higher-quality research is needed to identify effects per disease category and per health outcome and to learn which patients might benefit from specific functionalities.

Acknowledgments

This project was supported by grants K23HL150232 (principal investigator [PI]: SMB) and K23HL141447 (PI: RMC) from the National Heart, Lung, and Blood Institute of the National Institutes of Health and by grant NWA.1160.18.038 (PI: MH Cnossen, WP leader SCG) of the Netherlands Organization for Scientific Research (NWO) for the SYMPHONY consortium. The content is solely the responsibility of the authors and does not necessarily represent National Institutes of Health or NWO.

Authors' Contributions

All authors were involved in the design of the research protocol. MB performed the search strategy. MRB and SMB assessed all titles, abstracts, and full texts for eligibility. SCG helped resolve the discussion if necessary. MRB and SMB performed data extraction and synthesis. All authors provided feedback on the manuscript and approved the final manuscript.

Conflicts of Interest

SCG received an unrestricted medical research grant from Sobi. The authors have no futher interests to declare.

Multimedia Appendix 1
The search strategy used, that includes terms related to the search categories: “patient,” “intervention,” and “outcome.”.
[DOCX File , 19 KB - jmir_v24i12e43086_app1.docx]

Multimedia Appendix 2
The 4 modified Joanna Briggs Institute critical appraisal tools used in this study.
[DOCX File , 24 KB - jmir_v24i12e43086_app2.docx]

Multimedia Appendix 3
Proportion of beneficial effects reported per health outcome of all studies, presented per disease category.
[DOCX File , 18 KB - jmir_v24i12e43086_app3.docx]

Multimedia Appendix 4
The proportion of beneficial effects reported per health outcome of 16 high quality studies, presented per disease category.
[DOCX File , 26 KB - jmir_v24i12e43086_app4.docx]

References

1. Noncommunicable diseases. World Health Organization. 2021 Sep 16. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2022-12-01]
2. Benziger CP, Roth GA, Moran AE. The global burden of disease study and the preventable burden of NCD. Glob Heart 2016 Dec;11(4):393-397 [FREE Full text] [doi: 10.1016/j.gheart.2016.10.024] [Medline: 27938824]
3. Buttorff C, Ruder T, Bauman M. Multiple Chronic Conditions in the United States. Santa Monica, CA, USA: RAND Corporation; 2017.
4. Schulman-Green D, Jaser S, Martin F, Alonzo A, Grey M, McCorkle R, et al. Processes of self-management in chronic illness. J Nurs Scholarsh 2012 Jun;44(2):136-144 [FREE Full text] [doi: 10.1111/j.1547-5069.2012.01444.x] [Medline: 22551013]
5. Chen S, Kuhn M, Prettner K, Bloom DE. The macroeconomic burden of noncommunicable diseases in the United States: estimates and projections. PLoS One 2018 Nov 1;13(11):e0206702 [FREE Full text] [doi: 10.1371/journal.pone.0206702] [Medline: 30383802]
6. The Complexities of Physician Supply and Demand: Projections From 2019 to 2034. Association of American Medical Colleges. 2021 Jun. URL: https://tinyurl.com/yfren872 [accessed 2022-09-01]
7. European Commission, Directorate-General for Employment, Social Affairs and Inclusion, McGrath J. Analysis of Shortage and Surplus Occupations 2020. Luxembourg, Luxembourg: Publications Office of the European Union; 2020.
8. Medical staffing in England: a defining moment for doctors and patients. British Medical Association. 2021 Jul 11. URL: https://www.pslhub.org/learn/improving-patient-safety/workforce-and-resources/safe-staffing-levels/medical-staffing-in-england-a-defining-moment-for-doctors-and-patients-bma-11-july-2021-r4856/ [accessed 2022-09-01]
9. COVID-19 continues to disrupt essential health services in 90% of countries. World Health Organization. 2021 Apr 23. URL: https://www.who.int/news/item/23-04-2021-covid-19-continues-to-disrupt-essential-health-services-in-90-of-countries#:~:text=The%20second%20round%20of%20a,since%20the%20first%20survey%20conducted [accessed 2022-09-01]
10. Lancaster K, Abuzour A, Khaira M, Mathers A, Chan A, Bui V, et al. The use and effects of electronic health tools for patient self-monitoring and reporting of outcomes following medication use: systematic review. J Med Internet Res 2018 Dec 18;20(12):e294 [FREE Full text] [doi: 10.2196/jmir.9284] [Medline: 30563822]
11. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
12. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 8;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
13. 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]
14. Mc Namara KP, Versace VL, Marriott JL, Dunbar JA. Patient engagement strategies used for hypertension and their influence on self-management attributes. Fam Pract 2014 Aug;31(4):437-444. [doi: 10.1093/fampra/cmu026] [Medline: 24895137]
15. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
16. Peters AE, Keeley EC. Patient engagement following acute myocardial infarction and its influence on outcomes. Am J Cardiol 2017 Nov 01;120(9):1467-1471. [doi: 10.1016/j.amjcard.2017.07.037] [Medline: 28843395]

17. What are the differences between EHMRs, EHRs and PHRs? HealthIT.gov. 2019. URL: https://www.healthit.gov/faq/what-are-differences-between-electronic-medical-records-electronic-health-records-and-personal [accessed 2022-12-01]
18. A HIMSS Guide to Participating in a Health Information Exchange: HIMSS Healthcare Information Exchange HIE Guide Work Group White Paper. Healthcare Information and Management Systems Society (HIMSS). 2009 Aug. URL: https://assets.hln.com/pdf/HIMSS-HIEGuideWhitePaper.pdf?utm_source=legacy&utm_medium=redirect&utm_campaign=webmig21&utm_term=pdf/HIMSS-HIEGuideWhitePaper.pdf [accessed 2022-09-01]
19. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
20. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Expert Panel. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc 2007;14(1):1-9 [FREE Full text] [doi: 10.1197/jamia.M2273] [Medline: 17077452]
21. Cronin RM, Jimison H, Johnson KB. Personal health informatics. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. Cham, Switzerland: Springer; 2021:363-389.
22. MedMij | Dutch Personal Health Environment. URL: https://medmij.nl/en/home/ [accessed 2022-12-01]
23. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
24. Andrikopoulou E, Scott P, Herrera H, Good A. What are the important design features of personal health records to improve medication adherence for patients with long-term conditions? A systematic literature review. BMJ Open 2019 Sep 26;9(9):e028628 [FREE Full text] [doi: 10.1136/bmjopen-2018-028628] [Medline: 31558449]
25. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
26. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
27. Goldzweig GL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
28. Han HR, Gleason KT, Sun CA, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
29. Brands M, Gouw S, Beestrum M, Fijnvandraat K, Badawy S. Patient-centered digital health interventions and their effects on health outcomes: a systematic review. CRD42020213285. PROSPERO. 2020. URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=213285 [accessed 2022-12-01]
30. 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]
31. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak 2015 Apr 18;15:32 [FREE Full text] [doi: 10.1186/s12911-015-0159-1] [Medline: 25927384]
32. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol 2011 Mar 15;173(6):676-682. [doi: 10.1093/aje/kwq433] [Medline: 21330339]
33. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
34. Vahdat S, Hamzehgardeshi L, Hessam S, Hamzehgardeshi Z. Patient involvement in health care decision making: a review. Iran Red Crescent Med J 2014 Jan;16(1):e12454 [FREE Full text] [doi: 10.5812/ircmj.12454] [Medline: 24719703]
35. Fitzgerald JT, Funnell MM, Hess GE, Barr PA, Anderson RM, Hiss RG, et al. The reliability and validity of a brief diabetes knowledge test. Diabetes Care 1998 May;21(5):706-710. [doi: 10.2337/diacare.21.5.706] [Medline: 9589228]
36. Shmueli A. The SF-36 profile and health-related quality of life: an interpretative analysis. Qual Life Res 1998 Apr;7(3):187-195. [doi: 10.1023/a:1024959508246] [Medline: 9584548]
37. Barroso NE, Hungerford GM, Garcia D, Graziano PA, Bagner DM. Psychometric properties of the Parenting Stress Index-Short Form (PSI-SF) in a high-risk sample of mothers and their infants. Psychol Assess 2016 Oct;28(10):1331-1335 [FREE Full text] [doi: 10.1037/pas0000257] [Medline: 26595220]
38. Adherence to Long-Term Therapies: Evidence for Action. World Health Organization. 2003. URL: http://apps.who.int/iris/bitstream/handle/10665/42682/9241545992.pdf?sequence=1 [accessed 2022-09-01]

39. Data collection form for intervention reviews for RCTs and non-RCTs - template. Version 3. Cochrane Collaboration. 2014. URL: https://dplp.cochrane.org/sites/dplp.cochrane.org/files/public/uploads/CDPLPG%20data%20collection%20form%20for%20intervention%20reviews%20for%20RCTs%20and%20non-RCTs.doc [accessed 2022-09-01]
40. Critical Appraisal Tools. Joanna Briggs Institute. URL: https://jbi.global/critical-appraisal-tools [accessed 2022-09-01]
41. Bailey SC, Wallia A, Wright S, Wismer GA, Infanzon AC, Curtis LM, et al. Electronic health record-based strategy to promote medication adherence among patients with diabetes: longitudinal observational study. J Med Internet Res 2019 Oct 21;21(10):e13499 [FREE Full text] [doi: 10.2196/13499] [Medline: 31638592]
42. Boogerd E, Maas-Van Schaaijk NM, Sas TC, Clement-de Boers A, Smallenbroek M, Nuboer R, et al. Sugarsquare, a Web-based patient portal for parents of a child with type 1 diabetes: multicenter randomized controlled feasibility trial. J Med Internet Res 2017 Aug 22;19(8):e287 [FREE Full text] [doi: 10.2196/jmir.6639] [Medline: 28830853]
43. 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]
44. Chung S, Panattoni L, Chi J, Palaniappan L. Can secure patient-provider messaging improve diabetes care? Diabetes Care 2017 Oct;40(10):1342-1348. [doi: 10.2337/dc17-0140] [Medline: 28807977]
45. Conway NT, Allardice B, Wake DJ, Cunningham SG. User experiences of an electronic personal health record for diabetes. J Diabetes Sci Technol 2019 Jul;13(4):744-750 [FREE Full text] [doi: 10.1177/1932296818818837] [Medline: 30585085]
46. Devkota B, Salas J, Sayavong S, Scherrer JF. Use of an online patient portal and glucose control in primary care patients with diabetes. Popul Health Manag 2016 Apr;19(2):125-131. [doi: 10.1089/pop.2015.0034] [Medline: 26237200]
47. Dixon BE, Alzeer AH, Phillips EO, Marrero DG. Integration of provider, pharmacy, and patient-reported data to improve medication adherence for type 2 diabetes: a controlled before-after pilot study. JMIR Med Inform 2016 Feb 08;4(1):e4 [FREE Full text] [doi: 10.2196/medinform.4739] [Medline: 26858218]
48. Graetz I, Huang J, Brand RJ, Hsu J, Yamin CK, Reed ME. Bridging the digital divide: mobile access to personal health records among patients with diabetes. Am J Manag Care 2018 Jan;24(1):43-48 [FREE Full text] [Medline: 29350505]
49. Graetz I, Huang J, Muelly ER, Fireman B, Hsu J, Reed ME. Association of mobile patient portal access with diabetes medication adherence and glycemic levels among adults with diabetes. JAMA Netw Open 2020 Mar 05;3(2):e1921429 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21429] [Medline: 32074289]
50. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2008 Sep 08;168(16):1776-1782 [FREE Full text] [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]
51. Lau M, Campbell H, Tang T, Thompson DJ, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Mar;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
52. Lyles CR, Sarkar U, Schillinger D, Ralston JD, Allen JY, Nguyen R, et al. Refilling medications through an online patient portal: consistent improvements in adherence across racial/ethnic groups. J Am Med Inform Assoc 2016 Apr;23(e1):e28-e33 [FREE Full text] [doi: 10.1093/jamia/ocv126] [Medline: 26335983]
53. Martinez W, Hackstadt AJ, Hickson GB, Knoerl T, Rosenbloom ST, Wallston KA, et al. The my diabetes care patient portal intervention: usability and pre-post assessment. Appl Clin Inform 2021 May;12(3):539-550 [FREE Full text] [doi: 10.1055/s-0041-1730324] [Medline: 34192774]
54. McCarrier KP, Ralston JD, Hirsch IB, Lewis G, Martin DP, Zimmerman FJ, et al. Web-based collaborative care for type 1 diabetes: a pilot randomized trial. Diabetes Technol Ther 2009 Apr;11(4):211-217 [FREE Full text] [doi: 10.1089/dia.2008.0063] [Medline: 19344195]
55. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013 Jul 03;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
56. Price-Haywood EG, Luo Q. Primary care practice reengineering and associations with patient portal use, service utilization, and disease control among patients with hypertension and/or diabetes. Ochsner J 2017;17(1):103-111 [FREE Full text] [Medline: 28331456]
57. Price-Haywood EG, Luo Q, Monlezun D. Dose effect of patient-care team communication via secure portal messaging on glucose and blood pressure control. J Am Med Inform Assoc 2018 Jun 01;25(6):702-708 [FREE Full text] [doi: 10.1093/jamia/ocx161] [Medline: 29444256]
58. Quinn CC, Butler EC, Swasey KK, Shardell MD, Terrin MD, Barr EA, et al. Mobile diabetes intervention study of patient engagement and impact on blood glucose: mixed methods analysis. JMIR Mhealth Uhealth 2018 Mar 02;6(2):e31 [FREE Full text] [doi: 10.2196/mhealth.9265] [Medline: 29396389]
59. Reed M, Graetz I, Gordon N, Fung V. Patient-initiated e-mails to providers: associations with out-of-pocket visit costs, and impact on care-seeking and health. Am J Manag Care 2015 Dec 01;21(12):e632-e639 [FREE Full text] [Medline: 26760425]

60. Reed ME, Huang J, Brand RJ, Neugebauer R, Graetz I, Hsu J, et al. Patients with complex chronic conditions: health care use and clinical events associated with access to a patient portal. PLoS One 2019 Jun 19;14(6):e0217636 [FREE Full text] [doi: 10.1371/journal.pone.0217636] [Medline: 31216295]
61. Reed ME, Huang J, Millman A, Graetz I, Hsu J, Brand R, et al. Portal use among patients with chronic conditions: patient-reported care experiences. Med Care 2019 Oct;57(10):809-814. [doi: 10.1097/MLR.0000000000001178] [Medline: 31415340]
62. Riippa I, Linna M, Rönkkö I. The effect of a patient portal with electronic messaging on patient activation among chronically ill patients: controlled before-and-after study. J Med Internet Res 2014 Nov 19;16(11):e257 [FREE Full text] [doi: 10.2196/jmir.3462] [Medline: 25413368]
63. Riippa I, Linna M, Rönkkö I. A patient portal with electronic messaging: controlled before-and-after study. J Med Internet Res 2015 Nov 09;17(11):e250 [FREE Full text] [doi: 10.2196/jmir.4487] [Medline: 26553595]
64. Robinson SA, Zocchi MS, Netherton D, Ash A, Purington CM, Connolly SL, et al. Secure messaging, diabetes self-management, and the importance of patient autonomy: a mixed methods study. J Gen Intern Med 2020 Oct;35(10):2955-2962 [FREE Full text] [doi: 10.1007/s11606-020-05834-x] [Medline: 32440998]
65. Ronda MC, Dijkhorst-Oei LT, Rutten GE. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
66. Ronda MC, Dijkhorst-Oei LT, Rutten GE. Patients' experiences with and attitudes towards a diabetes patient web portal. PLoS One 2015 Jun 18;10(6):e0129403 [FREE Full text] [doi: 10.1371/journal.pone.0129403] [Medline: 26086272]
67. Sabo R, Robins J, Lutz S, Kashiri P, Day T, Webel B, et al. Diabetes engagement and activation platform for implementation and effectiveness of automated virtual type 2 diabetes self-management education: randomized controlled trial. JMIR Diabetes 2021 Mar 29;6(1):e26621 [FREE Full text] [doi: 10.2196/26621] [Medline: 33779567]
68. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
69. Seo D, Park YR, Lee Y, Kim JY, Park JY, Lee JH. The use of mobile personal health records for hemoglobin A1c regulation in patients with diabetes: retrospective observational study. J Med Internet Res 2020 Jun 02;22(6):e15372 [FREE Full text] [doi: 10.2196/15372] [Medline: 32484447]
70. Sharit J, Idrees T, Andrade AD, Anam R, Karanam C, Valencia W, et al. Use of an online personal health record's Track Health function to promote positive lifestyle behaviors in Veterans with prediabetes. J Health Psychol 2018 Apr;23(5):681-690. [doi: 10.1177/1359105316681065] [Medline: 27920356]
71. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained use of patient portal features and improvements in diabetes physiological measures. J Med Internet Res 2016 Jul 01;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]
72. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
73. van Vugt M, de Wit M, Sieverink F, Roelofsen Y, Hendriks SH, Bilo HJ, et al. Uptake and effects of the e-vita personal health record with self-management support and coaching, for type 2 diabetes patients treated in primary care. J Diabetes Res 2016;2016:5027356 [FREE Full text] [doi: 10.1155/2016/5027356] [Medline: 26955640]
74. Vo MT, Uratsu CS, Estacio KR, Altschuler A, Kim E, Alexeeff SE, et al. Prompting patients with poorly controlled diabetes to identify visit priorities before primary care visits: a pragmatic cluster randomized trial. J Gen Intern Med 2019 Jun;34(6):831-838 [FREE Full text] [doi: 10.1007/s11606-018-4756-4] [Medline: 30746642]
75. Wald JS, Grant RW, Schnipper JL, Gandhi TK, Poon EG, Businger AC, et al. Survey analysis of patient experience using a practice-linked PHR for type 2 diabetes mellitus. AMIA Annu Symp Proc 2009 Nov 14;2009:678-682 [FREE Full text] [Medline: 20351940]
76. Zocchi MS, Robinson SA, Ash AS, Vimalananda VG, Wolfe HL, Hogan TP, et al. Patient portal engagement and diabetes management among new portal users in the Veterans Health Administration. J Am Med Inform Assoc 2021 Sep 18;28(10):2176-2183 [FREE Full text] [doi: 10.1093/jamia/ocab115] [Medline: 34339500]
77. Druss BG, Li J, Tapscott S, Lally CA. Randomized trial of a mobile personal health record for behavioral health homes. Psychiatr Serv 2020 Aug 01;71(8):803-809 [FREE Full text] [doi: 10.1176/appi.ps.201900381] [Medline: 32362226]
78. Aberger EW, Migliozzi D, Follick MJ, Malick T, Ahern DK. Enhancing patient engagement and blood pressure management for renal transplant recipients via home electronic monitoring and web-enabled collaborative care. Telemed J E Health 2014 Sep;20(9):850-854 [FREE Full text] [doi: 10.1089/tmj.2013.0317] [Medline: 25046403]
79. Ahmed S, Ernst P, Bartlett SJ, Valois MF, Zaihra T, Paré G, et al. The effectiveness of web-based asthma self-management system, my asthma portal (MAP): a pilot randomized controlled trial. J Med Internet Res 2016 Dec 01;18(12):e313 [FREE Full text] [doi: 10.2196/jmir.5866] [Medline: 27908846]

80. Apter AJ, Localio AR, Morales KH, Han X, Perez L, Mullen AN, et al. Home visits for uncontrolled asthma among low-income adults with patient portal access. J Allergy Clin Immunol 2019 Sep;144(3):846-53.e11 [FREE Full text] [doi: 10.1016/j.jaci.2019.05.030] [Medline: 31181221]
81. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, Localio AR, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973 [FREE Full text] [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
82. 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]
83. Kogut SJ, Goldstein E, Charbonneau C, Jackson A, Patry G. Improving medication management after a hospitalization with pharmacist home visits and electronic personal health records: an observational study. Drug Healthc Patient Saf 2014 Jan 17;6:1-6 [FREE Full text] [doi: 10.2147/DHPS.S56574] [Medline: 24465136]
84. Kim JW, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of personal health records and wearables on health outcomes and patient response: three-arm randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]
85. Lau AY, Arguel A, Dennis S, Liaw ST, Coiera E. "Why didn't it work?" Lessons from a randomized controlled trial of a web-based personally controlled health management system for adults with asthma. J Med Internet Res 2015 Dec 15;17(12):e283 [FREE Full text] [doi: 10.2196/jmir.4734] [Medline: 26678294]
86. Manard W, Scherrer JF, Salas J, Schneider FD. Patient portal use and blood pressure control in newly diagnosed hypertension. J Am Board Fam Med 2016;29(4):452-459 [FREE Full text] [doi: 10.3122/jabfm.2016.04.160008] [Medline: 27390376]
87. Toscos T, Coupe A, Wagner S, Ahmed R, Roebuck A, Flanagan M, et al. Engaging patients in atrial fibrillation management via digital health technology: the impact of tailored messaging. J Innov Card Rhythm Manag 2020 Aug;11(8):4209-4217 [FREE Full text] [doi: 10.19102/icrm.2020.110802] [Medline: 32874747]
88. Wagner PJ, Dias J, Howard S, Kintziger KW, Hudson MF, Seol Y, et al. Personal health records and hypertension control: a randomized trial. J Am Med Inform Assoc 2012;19(4):626-634 [FREE Full text] [doi: 10.1136/amiajnl-2011-000349] [Medline: 22234404]
89. Martínez Nicolás I, Lê Cook B, Flores M, Del Olmo Rodriguez M, Hernández Rodríguez C, Llamas Sillero P, et al. The impact of a comprehensive electronic patient portal on the health service use: an interrupted time-series analysis. Eur J Public Health 2019 Jun 01;29(3):413-418. [doi: 10.1093/eurpub/cky257] [Medline: 30544169]
90. Cahill JE, Lin L, LoBiondo-Wood G, Armstrong TS, Acquaye AA, Vera-Bolanos E, et al. Personal health records, symptoms, uncertainty, and mood in brain tumor patients. Neurooncol Pract 2014 Jun;1(2):64-70 [FREE Full text] [doi: 10.1093/nop/npu005] [Medline: 26034618]
91. Chiche L, Brescianini A, Mancini J, Servy H, Durand JM. Evaluation of a prototype electronic personal health record for patients with idiopathic thrombocytopenic purpura. Patient Prefer Adherence 2012;6:725-734 [FREE Full text] [doi: 10.2147/PPA.S36320] [Medline: 23077409]
92. Collins PW, Bolton-Maggs P, Stephenson D, Jenkins B, Loran C, Winter M. Pilot study of an Internet-based electronic patient treatment record and communication system for haemophilia, Advoy.com. Haemophilia 2003 May;9(3):285-291. [doi: 10.1046/j.1365-2516.2003.00747.x] [Medline: 12694519]
93. Coquet J, Blayney DW, Brooks JD, Hernandez-Boussard T. Association between patient-initiated emails and overall 2-year survival in cancer patients undergoing chemotherapy: evidence from the real-world setting. Cancer Med 2020 Nov;9(22):8552-8561 [FREE Full text] [doi: 10.1002/cam4.3483] [Medline: 32986931]
94. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Supporting lung cancer patients with an interactive patient portal: feasibility study. JMIR Cancer 2017 Aug 08;3(2):e10 [FREE Full text] [doi: 10.2196/cancer.7443] [Medline: 28790025]
95. Hall MJ, Herda MM, Handorf EA, Rybak CC, Keleher CA, Siemon M, et al. Direct-to-patient disclosure of results of mismatch repair screening for Lynch syndrome via electronic personal health record: a feasibility study. Genet Med 2014 Nov;16(11):854-861 [FREE Full text] [doi: 10.1038/gim.2014.42] [Medline: 24784156]
96. Hong MK, Wilcox L, Feustel C, Wasileski-Masker K, Olson TA, Simoneaux SF. Adolescent and caregiver use of a tethered personal health record system. AMIA Annu Symp Proc 2016 Feb 10;2016:628-637 [FREE Full text] [Medline: 28269859]
97. Kidwell KM, Peugh J, Westcott E, Nwankwo C, Britto MT, Quinn CT, et al. Acceptability and feasibility of a disease-specific patient portal in adolescents with sickle cell disease. J Pediatr Hematol Oncol 2019 Oct;41(7):561-567 [FREE Full text] [doi: 10.1097/MPH.0000000000001341] [Medline: 30339657]
98. O'Hea EL, Creamer S, Flahive JM, Keating BA, Crocker CR, Williamson SR, et al. Survivorship care planning, quality of life, and confidence to transition to survivorship: a randomized controlled trial with women ending treatment for breast cancer. J Psychosoc Oncol 2022;40(5):574-594. [doi: 10.1080/07347332.2021.1936336] [Medline: 34151734]
99. Pai HH, Lau F, Barnett J, Jones S. Meeting the health information needs of prostate cancer patients using personal health records. Curr Oncol 2013 Dec;20(6):e561-e569 [FREE Full text] [doi: 10.3747/co.20.1584] [Medline: 24311957]

100. Tarver WL, Robb BW, Haggstrom DA. Usefulness and usability of a personal health record and survivorship care plan for colorectal cancer survivors: survey study. JMIR Cancer 2019 Aug 20;5(2):e10692 [FREE Full text] [doi: 10.2196/10692] [Medline: 31432780]
101. Wiljer D, Leonard KJ, Urowitz S, Apatu E, Massey C, Quartey NK, et al. The anxious wait: assessing the impact of patient accessible EHRs for breast cancer patients. BMC Med Inform Decis Mak 2010 Sep 01;10:46 [FREE Full text] [doi: 10.1186/1472-6947-10-46] [Medline: 20809950]
102. Williamson RS, Cherven BO, Gilleland Marchak J, Edwards P, Palgon M, Escoffery C, et al. Meaningful use of an electronic personal health record (ePHR) among pediatric cancer survivors. Appl Clin Inform 2017 Mar 15;8(1):250-264 [FREE Full text] [doi: 10.4338/ACI-2016-11-RA-0189] [Medline: 28293684]
103. Anand T, Nitpolprasert C, Kerr SJ, Apornpong T, Ananworanich J, Phanuphak P, et al. Implementation of an online HIV prevention and treatment cascade in Thai men who have sex with men and transgender women using Adam's Love Electronic Health Record system. J Virus Erad 2017 Jan 01;3(1):15-23 [FREE Full text] [Medline: 28275454]
104. Bidmead E, Marshall A. A case study of stakeholder perceptions of patient held records: the Patients Know Best (PKB) solution. Digit Health 2016 Sep 21;2:2055207616668431 [FREE Full text] [doi: 10.1177/2055207616668431] [Medline: 29942567]
105. Crouch PB, Rose CD, Johnson M, Janson SL. A pilot study to evaluate the magnitude of association of the use of electronic personal health records with patient activation and empowerment in HIV-infected veterans. PeerJ 2015 Mar 19;3:e852 [FREE Full text] [doi: 10.7717/peerj.852] [Medline: 25802815]
106. Druss BG, Ji X, Glick G, von Esenwein SA. Randomized trial of an electronic personal health record for patients with serious mental illnesses. Am J Psychiatry 2014 Mar;171(3):360-368. [doi: 10.1176/appi.ajp.2013.13070913] [Medline: 24435025]
107. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in electronic health record patient portal use in nephrology clinics. Clin J Am Soc Nephrol 2015 Nov 06;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
108. Kahn JS, Hilton JF, Van Nunnery T, Leasure S, Bryant KM, Hare CB, et al. Personal health records in a public hospital: experience at the HIV/AIDS clinic at San Francisco General Hospital. J Am Med Inform Assoc 2010;17(2):224-228 [FREE Full text] [doi: 10.1136/jamia.2009.000315] [Medline: 20190069]
109. Keith McInnes D, Shimada SL, Rao SR, Quill A, Duggal M, Gifford AL, et al. Personal health record use and its association with antiretroviral adherence: survey and medical record data from 1871 US veterans infected with HIV. AIDS Behav 2013 Nov;17(9):3091-3100. [doi: 10.1007/s10461-012-0399-3] [Medline: 23334359]
110. McInnes DK, Shimada SL, Midboe AM, Nazi KM, Zhao S, Wu J, et al. Patient use of electronic prescription refill and secure messaging and its association with undetectable HIV viral load: a retrospective cohort study. J Med Internet Res 2017 Mar 15;19(2):e34 [FREE Full text] [doi: 10.2196/jmir.6932] [Medline: 28202428]
111. Kiberd J, Khan U, Stockman C, Radhakrishnan A, Phillips M, Kiberd BA, et al. Effectiveness of a web-based eHealth portal for delivery of care to home dialysis patients: a single-arm pilot study. Can J Kidney Health Dis 2018 Sep 7;5:2054358118794415 [FREE Full text] [doi: 10.1177/2054358118794415] [Medline: 30210802]
112. Lee J, Kim JG, Jin M, Ahn K, Kim B, Kim S, et al. Beneficial effects of two types of personal health record services connected with electronic medical records within the hospital setting. Comput Inform Nurs 2017 Nov;35(11):574-581. [doi: 10.1097/CIN.0000000000000362] [Medline: 28548973]
113. Miller DM, Moore SM, Fox RJ, Atreja A, Fu AZ, Lee JC, et al. Web-based self-management for patients with multiple sclerosis: a practical, randomized trial. Telemed J E Health 2011;17(1):5-13 [FREE Full text] [doi: 10.1089/tmj.2010.0133] [Medline: 21214498]
114. Navaneethan SD, Jolly SE, Schold JD, Arrigain S, Nakhoul G, Konig V, et al. Pragmatic randomized, controlled trial of patient navigators and enhanced personal health records in CKD. Clin J Am Soc Nephrol 2017 Sep 07;12(9):1418-1427 [FREE Full text] [doi: 10.2215/CJN.02100217] [Medline: 28778854]
115. Plimpton E. A quality improvement project to increase patient portal enrollment and utilization in women living with HIV at risk for disengagement in care. J Assoc Nurses AIDS Care 2020;31(1):60-65. [doi: 10.1097/JNC.0000000000000153] [Medline: 31834101]
116. Reich J, Canakis A, Shankar D, Harrington J, Apte M, Weinberg J, et al. The use of an EHR patient portal (Mychart-Epic) in patients with inflammatory bowel disease. Crohn's Colitis 360 2019 Oct;1(3):otz039. [doi: 10.1093/crocol/otz039]
117. Nielsen AS, Halamka JD, Kinkel RP. Internet portal use in an academic multiple sclerosis center. J Am Med Inform Assoc 2012;19(1):128-133 [FREE Full text] [doi: 10.1136/amiajnl-2011-000177] [Medline: 21571744]
118. Son H, Nahm ES. Older adults' experience using patient portals in communities: challenges and opportunities. Comput Inform Nurs 2019 Jan;37(1):4-10. [doi: 10.1097/CIN.0000000000000476] [Medline: 30188353]
119. Tom JO, Mangione-Smith R, Solomon C, Grossman DC. Integrated personal health record use: association with parent-reported care experiences. Pediatrics 2012 Jul;130(1):e183-e190. [doi: 10.1542/peds.2011-1786] [Medline: 22689872]
120. van den Heuvel SC, Meije D, Regeer EJ, Sinnema H, Riemersma RF, Kupka RW. The user experiences and clinical outcomes of an online personal health record to support self-management of bipolar disorder: a pretest-posttest pilot study. J Affect Disord 2018 Oct 01;238:261-268. [doi: 10.1016/j.jad.2018.05.069] [Medline: 29894931]

121. van der Vaart R, Drossaert CH, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MA. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014 Mar 26;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
122. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
123. Dunsch F, Evans DK, Macis M, Wang Q. Bias in patient satisfaction surveys: a threat to measuring healthcare quality. BMJ Glob Health 2018 Apr 12;3(2):e000694 [FREE Full text] [doi: 10.1136/bmjgh-2017-000694] [Medline: 29662696]
124. Hanna L, Gill SD, Newstead L, Hawkins M, Osborne RH. Patient perspectives on a personally controlled electronic health record used in regional Australia. Health Inf Manag 2017 Jan;46(1):42-48. [doi: 10.1177/1833358316661063] [Medline: 27486184]
125. Fylan F, Caveney L, Cartwright A, Fylan B. Making it work for me: beliefs about making a personal health record relevant and useable. BMC Health Serv Res 2018 Jun 14;18(1):445 [FREE Full text] [doi: 10.1186/s12913-018-3254-z] [Medline: 29898730]
126. Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WS, Patel V, Moser RP, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017 Jun 02;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
127. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. J Fam Pract 2000 Sep;49(9):796-804. [Medline: 11032203]
128. Antonio MG, Petrovskaya O, Lau F. The state of evidence in patient portals: umbrella review. J Med Internet Res 2020 Nov 11;22(11):e23851 [FREE Full text] [doi: 10.2196/23851] [Medline: 33174851]

Abbreviations

ED: emergency department
JBI: Joanna Briggs Institute
PHR: personal health record
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial

Edited by A Mavragani; submitted 30.09.22; peer-reviewed by M Prietula, M Kapsetaki, F Lau, D Kalra; comments to author 24.10.22; revised version received 14.11.22; accepted 15.11.22; published 22.12.22.

Please cite as:
Brands MR, Gouw SC, Beestrum M, Cronin RM, Fijnvandraat K, Badawy SM
Patient-Centered Digital Health Records and Their Effects on Health Outcomes: Systematic Review
J Med Internet Res 2022;24(12):e43086
URL: https://www.jmir.org/2022/12/e43086
doi:10.2196/43086
PMID:36548034

Review

Patients Managing Their Medical Data in Personal Electronic Health Records: Scoping Review

Debby J Damen[1], PhD; Guus G Schoonman[1,2], MD, PhD; Barbara Maat[3], MD, PhD; Mirela Habibović[4], PhD; Emiel Krahmer[1], Prof Dr; Steffen Pauws[1,5], Prof Dr

[1]Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands

[2]Department of Neurology, Elisabeth-TweeSteden Hospital, Tilburg, Netherlands

[3]Department of Pharmacy, Elisabeth-TweeSteden Hospital, Tilburg, Netherlands

[4]Department of Medical and Clinical Psychology, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

[5]Department of Remote Patient Management & Connected Care, Philips Research, Eindhoven, Netherlands

Corresponding Author:
Debby J Damen, PhD
Department of Communication and Cognition
Tilburg School of Humanities and Digital Sciences
Tilburg University
Warandelaan 2
Tilburg, 5037 AB
Netherlands
Phone: 31 0630215926
Email: d.j.damen@tilburguniversity.edu

Abstract

Background: Personal electronic health records (PEHRs) allow patients to view, generate, and manage their personal and medical data that are relevant across illness episodes, such as their medications, allergies, immunizations, and their medical, social, and family health history. Thus, patients can actively participate in the management of their health care by ensuring that their health care providers have an updated and accurate overview of the patients' medical records. However, the uptake of PEHRs remains low, especially in terms of patients entering and managing their personal and medical data in their PEHR.

Objective: This scoping review aimed to explore the barriers and facilitators that patients face when deciding to review, enter, update, or modify their personal and medical data in their PEHR. This review also explores the extent to which patient-generated and -managed data affect the quality and safety of care, patient engagement, patient satisfaction, and patients' health and health care services.

Methods: We searched the MEDLINE, Embase, CINAHL, PsycINFO, Cochrane Library, Web of Science, and Google Scholar web-based databases, as well as reference lists of all primary and review articles using a predefined search query.

Results: Of the 182 eligible papers, 37 (20%) provided sufficient information about patients' data management activities. The results showed that patients tend to use their PEHRs passively rather than actively. Patients refrain from generating and managing their medical data in a PEHR, especially when these data are complex and sensitive. The reasons for patients' passive data management behavior were related to their concerns about the validity, applicability, and confidentiality of patient-generated data. Our synthesis also showed that patient-generated and -managed health data ensures that the medical record is complete and up to date and is positively associated with patient engagement and patient satisfaction.

Conclusions: The findings of this study suggest recommendations for implementing design features within the PEHR and the construal of a dedicated policy to inform both clinical staff and patients about the added value of patient-generated data. Moreover, clinicians should be involved as important ambassadors in informing, reminding, and encouraging patients to manage the data in their PEHR.

(J Med Internet Res 2022;24(12):e37783) doi:10.2196/37783

KEYWORDS

patient-generated data; patient portal; personal electronic health record; patient activation; patient engagement

Introduction

Background

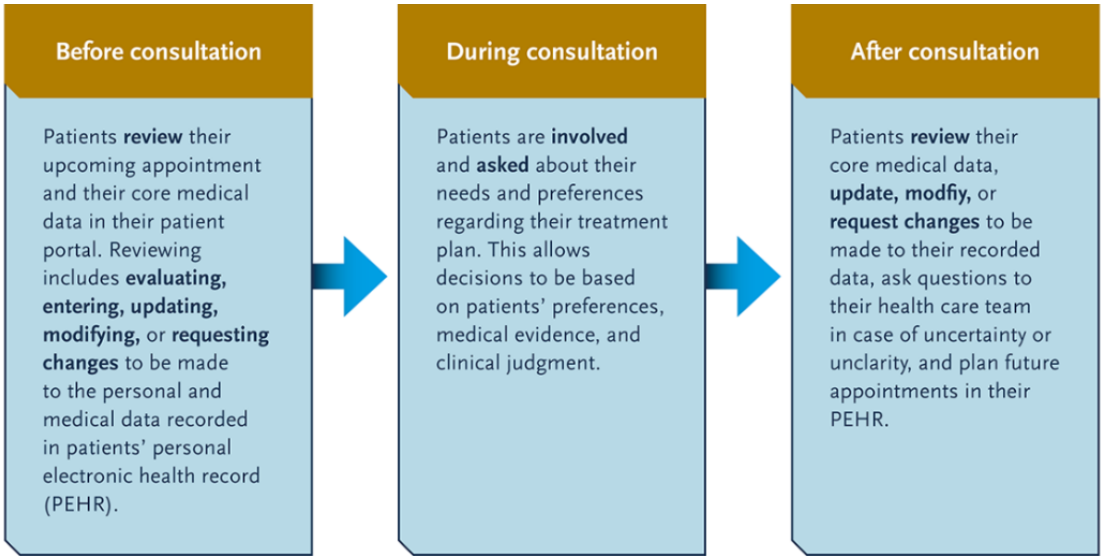
The beginning of most outpatient consultations is characterized by physicians going over the personal and medical information that is recorded in their patients' personal electronic health records (PEHRs). This includes information about their patients' current health problems and information about their vital signs, medication use, or known allergies. An up-to-date and accurate overview of this personal and medical information gives physicians a better sense of who is sitting in front of them and allows them to make appropriate and safe treatment-related decisions that correspond to their patients' needs. In most cases, clinicians are responsible for updating their patients' personal and medical data at the start of each consultation. However, this task can take up to 40% of the physicians' time, which would rather be spent on direct patient care [1,2]. Instead of only physicians managing their patients' personal and medical data (*core medical data*), patients can also play a role by entering, reviewing, and updating this information in their PEHR before or after each outpatient visit by themselves. Research shows that this active patient engagement is associated with various beneficial health-related outcomes, such as an increase in patients' self-care and medication adherence, improved patient-physician relations, shared decision-making, and even improved clinical outcomes for patients with chronic illnesses [3-5]. It is for this reason that health care services strive to engage patients in the self-entry and self-management of their health care data by using technology such as patients' PEHRs [6].

Over the past decade, identifying what determines whether patients are likely to engage with their PEHRs and how their engagement affects their clinical care has been a frequent topic of discussion [7-14]. The consensus is that less than half of the user population adopts a PEHR, and even less than one-third of the users actually use their PEHR records and manage their personal and medical data, with patients' data management declining as age increases, lower digital skills, and being unable to fully understand and use health information in treatment-related decisions [15-18]. Studies have also shown that patients are less likely to self-manage their medical data when they find it difficult or unpleasant to use the data management tools [11,19-23] or when the practice is not endorsed by their health care providers [21,24].

Although previous syntheses of the literature have been valuable in identifying the scope and potential causes of patients' disengagement [7-10,13,14,25], they have some limitations. First, the most recent review [10] synthesized knowledge from studies published till 2018 and retrieved them from a very limited set of 3 databases. Second, previous reviews have focused only on consumers' perceptions [7,10,13], patients aged ≥50 years [14], randomized controlled trials [8], or English publications [7,9,10,14], without providing an all-encompassing view on the patient-, care-, and system-related factors that drive or prevent patients' data management. Most importantly, previous literature refrains from providing sufficient information about patients' actual levels of engagement with their core medical data in their PEHR. The facilitators of and barriers to patients' personal data management have previously been considered in relation to patients' (future) portal adoption or access [25-27] or by basing patients' level of engagement on log-in frequencies or the number of times they view a certain page in their PEHR [7-10,12-14]. In these cases, we do not know the extent to which patients who access their PEHR feel coresponsible or "empowered" [28] to actually use their PEHR in a meaningful way. We define meaningful use as patients actively sharing, reviewing, updating, or modifying their personal and medical data in their PEHR throughout their entire care journey (Figure 1). Our definition does not include patients who only access their portal and passively view the recorded information, but it does include patients who evaluate the information recorded in their PEHR. Certainly, patients are meaningfully using their PEHR when they closely examine (evaluate) their core medical data and decide to leave the information as it is, because they believe it to be correct and complete (Figure 1). However, we know that PEHRs often lack sufficient or up-to-date core medical information [29]. Therefore, in this review, our aim is to synthesize the existing literature by focusing on instances in which patients take actual action to provide or update their core medical data in their PEHR. This focus on data generation (sharing) and management (updating and modifying) allows us (1) to determine what drives patients toward or prevents patients from maintaining an up-to-date record and (2) to examine the associated impact that this active data management has on patients' health and health care–related services.

To identify what may drive patients toward or prevent patients from taking on an active rather than a passive role when it comes to the management of their core medical data, we need to identify not only the type of data management activities patients perform within their portal but also the type of data that patients manage and how frequently they do so. Patients can engage differently with their PEHR depending on the personal and medical data they wish to share or update. Patients may be less inclined to share or update information about error-prone and sensitive data elements than to share or update personal and medical data that they are more confident or knowledgeable about. To date, it remains unknown whether the type of core medical information affects patients' personal data management.

Figure 1. Active patient engagement in terms of patients generating and managing their personal and medical data throughout their care journey. This figure was partially replicated and adapted from Carman et al [30]. PEHR: personal electronic health record.



Objectives

In this scoping review, we aimed to address the limitations of previous syntheses by exploring the barriers and facilitators that patients face when they decide to actively review, enter, update, or modify their core medical data in their PEHR throughout their care journey (Figure 1). We aimed to (1) identify the extent to which patients feel motivated or coresponsible for sharing, updating, and modifying their core medical data in their PEHR, and (2) examine the extent to which this engagement with a PEHR impacts the quality and safety of care and patients' satisfaction with the care delivered. Answers to these questions will result in clear recommendations on how to maximally stimulate active patient involvement with PEHRs.

Methods

Search Strategy and Eligibility Criteria

This scoping review was conducted and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews [31]; Multimedia Appendix 1). The search protocol was preregistered with the Open Science Framework [32]. In April 2020, the MEDLINE (PubMed), Embase, CINAHL, PsycINFO, Cochrane Library, Web of Science, and Google Scholar web-based databases were searched to retrieve studies concerning patients' management of their core medical data in an electronic patient portal. In March 2022, the MEDLINE database was re-searched to retrieve records that were published between April 2020 and March 2022. The reference lists of all primary and review articles were hand searched. Literature reviews were excluded, but practice briefs, fact sheets, white papers, and peer-reviewed publications (including conference proceedings) that focused on any type of population or study design (eg, qualitative, quantitative, or mixed methods studies) were included. The databases were searched for English or Dutch articles published between January 2000 and February 2020. We chose January 2000 as the starting point of the search because the 3 known early adopters of a web-based patient portal, the Palo Alto Medical Foundation ("MyChart"), the Beth Israel Deaconess Medical Center ("PatientSite"), and the Boston Children's Hospital ("Indivo"), implemented their patient portals between the end of 1999 and the beginning of 2000 [33]. Our search strategy was developed in collaboration with an experienced research librarian (Multimedia Appendix 2) and targeted words related to electronic health records (eg, *patient portal* and *electronic health record*) combined with Medical Subject Headings terms related to patient engagement (eg, *patient participation*, *patient education*, *patient involvement*, and *patient engagement*) and the type of data being managed (eg, *medication reconciliation*, *medication verification*, *allergies*, and *intoxications*). To be included in the review, papers needed to focus on patients who actively handled their personal and medical data in a web-based patient portal (ie, entering, updating, or modifying; Figure 1) and identify either patient-, care-, or system-related determinants that influence this active patient involvement, or focus on the (perceived or examined) benefits or costs related to active patient involvement with a PEHR. Articles were excluded when they only included patients' management of their core medical data in a PEHR as a secondary concept. Table 1 provides an overview of the checklist for full articles.

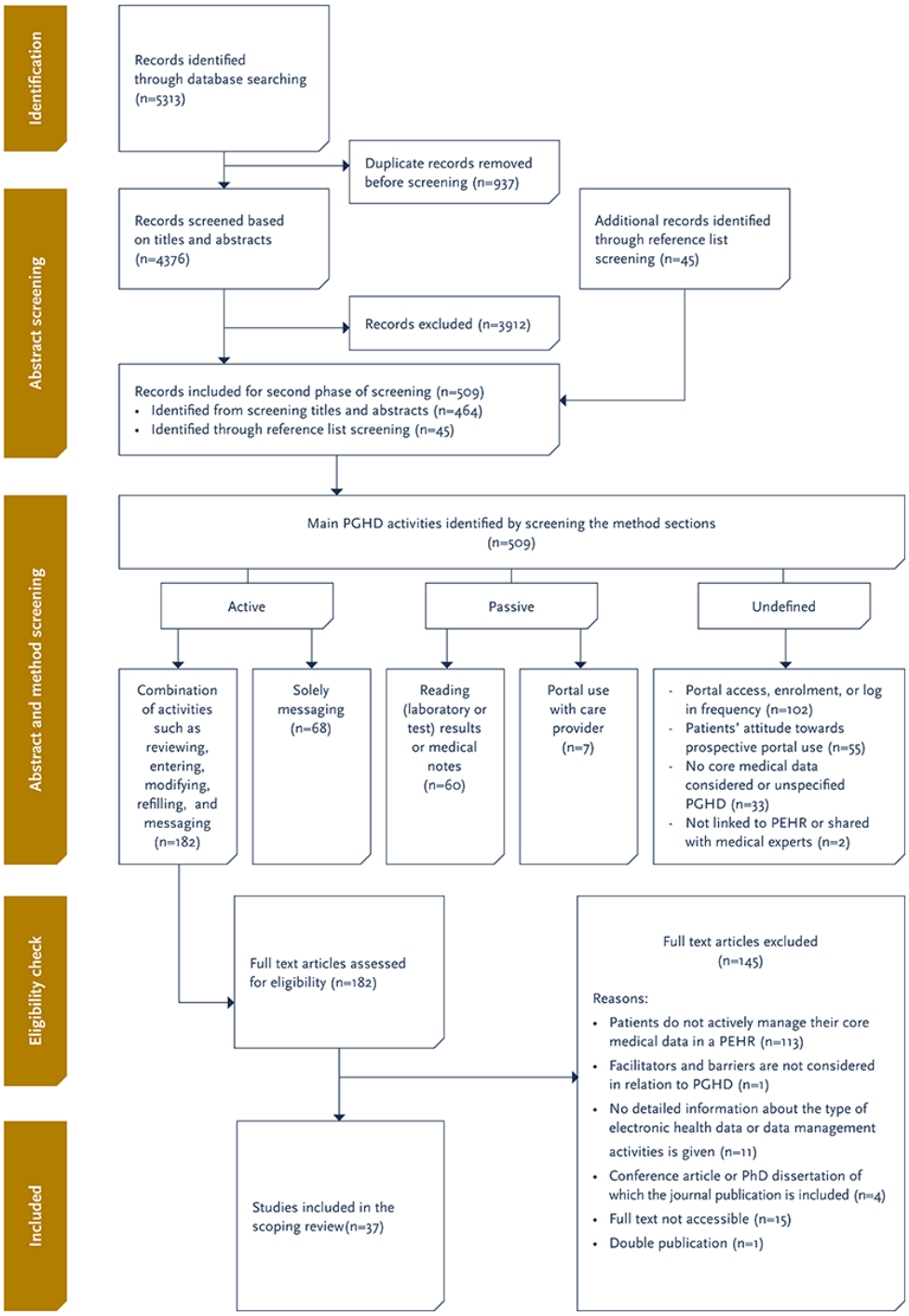
Table 1. Selection checklist for full articles.

Item	Inclusion
Report characteristics	
Type of publication	Practice briefs, fact sheets, white papers, and peer-reviewed publications and conference proceedings. Exclude when the articles are systematic or scoping reviews; meta-analyses
Date of publication	Between 2000 and February 2020; MEDLINE: re-searched in March 2022
Study details	
Type of study or intervention	All types of studies are allowed to be included in this review (eg, randomized controlled trial, non-randomized controlled trial, evaluation/usability, experimental, cohort/longitudinal, developmental, and pre-post design)
Type of health data being managed	Core medical data being managed in a personal electronic health record (eg, medication regimen, vaccinations, allergies, medical and family history, and intoxications)
Population	Both patients and clinicians

Screening Rounds and Data Extraction

The flowchart for the inclusion of articles in the scoping review is presented in Figure 2. The eligibility screening and data extraction form is presented in Multimedia Appendix 3. Searching the databases resulted in 5313 records that were imported into the reference manager, Mendeley (Elsevier). After duplicates were removed, 4376 (5313/4376, 82%) unique records were retained. The first author (DJD) used Mendeley to screen the identified records based on their titles and abstracts. A total of 45 (1%) additional records were identified through the screening of reference lists. This initial screening resulted in 509 records that were identified to be eligible for the review. However, after this initial screening, it remained unclear what kinds of activities patients performed within the PEHRs. Therefore, we diverged from our preregistered review protocol by administering an additional screening round. In this round, the first author (DJD) screened the Methods section of the 509 records to identify what kind of patient-generated medical data activities were included. This screening method identified 7 activities (Figure 2): active (ie, generating data, refilling, and messaging), passive (ie, viewing and portal use with health care provider), and undefined data management activities (ie, prospective use, portal access, log-in frequency, and portal enrollment). The first author (DJD) categorized the records into these 7 categories, and the second author (GGS) screened and reviewed a subset (51/509, 10%) of these records. Both authors discussed the screening method and the categorized subset until a consensus was reached. After the screening of the Method sections, 182 articles were found to be eligible for full-article screening. The full texts of these 182 records were subsequently screened by 4 authors (DJD, GGS, BM, and SP) in equally divided subsets. This resulted in 37 (20%) records that met the criteria for inclusion in this scoping review. The first (DJD) and second (GGS) authors then rated a subset of a mix of inclusions and exclusions, but no problematic cases were identified. The first author (DJD) then commenced with extracting the data from the 37 (20%) records according to the data extraction form (Multimedia Appendix 3).

Figure 2. Flowchart for the identification, screening, and inclusion of articles in this scoping review. PGHD: patient-generated health data; PEHR: personal electronic health record.



Results

Description of the Included Studies

The general characteristics of the 37 included records are presented in Table 2. We rejected articles that only addressed patients who passively reviewed their data without making actual changes to their records (eg, the studies by Apter et al [34] and Jhamb et al [35]). We categorized the included studies as reporting on one or more of the following three categories (Table 3) [33,36]: (1) information about patients' portal use, including the frequency of patients entering, updating, or modifying their core medical data; (2) patient and provider (perceived) facilitators of and barriers to the activities described in the first category, including usability, prototyping, and pilot studies in which portal features or tools were tested with specific end users; and (3) the impact of patients' active involvement in the management of their data on patient care, including studies that focused on the quality of the data entered and the (perceived or examined) effects of patient-generated or patient-managed data on the quality, safety, cost-effectiveness, and patient or health care provider satisfaction of health care services. In further sections, we will report the findings of the included studies based on these categories.

Table 2. Study characteristics of the records included in the scoping review.

Number	Study	Country	Study aim	Sample	Type of data	Data activity	Portal	Data entry tools
1	Ali et al [37], 2018	United States	Evaluating the usability of a portal	Patients or caretakers of patients (n=23) with chronic conditions (diabetes, cancer, ulcerative colitis, or thalassemia)	Medical history	Reviewing and entering data	myNYP	None
2	Ancker et al [38], 2019	United States	Describing portal adoption rates and characteristics of patients who enter health data and their association with clinical outcomes	Patients with diabetes (n=53), of which 23 were pregnant and 30 were nonpregnant, and their physicians in obstetrics-gynecology (n=12) or internal medicine (n=4)	Blood glucose values	Entering data	Weill Cornell Connect (EpiCare)	None
3	Arsoniadis et al [39], 2015	United States	Evaluating the quality of patient-generated health data with a health history tool accessible via the web or a tablet	Patients (n=146) with an appointment at a surgery clinic, of whom 50 completed the intervention	Medical history, surgical history, and social history (including questions related to tobacco use, alcohol consumption, illicit substance use, and sexual history)	Entering data	EpiCare	Questionnaires
4	Bajracharya et al [40], 2019	United States	Evaluation of the family history module implemented in a patient portal and patients' adoption of and experiences with the module	Patients (n=4223)	Family health history	Reviewing and entering and modifying data	PatientSite (electronic medical record of the Beth Israel Deaconess Medical Center)	Questionnaires
5	Bryce et al [41], 2008	United States	Exploring the usability of patient portal features and users' intentions to pay fees for portal use for a diabetes management portal	Patients (n=39) with diabetes, with 21 patients allocated to the preportal group and 18 to the portal users group	Vital signs (blood glucose values)	Entering data	HealthTrak	Calculator
6	Chrischilles et al [42], 2014	United States	Exploring how patient-generated health data affects medication use safety among older adults	Nonclinical population (n=1075) with variety in medical backgrounds; most participants were experiencing stomach-related problems; 802 participants were allocated to use a patient portal, and 273 were allocated to a control group	List of allergies, medication list, problem list, and medical history	Entering data	Iowa PHR[a] (stand-alone patient portal)	None
7	Cohn et al [43], 2010	United States	Evaluating the usability and analytic validity of the Health Heritage tool that helps patients to collect their family health history	Mixture of nonclinical and clinical participants (n=109), of which 54 were allocated to the intervention arm (Health Heritage) and 55 to the usual care arm	Family health history	Entering data	Health Heritage (stand-alone tool)	None

Number	Study	Country	Study aim	Sample	Type of data	Data activity	Portal	Data entry tools
8	Polubriaginof and Pastore [29], 2016	United States	Comparing the accuracy and completeness of a tablet-administered problem list questionnaire to a problem list that was self-reported by patients	Patients with variety in medical backgrounds (n=1472); details were given for patients with hypercholesterolemia and diabetes	Problem list, medical history, family health history, and risk factors	Entering data	LMR[b]	Tablet questionnaire administered via the Hughes RiskApps life cycle cost software
9	Dullabhet et al [44], 2014	United States	Exploring how patients can be engaged to provide feedback on electronic health record content and how this feedback affects the accuracy of medical records	Patients (n=457) with chronic conditions (obstructive pulmonary disease, asthma, hypertension, diabetes, or heart failure); the number of providers and pharmacists interviewed is not provided	Medication list	Reviewing and modifying data	MyGeisinger (Geisinger Health System)	Web-based feedback forms
10	Eschler et al [45], 2016	United States	Exploring the usability of a patient portal, whether and how it helps patients to remember important health tasks, and whether it enhances patient engagement and agency in managing a chronic illness	Patients with diabetes and parents managing asthma for child dependents (n=19)	Immunization record	Reviewing and entering data	Three paper prototypes that represented features of a regional health cooperative portal's interface were used	None
11	Hanauer et al [46], 2014	United States	Exploring the frequency, type, reasons, and outcomes of patient-initiated amendment requests	Patients (n=181) for whom amendment requests were made to various clinical departments and divisions but whose medical conditions were unspecified	Medical history, social history, intoxications, family health history, clinic notes, discharge summaries, and emergency department notes	Reviewing and modifying data	MyChart (Epic)	To initiate a chart amendment request, the patient had to contact the information management department by phone, by mail, fax or in person and obtain an amendment request form
12	Heyworth et al [47], 2013	United States	Testing a medication reconciliation tool to improve medication safety among patients who were recently discharged from the hospital	Patients (n=25) with chronic conditions (eg, diabetes, hypertension, prior myocardial infarction or stroke, hyperlipidemia, and heart disease)	Medication list	Reviewing and entering and modifying data	My HealtheVet (The Veterans Health Administration)	Secure Messaging for Medication Reconciliation Tool within the portal
13	Hill et al [48], 2018	United States	Exploring health care providers perceived advantages and disadvantages of PHR portal use	Health care providers (n=26) who treat patients with spinal cord injuries and disorders	Vital signs (blood pressure, pulse rate, and weight), medical history, immunization record, and medication list	Reviewing and entering data	My HealtheVet (The Veterans Health Administration)	None

Number	Study	Country	Study aim	Sample	Type of data	Data activity	Portal	Data entry tools
14	Laranjo et al [49], 2017	Portugal	Examining portal use, associated patient demographics, and clinical variables	Patients (n=109,619), of whom 18,504 were portal users	Vital signs (height, weight, blood pressure, glycemia, cholesterol, and triglycerides levels) and allergies	Entering data	Tethered PHR provided by the National Health Service	None
15	Lemke et al [50], 2020	United States	Exploring primary care physicians' experiences with the Genetic and Wellness Assessment tool for capturing patients' family health history	Health care providers (n=24) who specialized in internal medicine, family medicine, or obstetrics/gynecology	Family health history	Entering data	Epic	Genetic and Wellness Assessment tool
16	Lesselroth et al [51], 2009	United States	Exploring the extent to which kiosk technology improves the reporting of patients' medication history	Patients (n=17,868) visiting a chemotherapy facility	Medication list and list of allergies	Reviewing and entering and modifying data	See Data Entry Tools	Automated Patient History Intake Device accessed via computer terminal kiosk in the clinical waiting room
17	Murray et al [52], 2013	United States	To examine the capacity of 3 different electronic tools for collecting patients' family health history	Patients (n=959) scheduled for an annual examination visit, of which 663 were allocated to the intervention arms (interactive voice response technology, patient portal, and waiting room laptop computer)	Family health history	Reviewing and entering data	Patient Gateway, LMR	The Surgeon General: My Family Health Portrait
18	Nagykaldi et al [53], 2012	United States	Examining the behavior and experiences of patients and primary care clinicians with regard to the Wellness Portal	Patients in primary care (n=560) who were in the randomized controlled trial; 3 clinicians, 2 office staff, and 6 patients in the pilot testing of the portal	Vital signs (weight), preventive services (mammography, diabetes education, and smoking counseling), wellness plan, symptom diary, medical history, medication list, problem list, list of allergies, and immunization record	Reviewing and entering data	Wellness Portal linked to the Preventive Services Reminder System	None
19	Nazi et al [54], 2013	United States	Exploring Veterans' perspectives on receiving access to their personal medical information, which of its data elements they find most valuable, and how it affects their satisfaction, self-management, communication, and health care quality	Military service Veterans in the United States (n=688)	Medication list, list of allergies, and vital signs (eg, blood pressure, blood sugar, and cholesterol)	Entering data	My*HealtheVet* and Veterans Information System Technology Architecture	None

Number	Study	Country	Study aim	Sample	Type of data	Data activity	Portal	Data entry tools
20	Park et al [55], 2018	Korea	Evaluating how and which users are generating and managing their personal and medical data	Patients with diabetes (n=16,729) and general users of the app (n=1536)	Vital signs (blood pressure, blood glucose levels, and weight); the functions list of allergies, medical history, and medication list were excluded because the number of users was relatively small (n=116)	Entering data	Mobile PHR known as My Chart in My Hand	None
21	Powell and Deroche [56], 2020	United States	Exploring the determinants of portal use among patients with multiple chronic conditions	Patients with multiple morbidities (n=500) with diabetes, heart failure, hypertension, and coronary artery disease	Vital signs (eg, weight and blood pressure)	Entering data	FollowMyHealth (AllScripts)	None
22	Prey et al [57], 2018	United States	Exploring the extent to which an electronic home medication review tool engaged patients in the medication reconciliation process and how this affected medication safety during hospitalization	Patients (n=65) arriving at the emergency department and their health care providers (n=20)	Medication list	Reviewing and entering and modifying data	AllScripts	Internally developed home medication review tool
23	Raghu et al [58], 2015	United States	Exploring the extent to which secure messaging helps patients to update their medication list in an ambulatory care setting	Patients (n=18,702) of a clinical practice that focused on surgical care for adults, of which 7818 had portal access	Medication list	Reviewing and entering data	Not specified	A secure messaging feature (alongside phone calls) was used by patients to update their medication list
24	Schnipper et al [59], 2012	United States	Investigating the extent to which a PHR-linked medications review module affects medication accuracy and safety	Patients in primary care (n=541), of which 267 were in the intervention arm	Intervention arm: medication list, list of allergies, and diabetes management information; control arm: family health history	Reviewing and modifying data	Patient Gateway, LMR	Patient Gateway medications module; electronic journals
25	Seeber et al [60], 2017	Germany	Validating the accuracy of VaccApp in helping parents to report their children's vaccine history	Parents (n=456) of infants and children with suspected vaccine-preventable diseases (eg, influenza-like illness or infections of the central nervous system)	Immunization record	Reviewing and entering data	Vaccination app (VaccApp)	None
26	Sun et al [61], 2019	United States	Exploring how patients with type 2 diabetes use their patient portals and what determines their portal use	Parents (n=456) of children with diabetes, of which 178 used the app	Medication list, list of allergies, and medical history	Reviewing and entering data	Epic	Questionnaire for recording medical history

Number	Study	Country	Study aim	Sample	Type of data	Data activity	Portal	Data entry tools
27	Tsai et al [62], 2019	United States	Exploring the characteristics of portal users and the activities that users perform within their patient portals	Patients (n=505,503), of which 109,200 were registered for a portal	Problem list, medication list, and list of allergies	Reviewing and entering and modifying data	MyChart (Epic)	None
28	Wald et al [63], 2010	United States	Exploring patients' and health care providers' experiences of using previsit electronic journals to record core medical data and survey data	Patients in primary care (n=2027 in the intervention arm and n=2345 in the postintervention survey) and 84 physicians	Arm 1: medication list, list of allergies, and diabetes items; arm 2: health maintenance, personal history, and family health history	Reviewing and entering and modifying data	Patient Gateway, LMR	Previsit electronic journals with tailored and untailored questions
29	Yu et al [64], 2015	United States	Exploring and identifying the needs and preferences of individuals with dexterity impairments when they use iMHere.	Patients with dexterity impairments (n=9)	Medication list and problem list	Entering reasons for taking medication and modifying medication reminders	Interactive mobile health and rehabilitation apps. iMHere is a system that connects smartphone apps to clinicians' web-based portal.	MyMeds app (medication management) and SkinCare app (monitoring and reporting skin breakdown)
30	Zettel-Watson and Tsukerman [65], 2016	United States	Exploring the use patterns among users of web-based health management tools and identifying barriers to use among nonusers	Nonclinical population (n=166)	Vital Signs (cholesterol, blood pressure, and glucose levels; uploading data from a monitoring device)	Reviewing and entering data	Most participants used tools provided by their physician's office, hospital, or insurance company (type of records unspecified)	None
31	Siek et al [66], 2011	United States	Testing the usability of an open source, web-based personal health app that provides older adults and their caregivers the ability to manage their personal health information during care transitions	Older adult patients with multiple morbidities (n=31)	Medication list	Reviewing and entering data	Colorado Care Tablet, personal health app	Pharmacy fulfillment and barcode scanning and a Prepare For Appointments wizard
32	Lober et al [67], 2006	United States	Exploring the barriers that older adults and disabled persons face when using PHRs	Nonclinical population (n=38) specified as low-income older adults with disabilities residing in a publicly subsidized housing project	Family health history, list of allergies, medication list, medical history, and immunization record	Reviewing and entering and modifying data	Personal Health Information Management System	A nurse was available to help with data entry
33	Arar et al [68], 2011	United States	To assess the facilitators of and barriers to Veterans' use of the Surgeon General's web-based tool to capture their family health history	Veterans (n=35)	Family health history	Entering data	My HealtheVet (The Veterans Health Administration)	The Surgeon General: My Family Health Portrait

Number	Study	Country	Study aim	Sample	Type of data	Data activity	Portal	Data entry tools
34	Wu et al [69], 2014	United States	Assessing the content and quality of the MeTree family health history tool	Patients in primary care (n=1184)	Family health history	Entering data	MeTree	None
35	Cimino et al [70], 2002	United States	Exploring patients' portal use, the cognitive effects of portal use and how it affects the patient–health care provider relationship	Patients (n=12) and health care providers (n=3)	Vital signs (height, weight, blood pressure, pulse, and temperature) and diabetes diary	Reviewing and entering data	Patient Clinical Information System, New York Presbyterian Hospital clinical data repository	None
36	Witry et al [71], 2010	United States	Exploring family practice physician and staff views on the (dis)advantages of PHR use	Health care providers (n=28) of a family medicine department	Medical history, medication list, and vital signs (blood pressure and glucose levels)	Entering data	Not specified	None
37	Kim and Johnson [72], 2004	United States	Exploring whether and how different types of data entry methods used by PHRs affect the accuracy of patient-generated data	Patients with disorders requiring treatment with thyroid hormone preparations (n=14)	Problem list and medication list	Reviewing and entering data	Password-protected website used to test data entry methods	Free-text entry (recall or abstraction) and selection methods

[a]PHR: patient health record.

[b]LMR: longitudinal medical record.

Table 3. Categorization of patient management papers and study type (N= 37).

Categories	Records[a], n (%)	Study types and references
Frequency of portal use	27 (73)	• Observational [38,42,49,55,56,58,62,70] • Content analysis [44,46,51,63] • RCT[b] [42,53,59,63] • RT[c] [57] • NRT[d] [52] • Cohort [43,61] • Interview [44,47,50] • Usability [47] • Survey [54,65,70]
Facilitators and barriers		
Patient-related	33 (89)	• Observational [38,42,49,55,56,58,62,63,70] • Content analysis [39,44,46,69] • RCT [42,53,63] • RT [57] • Cohort [61] • Interview [44,47,50,66,68,71] • Usability [47,66,67] • Prototype testing [45] • Survey [40,54,65,68,70]
Provider-related	7 (19)	• Content analysis [39,46,51] • Interview [48,50,71] • RCT [53]
System-related	28 (76)	• Observational [55,63] • Content analysis [44,46,51] • NRT [72] • RCT [42,53,63] • Cohort [61] • Interview [44,47,50,66,68,71] • Prototype testing [45] • Usability [37,41,48,64,66,67] • Survey [40,54,65,68]
Impact on patient care	26 (70)	• Observational [29,38,42,63] • RCT [42,53,59,63] • NRT [52,72] • RT [57] • Cohort [43,60] • Interview [44,47,48,50,68] • Content analysis [39,44,46,51,69] • Usability [47] • Survey [40,54]

[a]The total number of records exceeds the total number of included studies because records contributed to more than one category.

[b]RCT: randomized controlled trial.

[c]RT: randomized trial.

[d]NRT: nonrandomized trial.

Actual Use Information

Few Registered Users Enter Core Medical Data

Figure 3 and Table 4 display the distribution of the core medical data components managed (entered, modified, or updated) by the patients in the included records. In more than half (25/37, 68%) of the included records, patients performed predefined data management tasks in which the usability of the tool or the effects of patients' data management on data quality were explored, and 3 records explicitly reported that their patients wanted to update more information than they were allowed to [40,44,45]. Reviewing the 13 papers in which patients' data management was not constrained by task demands [41,46,49,53-56,58,61,62,65,66,70] showed that the percentage of patients making changes to their core medical data ranged from 0.2% [46] to 22% [54] of registered users. Patients appreciated having insight into their recorded data but were otherwise not adding or updating this information [46,56]. A study investigating the number and content of amendment requests showed that over a period of 6 years, the number of

patients requesting changes to their core medical data was extremely small relative to the number of patients requesting access to their patient records (0.2% of the access requests) [46]. Even when patients did request changes to their medical records (N=818), these changes were mostly related to clinical notes (308/818, 37.7%) and discharge summaries (84/818, 10.3%) [46] and not to the core medical data components (eg, admission history and physical; 19/818, 2.3%). In line with this, studies have shown that portal features that only allowed patients to view their medical information [54,61,62,70] or to message their health care provider [41,54,56] were more frequently used than features that allowed the self-entry of medical data. These passive features were valued more than self-entry features [41,54]. When patients did use self-entry features, they seemed to prefer to enter information about their vital signs (eg, blood pressure, blood glucose values, and weight) compared with other core medical data components [41,49,53,55,65,70].

Figure 3. Distribution of the core medical data components managed (entered, updated, and modified) by patients. PEHR: personal electronic health record.

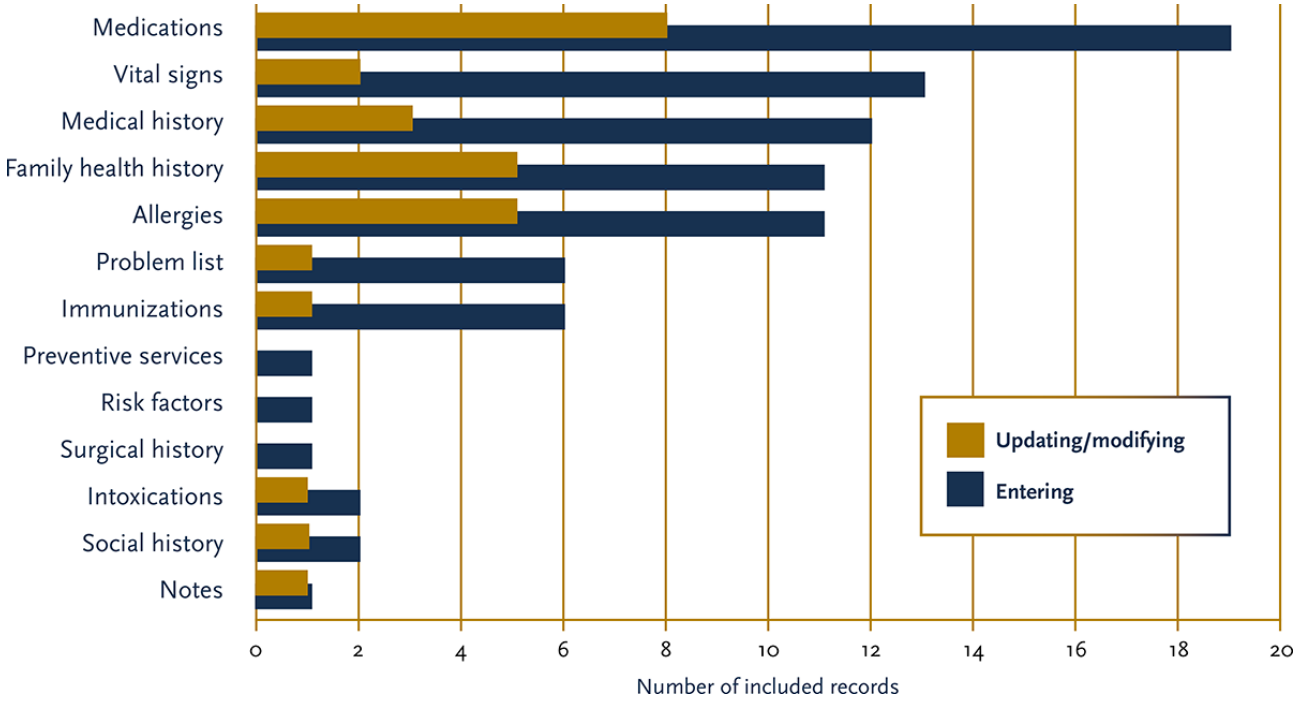


Table 4. Distribution of core medical data components managed and associated tasks across the included records.

Data component and activity, constrained or unconstrained by task demands	Records, n (%)	References
Generating core medical data (entering and sharing data)		
Constrained	24 (64.8)	[29,37-40,42-45,47,48,50-52,57,59,60,63,64,67-69,71,72]
Unconstrained	13 (35.1)	[41,46,49,53-56,58,61,62,65,66,70]
Medications		
Constrained	12 (32.4)	[42,44,47,48,51,57,59,63,64,67,71,72]
Unconstrained	7 (18.9)	[53-55,58,61,62,66]
Vital signs		
Constrained	5 (13.5)	[38,48,59,63,71]
Unconstrained	8 (21.6)	[41,49,53-56,65,70]
Medical history (including personal history)		
Constrained	8 (21.6)	[29,37,39,42,48,63,67,71]
Unconstrained	4 (10.8)	[46,53,55,61]
Family health history		
Constrained	10 (27)	[29,40,43,50,52,59,63,67-69]
Unconstrained	1 (2.7)	[46]
Allergies		
Constrained	5 (13.5)	[42,51,59,63,67]
Unconstrained	6 (16.2)	[49,53-55,61,62]
Problems list (including symptom diary and health conditions and issues)		
Constrained	4 (10.8)	[29,42,64,72]
Unconstrained	2 (5.4)	[53,62]
Immunizations		
Constrained	5 (13.5)	[39,45,48,60,67]
Unconstrained	1 (2.7)	[53]
Preventive services		
Constrained	0 (0)	—
Unconstrained	1 (2.7)	[53]
Risk factors		
Constrained	1 (2.7)	[29]
Unconstrained	0 (0)	—
Surgical history		
Constrained	1 (2.7)	[39]
Unconstrained	0 (0)	—
Intoxications		
Constrained	1 (2.7)	[39]
Unconstrained	1 (2.7)	[46]
Social history		
Constrained	1 (2.7)	[39]
Unconstrained	1 (2.7)	[46]
Clinical notes, discharge summaries, and emergency department notes		
Constrained	0 (0)	—

Data component and activity, constrained or unconstrained by task demands	Records, n (%)	References
Unconstrained	1 (2.7)	[46]
Managing core medical data (updating, modifying, and requesting changes to data)		
Constrained	8 (21.6)	[40,44,47,51,57,59,63,67]
Unconstrained	2 (5.4)	[46,62]
Medications		
Constrained	7 (18.9)	[44,47,51,57,59,63,67]
Unconstrained	1 (2.7)	[62]
Vital signs		
Constrained	2 (5.4)	[59,63]
Unconstrained	0 (0)	—
Medical history (including personal history)		
Constrained	2 (5.4)	[63,67]
Unconstrained	1 (2.7)	[46]
Family health history		
Constrained	4 (10.8)	[40,59,63,67]
Unconstrained	1 (2.7)	[46]
Allergies		
Constrained	4 (10.8)	[51,59,63,67]
Unconstrained	1 (2.7)	[62]
Problem list (including symptom diary and health conditions and issues)		
Constrained	0 (0)	—
Unconstrained	1 (2.7)	[62]
Immunizations		
Constrained	1 (2.7)	[67]
Unconstrained	0 (0)	—
Intoxication		
Constrained	0 (0)	—
Unconstrained	1 (2.7)	[46]
Social history		
Constrained	0 (0)	—
Unconstrained	1 (2.7)	[46]
Clinical notes, discharge summaries, and emergency department notes		
Constrained	0 (0)	—
Unconstrained	1 (2.7)	[46]

Continued Use Drops as Time Increases

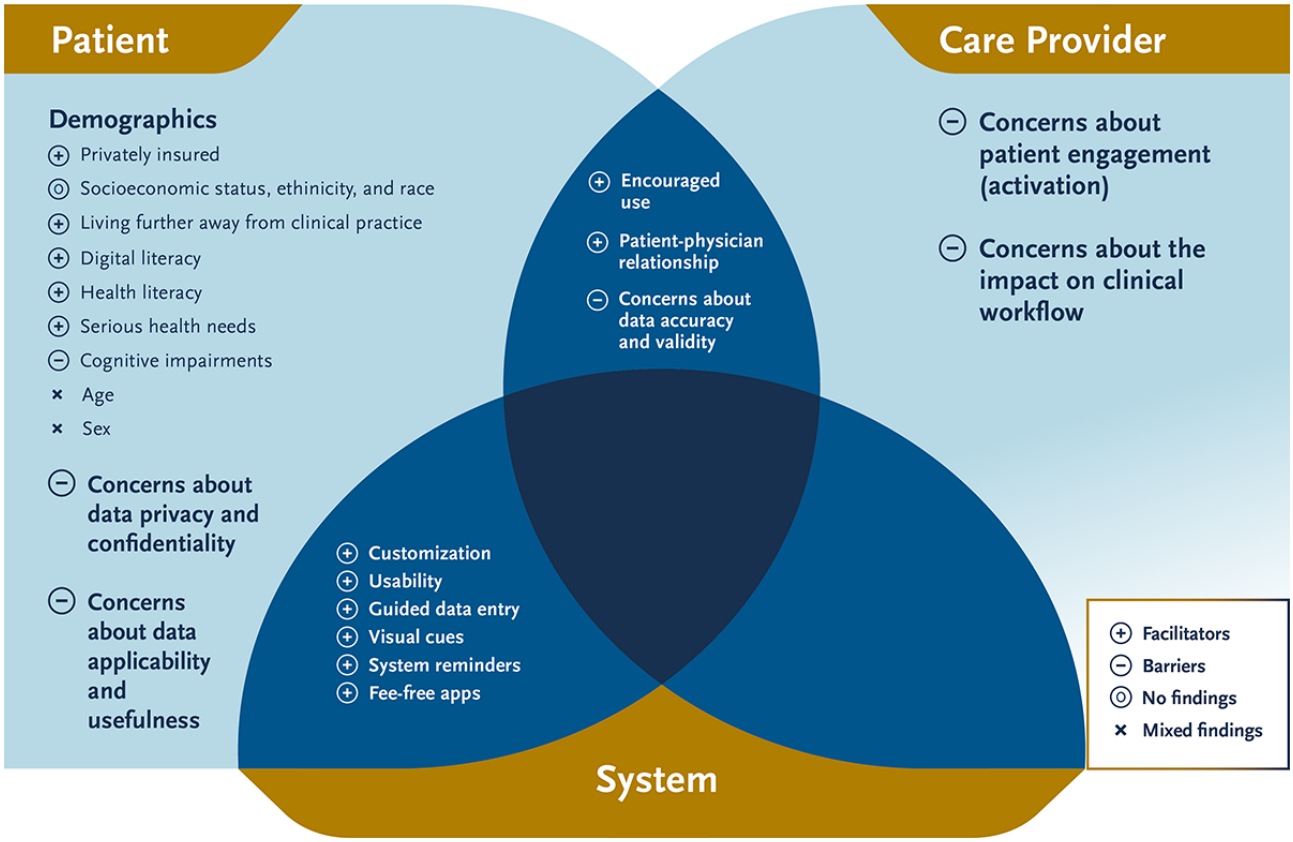
Of the 37 included studies, 23 (62%) provided information about the frequency of patients' portal uptake [38-40,42-47,49-51,53-58,61-63,65,70]. Most of the sample (>50%) used the portal's features [42,47,53,54,70] or specific tools [57], such as an app [43], electronic journal [63], or a computer terminal kiosk in the lobby [51], to enter or update their core medical data in only 9 (24%) of these records. In the remaining studies, a minority of patients (ranging from 0.04% to 44.16% of the population) used the portal's features [45,46,49,55,56,58,61,62,65], an implemented flow sheet [38], a questionnaire [39], a feedback form [44], or a family health history module [50] to manage their core medical data. Most of these records identified patients' use patterns at a specific time point, and only 19% (7/37) of the records explicitly considered patients' frequency of portal use over time [42,49,53-55,61,70]. These latter studies showed that although active portal users usually have more multiple inputs than passive users [42,49], continued use is very limited. Users who

manage their data for longer than a year represent only 5% to 9% of the user population [42,53-55,61], and continued use further decreases as time increases [45,55,61,70]. In the remainder of this paper, we explore what prevents patients from actively managing or helps patients to actively manage their core medical data.

Factors Affecting Active Data Management

We categorized the facilitators and barriers associated with patients actively managing their core medical data through a patient portal into one of the three categories: those dealing with patient characteristics, those dealing with health care provider characteristics, or those dealing with system characteristics. A brief overview of how the important factors affecting patients' personal data management are related to each other is presented in Figure 4.

Figure 4. Patient-related, health care provider–related, and system-related factors affecting patients' management of their personal and medical data.



Patient-Related Determinants

Overview

We identified the following 6 themes that determined whether patients entered, updated, or modified their core medical data: patient demographics; digital and health literacy; concerns related to the accuracy, validity, privacy**,** and confidentiality of recorded data; misconceptions about the applicability; and usefulness of patient-entered data.

Patient Demographics

There is little consensus on whether and how a patient's age or sex influence active data management. While 6 retrospective studies indicated that younger patients are more likely to manage their core medical data [38,42,49,58,61,65], 4 similar studies showed the exact opposite pattern [55-57,62]. In all records, comparisons were predominantly made within rather than across age categories. Taken together over all included records, we see that the age of active portal users ranges from approximately 30 to 70 years [38,42,55,61,62,65], with the most active users being more likely to be in their 30s or 60s [62]. In terms of patients' sex, in 4 retrospective studies, active portal users were more likely to be male than female [42,49,55,61], but 2 other similar studies showed the opposite [62,65]. Thus, age and sex are not very indicative of patients' level of involvement in the generation and management of their core medical data. It may be more informative to look at other patient demographics.

A total of 5 (13.5%) retrospective studies showed that compared with inactive or less active users, active portal users are more likely to be privately insured [58], to have a higher median household income and education level [61], to live farther away from a clinical practice [56], or to reside in urban centers [49,61]. Furthermore, 3 retrospective use pattern studies did not find any significant differences in socioeconomic status, race, or ethnicity of active versus nonactive users [38,42,62]. In 2 other retrospective studies [42,57] and 1 cluster randomized controlled trial [53], active users were found to be digitally competent with a computer or tablet and were already using technology to improve their health [53]. In addition, 3 retrospective user evaluations showed that active users wanted to ensure that their provider had the most accurate and complete information [40] and reported to have already managed their

medical data offline [42] or on the web [65]. We also found that active use might depend on patients' medical condition and health needs, as user pattern studies have shown that active users have a more serious health condition [38,42,53,56,57,61,70] and more clinical encounters [38,62] than other users. In a related vein, a randomized pilot study showed that active users were more interested in improving their understanding of their medical problems and treatments [54]. A usability study showed that cognitive impairments (eg, Alzheimer disease and dementia) and physical limitations (eg, hearing and vision impairments and joint diseases) negatively affected patients' ability to independently manage their medical data in an electronic system [67].

Digital and Health Literacy

Limited internet or computer access, digital illiteracy, and computer anxiety are barriers to patients entering and modifying their core medical data electronically [67,68]. Interviewed users of a web-based family health history tool reported that a lack of knowledge about how to use a computer or web-based technology might limit patients' ability to manage their data electronically without assistance, especially when tasks become more complex [68]. In addition, older adult patients with disabilities reported that their lack of understanding or knowledge of the terminology used for core medical data and how they should report it prevented their data entry [67]. This negative impact of health literacy on active data management was also addressed by interviewed primary care physicians evaluating another implemented family health history tool [50] and by patients recording their family health history in a retrospective data analysis [69] and a user evaluation study [40].

Concerns About Data Accuracy and Validity

An interesting factor that might explain whether patients manage their core medical data is their belief and reassurance that they are not bypassing clinical staff by directly entering or modifying their data in their record [44,45,66]. Patients with multiple morbidities [66] and patients with diabetes or parents managing asthma for their children [45] reported that they preferred having health care providers updating their medical record on their behalf, in fear that their own modifications might alter their physicians' information. In addition, interviewed patients with chronic conditions (ie, chronic obstructive pulmonary disease, asthma, hypertension, diabetes, or heart failure) who were reviewing and modifying their medication list indicated that they found it reassuring to know that all recommended changes were first checked by their provider before they were actually recorded in their medical records [44]. This reassurance can be corroborated by implementing visual features or cues into the interface that convey that patients are modifying personal information that is independent from their physician's records [66]. Patients might also fear that they will provide inaccurate information to their caregivers because they cannot reliably recall medical information such as their family health history [40,43]. Patients who generated their family health history using prepopulated questionnaires stressed that they wanted to include this uncertainty in their records, explicitly stating that they would be more willing to share medical information if they could provide more contextual information to the reported data [40].

Concerns About Data Privacy and Confidentiality

Concerns about data loss and breach of privacy further prevent patients from maintaining their medical records electronically [40,65,68,71]. Patients seek the assurance of data confidentially and protection of their privacy. In a focus group interview, health care providers voiced that patients might fear that their identity might be stolen or that they might purposely omit medical information in fear that it might affect their health insurance or future employment [71]. This concern was indeed confirmed by patients evaluating an implemented family history module in a survey [40] and interview study [68] and by a nonclinical population reporting on their experience with web-based health management tools [65]. Owing to privacy and autonomy concerns, patients do not prefer to share identifiable information, such as their relatives' names and ages [40].

Perceived Applicability and Usefulness

(Mis)conceptions about the applicability and usefulness of patient-generated health data may also prevent patients from taking on a more active role in the management of their personal and medical data via a PEHR. As was mentioned by interviewed patients [66] and interviewed health care providers [71], patients may not see the need to manage their medical information in a web-based portal, as they assume that their providers have access to and share more medical information among specialists than they actually do. Moreover, patients reported that not knowing the benefit of managing and updating medical information [65] or not knowing whether their health care provider actually used the information and found it to be useful [63] prevent their active participation.

Health Care Provider–Related Determinants

Overview

Encouraged use by health care providers and the patient-clinician relationship are identified as the 2 important factors determining whether patients actively manage their core medical data. However, we noticed that health care professionals' recommendations to use the system are dependent on whether they believe that there are benefits associated with patient-entered data in terms of data quality and reliability and cost-effectiveness.

Encouraged Use

Being encouraged by health care providers to manage core medical data plays an important role in the adoption and continued use of PEHRs among patients. First, in both a qualitative content analysis of patient-initiated amendment requests [46] and in a retrospective use pattern study by Ancker et al [38] in which patients managed their blood glucose values, it was suggested that the low amount of generated data was caused by patients not knowing whether they could make changes to their records or how they should go about it. Second, most (84%) respondents voiced that they used web-based health management tools because they were recommended to do so by their clinician [65]. Clinicians also realized that their own

recommendations are important and that reminding patients to use the tools is an important activator of portal use [53]. Clinicians even went so far as to suggest that portal use could be a prerequisite for receiving regular care [53]. In addition, showing the added value of patient-generated health data during an outpatient visit might stimulate patient participation [45,65,67]. Patients with multiple morbidities in a retrospective user pattern study indicated they would stop using tools to record and maintain their core medical data if they did not have someone showing them how to use them, especially when they found it to be difficult to use the tools [65]. In particular, older patients with disabilities both seek and need assistance when it comes to entering and modifying their electronic core medical data [67].

We identified several beliefs that health care providers have about patient-generated and patient-managed medical data that may determine whether they are likely to encourage or assist their patients in managing their core medical data in their PEHR. First, health care providers are often unaware of the benefits that are associated with patients' management of their own data [71]. Second, health care providers do not believe that their patients are motivated [71] or able to provide and maintain accurate and reliable information [44,48,71]. Moreover, health care providers may believe that reviewing patient-entered data may have a significant impact on time spent on outpatient visits and practice workflow [39,46,48,50]. Interviewed physicians who treated patients with spinal cord injuries and disorders voiced concerns that a patient's medical and emotional state may affect their ability to record their data in a reliable fashion and that if patients misinterpret data retrieved from the portal, it might negatively affect their own documentation [48] or treatment information [71]. Pharmacists [44] and family physicians [71] were also skeptical about their patients' ability to enter core medical data accurately. Physicians of a family medicine department explicitly voiced concerns that patient-entered data might be subjective and that health care providers should, therefore, always be in control of data input. Physicians stated that their patients may not know what is appropriate to put in their health records, causing them to enter information that is verified by a professional. They even believed that allowing their patients to enter information into their medical records might facilitate narcotics abuse because patients could inappropriately request or elicit prescriptions [71]. Furthermore, the time saved by having patients enter their own data may be counterbalanced by the time it takes for providers to review patient data [39,46]. Health care providers who treated patients with spinal cord injuries and disorders stressed that checking the patient portals impacts their time and workflow [48]. This view was shared by health care providers who specialized in internal (family) medicine, obstetrics, and gynecology in a study that explored their initial experiences with a family history screening tool implemented in a patient portal. Physicians reported a lack of time for using the tool and stressed that patient-generated and -managed data may only benefit their workflow if patients are able to fill out all the information before their outpatient visit [50].

Patient-Clinician Relationship

Patients testing a medication reconciliation tool via a secure messaging feature within the portal indicated that they appreciated the possibility of communicating directly with health care providers when they had questions about their medications or wanted to request refills. Most (90%) users said they would use the tool again, frequently emphasizing how it allowed them to have instant access to their health care provider [47]. On a related note, patients may refrain from managing their medical data if they want to avoid communicating with their clinicians. Patients with diabetes and parents managing asthma for dependent children voiced that they would rather not use the secure message feature when they did not trust or like their health care provider [45]. This study recommends design implications for the portal that could amplify the positive aspects of the patient–health care provider relationship, such as profile pictures accompanying health care providers' messages or allowing patients to view or hide profiles from a care team in the portal.

System-Related Determinants

Overview

Patients' satisfaction with the system used to collect and maintain their core medical data is an important factor that stimulates active data management [44,64]. A total of 6 main themes emerged from the data extraction that concerned system-related facilitators and barriers affecting patients' satisfaction with the tools used to record their medical core data: the level of customization, usability of the system or tool, guided versus free data entry, presence of visual cues, reminders, and fee-free access to the system/tool.

Customization

A total of 4 studies stressed the importance of offering a level of customization to patient portals [45,63,64,66]. To increase the usability of the system, patients could be allowed to prioritize frequently used portal features [45,63] by, for instance, adding these features to the front page of their portal [45]. Patients also prefer to personalize the system by assigning a personally selected background [64] or self-selected icons for portal features [66], increasing or decreasing the size of these buttons/icons [64], and changing the background and text colors to improve the readability of the portal [64].

Usability

Patients' (continued) use of their electronic patient portal to generate and update their core data depends on the perceived complexity and thus the usability of the system or tools used [37,45,47,63,64,66,68]. Failure to record and maintain core medical data might result from patients not finding the area where it should be recorded [45] or because patients might misinterpret medical terms or encounter terms within the portal that they do not understand, causing frustration and self-doubt [37]. In general, participants prefer to have clear on-screen instructions and directions [53,64,66,68] and short drop-down menus [53]. Using thematic colors also improves the usability of a system [64]. Patients also prefer to have access to previously entered data and to be allowed to mark this information as

unchanged when updating their core medical data in the system [63].

Guided Data Entry

Unless patients are being asked to enter information about simple diagnoses or prescriptions, systems should use guided entry of data elements [55,66,72]. Patients in 5% (2/37) of studies experienced problems during medication reconciliation when asked to enter their medication names into the system [55,66]. It was for this reason that they were reluctant to provide additional dosage and scheduling information [66]. Patients prefer a less textual way of adding medications to their list, voicing that free-text entry is too complex and time-consuming [66]. To aid the reviewing process, a prepopulated medication form [55] or a barcode scanning function [69] could be used, especially when patients need to report on a large number of medications [55]. Autofilling processes also give patients some reassurance about the accuracy of their data entry [66]. Free-text entries are undesirable when patients are asked to add information to their problem list, as they may be inclined to include extraneous information that does not contribute to the identification of a primary diagnosis [72]. However, in a study exploring patients' experiences with a family history tool [40], patients reported on the danger of using closed answer options. The patients expressed concerns that some answers did not allow for sufficient granularity and reliability, arguing that their family history was often far more complex than what they were allowed to record. These patients also preferred to receive more clarity and information about the diseases that they were asked to report. Allowing patients to provide contextual information when they have the desire to do so might reassure them about their answers' validity [40].

Visual Cues

Implementing visual feedback facilitates data entry by patients and patients' satisfaction with using the system. For instance, providing medication pictures alongside a selected medication assists patients' medication reconciliation [51] and allows them to confirm whether it is the correct medication to add [66]. In addition, patients prefer to receive clear feedback when performing an action within the system, such as seeing a medication being highlighted after they suggest it should be deleted from their list [66]. Visual feedback in the form of using red and green colors also helps patients to take further actions such as scheduling alerts to take the medication when a new medication is added to the list [64]. Using colors is also beneficial when they are used to demarcate separate body parts, helping patients to correctly specify the location of the problem skin areas [64].

Reminders

If reminded to do so, patients are more likely to use the portal before and after their outpatient visits [26]. Reminders generated through the portal stimulate patients to access their records [26] and enter information about their medications, allergies, and vital signs [54].

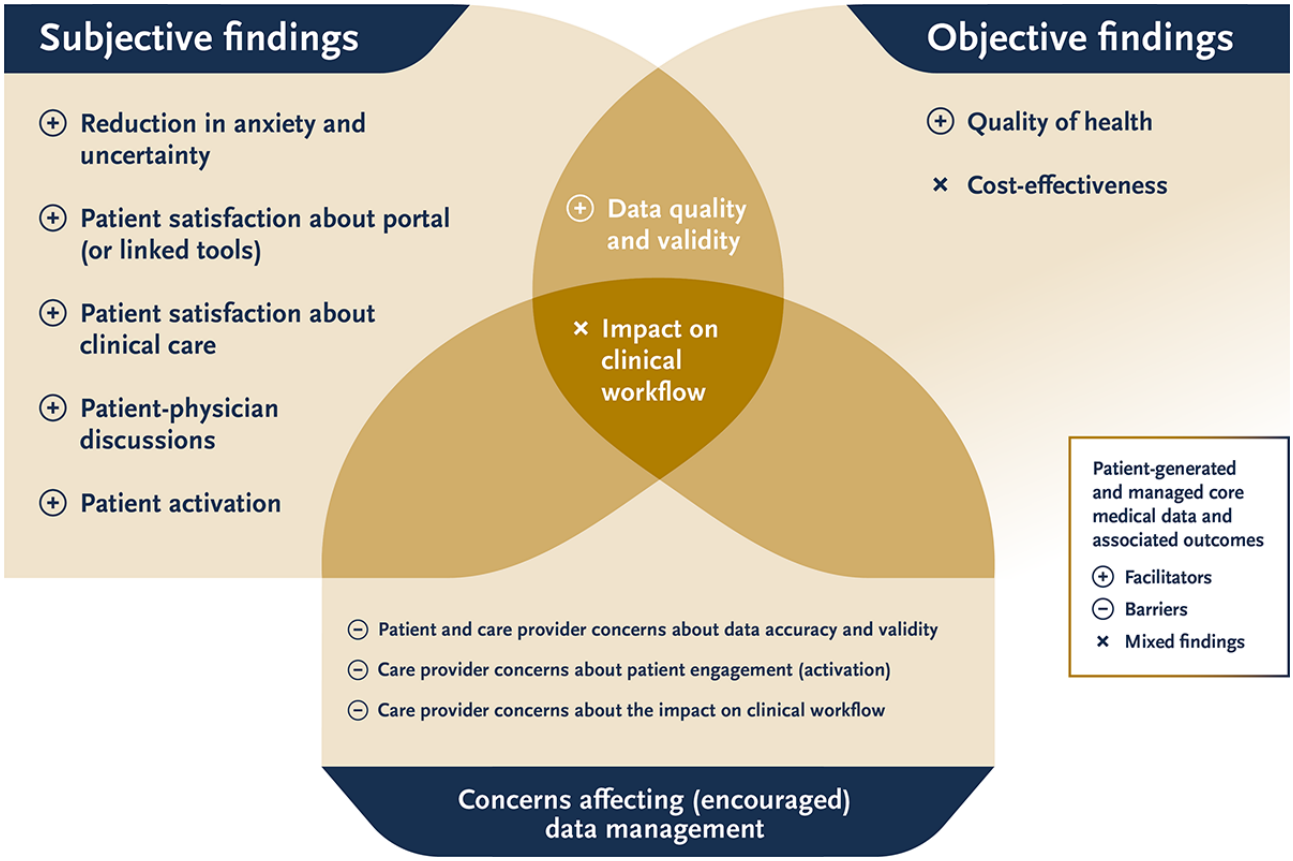
Fee-free Apps

Providing applications without charge [41] that can be downloaded by patients as well as by a more general group of users [55] stimulates the accumulation of patient-generated core medical data. A study that focused on patients' diabetes management [41] showed that patients believed that implementing fees for portal access would significantly reduce their tendency to use the portal for the self-management of their diseases. The implementation of portal fees seemed unfair according to patients because health systems also benefit from patients' self-management of their disease. Patients believed that introducing fees would increase inequities between patients who can and cannot afford using the portals, and they also feared that costs would increase when previously free services would start requiring payment [41].

Impact on Patient Health and Health Care Services

This section describes the impact of patients' data management on the quality and safety of patient care, psychological outcomes for patients, patient engagement, patient satisfaction, and clinical workflow. Figure 5 presents the important subjective and objective outcomes identified and how they are related to the concerns of both patients and health care professionals.

Figure 5. Impact of patient-generated health data (PGHD) on patients' health and health care–related services and how this impact is associated with the important concerns regarding PGHD raised by patients and health care providers.



Data Quality and Validity

Clinicians' concerns about the quality and validity of patient-entered data seem to be unfounded. Observational [29], experimental [52,57,72], usability [47], cohort [60], and content analysis [44,46,69] studies have shown that medical records are completer and more accurate when the data are generated by patients themselves. Patients are able to accurately self-report on their diagnoses [29,72], medications [29,44,47,57], medical or surgical history [46], family health history [52,69], or their children's vaccination history [60]. Patients request changes to their core medical data especially when this information is incomplete [46,47,59] or incorrect [46], and these requests are approved in approximately half [46] up to 80% [44] of cases. Studies have reported on improved medication reconciliation [44,47,51,57,59], arguing that patients' management of their medical data makes them more attentive to medication safety and monitoring [42,44,47] and even helps clinicians to identify (potential) lethal medication discrepancies [51]. In addition, the quality and validity of patients' problem lists [29], immunization records [60], and family health history [43,52,69] improves when patients enter and manage their own medical data. Clinicians even felt that the risks identified because of patients entering their family health history helped them to make informed changes to their patients' medical management [50]. Pharmacists reported being surprised to learn about patients' willingness and ability to report their medications accurately, even when patients were taking >20 medications or were taking medications that had been prescribed by physicians who were not part of the current health system [44]. Only 1 content analysis study did not show the added value of patient-generated data [39]. In this study, patients entered information about their medical, surgical, and social history, using closed question questionnaires with "yes" and "no" answer options. Patients were allowed to give additional information in the comments section. The researchers concluded that the new information added to a patient's record often lacked sufficient granularity to be found meaningful. However, they did not reflect on how the closed nature of the questionnaire could have contributed to this outcome.

Quality of Health

Another theme we identified was a significant objective [38,53] and subjective [42,63] improvement in patients' health because of them actively managing their medical data. First, an observational study of patients with diabetes who were uploading (and thus tracking) their blood glucose values showed a significant drop in their average BMI and mean glycated hemoglobin values compared with nonuploaders (nontrackers) [38]. Second, patients who entered and tracked their vital signs and preventive services were more likely to receive all recommended immunizations than control groups [53]. These objective findings are corroborated by patients' self-reports [42,63]. Older adults reported more changes in medication use and improved medication reconciliation behaviors than less active recorders and nonrecorders. These patients also reported more side effects [42]. In a similar vein, patients in primary care who entered and modified their lists of medications and allergies felt that their health care provider had more accurate information

about them and that this improved the quality of care at the visit [63].

Psychological Outcomes for Patients

Insight into medical data might reduce anxiety and uncertainty in patients. This point was explicitly raised by interviewed health care providers who were evaluating a tool that helped the patients under their care to report on their family health history to identify possible genetic diseases [50]. Patients felt less anxious when the tool identified no increased risk and they were able to discuss the findings with their clinician.

Patient Engagement

We identified two themes in this subsection: (1) the extent to which patients' data management improves patient-physician discussions and (2) feelings of ownership among patients and future patient participation.

Improved Patient-Physician Discussions

Patients who update their core medical data before an outpatient visit, feel better informed [44] and better prepared for the visit [44,63,70] and experience improvement in their interaction with their health care providers [50-52,54,59,63,65,70]. Patients indicate that they can provide more comprehensive information about complex and sensitive health issues at home than in their physician's office because in the latter case, they feel more stressed and uncomfortable [40]. Patients [43,52] and primary care physicians [50] believe that patients who update their family health history are more aware of its (medical) importance, facilitating both patient-physician [50,52] and patient-family [43,50] discussions about associated family history–related health risks and ways to improve their health. Patients who manage their vital signs data prepare their questions before visiting their provider [70], thereby improving treatment-related discussions and decisions [65,70]. Regarding medication reconciliation, nurse practitioners mentioned that allowing patients to review, update, and modify their medication lists improved their medication dispensing information and identification of errors [51,59]. In their turn, practitioners [51] and patients in primary care [59] stated that patients asked more questions about their regimens [51], were more likely to report adverse reactions [51] or to address medication-related problems and new symptoms [59], and requested more refills for medications that were nearing their expiration date [51]. Active patients feel more confident when asking questions about medications during their outpatient visits [44], and they recall more questions that they want their physicians to answer. Patients also feel that such preparation saves time during the visit [63] or even reduces the need for an outpatient visit [44]. This viewpoint is shared by primary care clinicians, who stress that they would recommend that other clinicians ask their patients to review, update, and modify their list of medications, allergies, and diabetes items before an outpatient visit [63].

Patient Activation

Patients who generate and manage their own medical data feel that they have more control over their health care and health-related decisions [40,44,53,65,70,71]. A randomized controlled trial comparing patients who managed their core medical data against nonactive patients showed that active patients were not only more confident and knowledgeable about their health in general and about making health-related decisions but were also more likely to actually take action to improve their health [53]. These findings are supported by studies that focus on patients who managed their family health history [40,68], vital signs [65,70,71], medical history [71], and medications [44,71]. Patients feel that their participation improves their clinician's knowledge [40,70]. Patients experience a sense of ownership when they manage their own medical data [70] and report that they consider their contributions to be valuable to an extent that makes them feel empowered [40] and motivated [68] to improve their health condition. This viewpoint is shared by family physicians [71] and health care providers who treat patients with spinal cord injuries and disorders [48]. These clinicians feel that if patients maintain their medical data, they may become more organized and adherent to medications [48] and improve their involvement in their care, which may result in better outcomes [71].

Patient Satisfaction

Patients were generally satisfied with the tools that they used to update their medical data [43,63,64,68]. Only 2 records discussed whether active management of data by patients affected patients' satisfaction with their clinical care [40,63]. One of these records measured patient satisfaction using a 1-item survey question [63], showing that 37.7% of the respondents were more satisfied with their visit after they had first entered or updated their medical information using electronic journals implemented in a patient portal. The second study found that their patients were more satisfied with reviewing their free-text responses after they had entered or updated their family health history in their web-based records [40]. In the comment section of that study [40], patients reported that they felt welcomed, cared for, and safe when asked to share their medical information.

Impact on Clinical Workflow and Costs

A study that interviewed health care providers who treated patients with spinal cord injuries and disorders found that health care providers believed that patient-generated health data collected via patient portals can improve the coordination of medical care, especially for those patients who receive health care in nonclinical settings [48]. However, we found mixed evidence concerning the effects of patients' active management of their medical data on clinical and patient throughput. Both clinicians [57,70] and patients [63,70] believed that asking patients to review and update their medical data before an outpatient visit positively affects clinical throughput because consultations can be executed more efficiently. For instance, pharmacists and physicians stated that they spent half of the usual amount of time on medication reconciliation on outpatient visits when patients generated this information themselves [44]. Active involvement of patients in the generation and management of their data may even reduce the need to schedule an outpatient visit [44], especially when physicians can address their patients' questions via a secure messaging feature [48]. However, interviewed family physicians were concerned that patient-generated data would negatively impact consultation

time if it required logging in and searching for relevant information [71].

We identified only 4 records that objectively measured the cost-effectiveness of patients' data management. A retrospective cross-sectional study investigating the impact of patients updating their medication list via a secure message feature showed that its use did not significantly decrease the cost burden of outpatient clinics [58]. However, another retrospective study found that asking patients to review and update their medical history via a computer terminal kiosk in the waiting room of a chemotherapy clinic reduced the medication reconciliation time by nearly 50% [51]. A retrospective longitudinal cohort study also found that active portal users were less likely to contact or visit their health care providers [61], whereas another retrospective analysis of portal use showed that nonusers visited the emergency room more often than active users, even though active users had more outpatient and inpatient visits [62].

Discussion

Principal Findings

This synthesis of literature explored the barriers and facilitators that patients face when they decide to generate and manage their core medical data in (tools linked to) their PEHRs. First, we found that a minority of registered users entered, updated, or modified their personal and medical data. More specifically, less than half of the registered users entered their data and less than a quarter of users updated or modified their already recorded data; continued use further dropped to <10% of the user population as time increased. Patients preferred to take on a passive rather than an active role regarding the self-management of their health information, and they seemed to prefer tracking vital signs above more complex medical information, such as medications and their family health history. We identified both patients' and health care professionals' (positive) perceptions about the validity, applicability, and confidentiality of patient-generated data as well as patients' digital and health literacy as important facilitators of patients' active management of their personal and medical data. However, we also found that patients' and health care providers' concerns about the validity and applicability of patient-generated data seem to be unfounded. Patients accurately reported on their diagnoses, medications, immunizations, medical history, and family health history, making their medical records more complete. Moreover, patients who managed their medical data felt more knowledgeable, more in control of their own health care, and more adherent to their treatment than less active patients. Both patients and clinicians felt that active patients were also more prepared for their clinical visits because they knew which questions they wanted answered by their health care provider. In the following sections, we propose recommendations that health care practices can adopt for stimulating patient participation in the generation and management of their electronic core medical data.

The Health Care Provider as Ambassador and Gatekeeper

Patients felt that they were bypassing clinical staff when they self-managed their medical data. Patients were concerned that they would provide their physicians with inaccurate information, especially when the nature of the medical information is complex and sensitive. Clear guidelines and information regarding the added value of patient-entered data for both patients and clinicians may reduce these concerns. Clinical staff are important ambassadors for informing their patients about the added value of patient-generated and management data and in reminding and encouraging their patients to prepare themselves for each visit by reviewing the medical data in their PEHRs. Moreover, we also found that self-management of medical data may be higher for those patients who feel that they are able to directly contact their provider for support. Design features within the PEHR systems that amplify the visibility of the health care providers' availability for support and guidance as well as visual feedback elements in the PEHR system that indicate to the patients that their entered or modified data will be checked by a professional may reassure patients that they are not altering their medical record without their provider's knowledge or approval.

Ethical and Comprehensive by Design

We also found that patients were generally concerned that their medical data were unprotected against unauthorized access and could, therefore, be used for non–health care–related purposes. Stressing data confidentiality and allowing patients to give their informed consent on an opt-in and opt-out basis may diminish their potential unease about confidentiality. Furthermore, we have also seen that customization features may enhance the self-management of core medical data because they make the system more understandable and easier to use. Helping patients to remember medical information by using prepopulated forms or guided data entry might further aid and encourage them to record information that might be inaccurate. This may also address health care providers' concerns that patients are not able to accurately report on their medical information.

Future Directions

On the basis of our findings and recommendations, we have outlined several priority questions for future studies (Textbox 1) that we address briefly in this section. The first 2 questions are related to the finding that health care providers play an important role in their patients' uptake and continued use of (tools linked to) their PEHRs to manage their core medical data. It is still not known what providers need for addressing their concerns about the validity and applicability of patient-generated data. Thus, we invite future studies to explore the needs of professionals in terms of (portal) assistance or (system) requirements so that they are willing to encourage the practice of patients' self-management medical data and their patients feel stimulated and supported to manage their core medical data during their entire care journey as a result.

Textbox 1. Priority questions for future research based on our 3 recommendations.

1. The health care provider as ambassador and gatekeeper

- What are the unmet needs of health care professionals with respect to encouraging and supporting their patients to share and manage their personal and medical data during their care journey?
- What are the unmet needs of patients in terms of feeling encouraged and supported by their health care providers to share and manage their personal and medical data during their care journey?

2. Ethical and comprehensive by design

- What do patients need in terms of assistance, support, and system requirements, to generate and manage their personal data during their care journey?
- To what extent does the type of personal and medical data affect patients' data management?

3. Stimulating the patient-provider partnership

- When do patients consider themselves to be "active" managers of their personal and medical data, and to what extent does this correspond to health care professionals' perspectives?
- To what extent do patients' perspectives on their personal data management activity and role preference affect their data management?

For fear of reporting inadequate information, patients prefer to report their core medical data in a structured, guided manner. Our review showed that this was the case for data that were perceived to be error-prone and sensitive, such as information about the types, names, and dosages of patients' medications or information about patients' family health history that would be used for genetic counseling. This finding corresponds to the findings of Esmaeilzadeh et al [73], who showed that individuals were more willing to share sensitive and private information about their mental or physical illnesses when they could enter this information by following a structured, organized, and predefined data entry model, as opposed to using an unstructured, text-heavy interface [73]. Taken together, this seems to indicate that guided data entry interfaces may stimulate patients to share personal health information they would not otherwise share because they do not feel confident or knowledgeable enough to share it or because confidentiality or privacy concerns prevent them from doing so. However, we also found that in case of sensitive information, patients may feel that closed answer options do not offer sufficient granularity and feel the need to add additional contextual information to their answers. Hence, we invite future studies to explore the extent to which patients' preference for structured data entry models is dependent on the type of data that they wish to record.

We have also shown that patients prefer to update and monitor data about their vital signs (eg, blood glucose levels and BMI) over updating information about their medications, allergies, intoxications, and social and family history. To the best of our knowledge, no studies to date have examined the reasons for these differences. On the basis of the findings of our review, we hypothesize that patients prefer to manage data about their vital signs to managing information about other core medical data because they are trackable over time and thereby give patients a more direct, visible insight into their health status compared with other core medical data. We encourage future studies to explore this explanation.

We have shown that the number of studies that focus on actual portal use—by exploring how patients use their portal, whether and when patients consider themselves to be active users, which data patients share, and how frequently they do this—remains scarce. Interestingly, it is not common practice for patient data management papers to describe in full detail whether, how, and how frequently and what type of medical information is entered, updated, or modified by patients. We believe that this is mainly caused by an undifferentiated definition of the term "active user." In the retrieved literature, users were predominantly considered to be active based solely on whether they activated their account [74], the number of times they logged in or accessed a certain page or implemented tool [75], or their self-reported (undefined and abstract) use of the portal [76]. Patients were described to be active when they performed an activity once [40,42,53,56-58,65,67,70], more than once [49], >3 times [38], >20 times [61], or more than once every 4 months [62]. It would be a promising endeavor for future research to define "active data management" from both the patients' and their care professionals' perspectives.

Our findings are in line with research that has investigated the extent to which patients participate in making decisions together with their physicians regarding treatment plans. Shared decision-making entails the collaborative exchange and discussion of health care information among patients and their health care providers, including information about patient preferences and the pros and cons of all possible treatment options [77,78]. Collaboration is the key here [79], meaning that both patients and health care providers are jointly responsible for reducing asymmetries in information exchange so that treatment decisions that patients can adhere to because they optimally align with their wishes and abilities are reached [80]. One line of research claims that not all patients have the desire to participate in decision-making processes [80-82] and that this is especially the case for older and less healthy patients who, ironically, might benefit the most from being involved [83]. Another line of research claims that most patients do in fact want to be informed and involved, but that they cannot fulfill this desire because it is not acknowledged or afforded to them by their health care provider [80,84]. Patients' preferred and assumed roles often do not match [85], leading to decisional role regret [86]. In many cases, physicians do not know their

patients well enough. Patients believe that the medical expertise and knowledge of their health care provider are more important than their own knowledge and preferences. Thus, our advice is to inform patients about the complementary value that they bring to the shared decision-making process and to improve patients' confidence in their capability to acquire and understand the information that is necessary to make informed decisions based on the available options [84,87]. Our literature review showed that these recommendations also apply when clinical staff want to involve patients in the management of their medical data. We invite future studies to explore the extent to which discrepancies in patients' preferred versus assumed roles in the management of their medical data affect their engagement and satisfaction with their clinical care.

Limitations

This scoping review has some limitations. We retrieved a limited set of highly heterogeneous papers because they provided detailed information about patients' actual data management activities. Despite the considerable heterogeneity in the study objectives, designs, and outcome measures used in these papers, we were able to identify key themes regarding the facilitators and barriers that patients face when they decide to generate and manage their medical data. In addition, this review concentrated on measurable uses of PEHRs (ie, entering, updating, and modifying data) to identify what stimulates or prevents patients' use. Although patients who evaluate their core medical data and subsequently decide not to add or modify information are actively engaging with their PEHR, we chose not to include this group because we would then need to rely on log-in frequencies to determine the patients' (level of) engagement with their health data. Not only may log-in frequencies be biased by false log-in data resulting from log-in problems, but they also do not inform us whether a log-in moment resulted in meaningful use of the portal. A promising endeavor for future studies would be to identify whether and how frequently patients review and approve of the core medical data recorded in their PEHR and which factors contribute to this type of use.

Conclusions

Most patients do not actively review and enter, update, or modify their medical data in a PEHR. Patients refrain from generating and managing their medical data, especially when medical information is complex and sensitive. The reasons for patients' passive behavior are their concerns about the validity, applicability, and confidentiality of patient-generated data, although we found that patient-generated data are often accurate and helpful in stimulating patient engagement and satisfaction. We have offered recommendations for implementing design features within the (tools linked to) PEHRs and the creation of a dedicated policy to inform both clinical staff and patients about the added value of patient-generated data, with clinicians being involved as important ambassadors in informing, reminding, and encouraging patients to manage the data in their PEHR.

Acknowledgments

This scoping review is part of the research project Patient-Generated Health Data: Engaging Patients to Improve Shared Decision-making and to Optimize Electronic Health Record Content, funded by the We Care partnership between the Tilburg University and the Elisabeth-TweeSteden Hospital in the Netherlands.

Authors' Contributions

All authors contributed to developing the aim of the scoping review and construing the study protocol. DJD took the lead by developing the search strategy, by retrieving and screening the identified records, by analyzing and interpretating the data for the article, and by drafting the first version of the manuscript. GGS, BM, and SP contributed by screening the identified records. All authors proofread the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Filled-in PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist [31].
[PDF File (Adobe PDF File), 102 KB - jmir_v24i12e37783_app1.pdf]

Multimedia Appendix 2
Search strategy for the MEDLINE, PsycINFO, CINAHL, Cochrane Library, Embase, Web of Science, and Google Scholar databases.
[PDF File (Adobe PDF File), 174 KB - jmir_v24i12e37783_app2.pdf]

Multimedia Appendix 3
Data extraction form.
[PDF File (Adobe PDF File), 83 KB - jmir_v24i12e37783_app3.pdf]

References

1. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Dec 06;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
2. Wenger N, Méan M, Castioni J, Marques-Vidal P, Waeber G, Garnier A. Allocation of internal medicine resident time in a Swiss hospital: a time and motion study of day and evening shifts. Ann Intern Med 2017 Apr 18;166(8):579-586. [doi: 10.7326/M16-2238] [Medline: 28135724]
3. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
4. Lau M, Campbell H, Tang T, Thompson DJ, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
5. Rief JJ, Hamm ME, Zickmund SL, Nikolajski C, Lesky D, Hess R, et al. Using health information technology to foster engagement: patients' experiences with an active patient health record. Health Commun 2017 Mar;32(3):310-319. [doi: 10.1080/10410236.2016.1138378] [Medline: 27223684]
6. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015 Oct;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]
7. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: a systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
8. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
9. Antonio MG, Petrovskaya O, Lau F. Is research on patient portals attuned to health equity? A scoping review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):871-883 [FREE Full text] [doi: 10.1093/jamia/ocz054] [Medline: 31066893]
10. Crameri K, Maher L, Van Dam P, Prior S. Personal electronic healthcare records: what influences consumers to engage with their clinical data online? A literature review. Health Inf Manag 2022 Jan;51(1):3-12. [doi: 10.1177/1833358319895369] [Medline: 31918578]
11. Daulby LM. Predictors of electronic personal health record adoption among health care consumers: a case for “meaningful use” engagement (PART 1). SJSU School of Information. 2014 Apr 14. URL: https://ischool.sjsu.edu/ciri-blog/predictors-electronic-personal-health-record-adoption-among-health-care-consumers-case [accessed 2020-04-22]
12. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient portals: who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501 [FREE Full text] [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
13. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
14. Wildenbos GA, Peute L, Jaspers M. Facilitators and barriers of electronic health record patient portal adoption by older adults: a literature study. Stud Health Technol Inform 2017;235:308-312. [doi: 10.3233/978-1-61499-753-5-308] [Medline: 28423804]
15. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
16. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]
17. Aljabri D, Dumitrascu A, Burton MC, White L, Khan M, Xirasagar S, et al. Patient portal adoption and use by hospitalized cancer patients: a retrospective study of its impact on adverse events, utilization, and patient satisfaction. BMC Med Inform Decis Mak 2018 Jul 27;18(1):70 [FREE Full text] [doi: 10.1186/s12911-018-0644-4] [Medline: 30053809]
18. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
19. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013 Feb 26;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
20. Kerns JW, Krist AH, Longo DR, Kuzel AJ, Woolf SH. How patients want to engage with their personal health record: a qualitative study. BMJ Open 2013 Jul 30;3(7):e002931 [FREE Full text] [doi: 10.1136/bmjopen-2013-002931] [Medline: 23901027]
21. Lam R, Lin VS, Senelick WS, Tran H, Moore AA, Koretz B. Older adult consumers' attitudes and preferences on electronic patient-physician messaging. Am J Manag Care 2013 Nov;19(10 Spec No):eSP7-eS11 [FREE Full text] [Medline: 24511886]

22. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
23. Weitzman ER, Kelemen S, Kaci L, Mandl KD. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. BMC Med Inform Decis Mak 2012 May 22;12:39 [FREE Full text] [doi: 10.1186/1472-6947-12-39] [Medline: 22616619]
24. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 01;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
25. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
26. Whetstone M, Goldsmith R. Factors influencing intention to use personal health records. Intl J Pharm Health Mrkt 2009 Apr 03;3(1):8-25. [doi: 10.1108/17506120910948485]
27. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, facilitators, and solutions to optimal patient portal and personal health record use: a systematic review of the literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
28. McAllister M, Dunn G, Payne K, Davies L, Todd C. Patient empowerment: the need to consider it as a measurable patient-reported outcome for chronic conditions. BMC Health Serv Res 2012 Jun 13;12:157 [FREE Full text] [doi: 10.1186/1472-6963-12-157] [Medline: 22694747]
29. Polubriaginof CG F, Pastore G P. Comparing patient-reported medical problems with the electronic health record problem list. Gen Med (Los Angeles) 2016;04(03):1000258. [doi: 10.4172/2327-5146.1000258]
30. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. Health Aff (Millwood) 2013 Feb 04;32(2):223-231. [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]
31. 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]
32. Damen DJ, Schoonman GG, Maat B, Habibović M, Krahmer EJ, Pauws S. Patients' barriers and facilitators to manage their core health data in an electronic health record. OSF Registries. 2020 Apr 9. URL: https://osf.io/xcuj2 [accessed 2021-11-11]
33. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
34. Apter AJ, Bryant-Stephens T, Perez L, Morales KH, Howell JT, Mullen AN, et al. Patient portal usage and outcomes among adult patients with uncontrolled asthma. J Allergy Clin Immunol Pract 2020 Mar;8(3):965-70.e4 [FREE Full text] [doi: 10.1016/j.jaip.2019.09.034] [Medline: 31622684]
35. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in electronic health record patient portal use in nephrology clinics. Clin J Am Soc Nephrol 2015 Nov 06;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
36. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
37. Ali S, Romero J, Morrison K, Hafeez B, Ancker J. Focus section health IT usability: applying a task-technology fit model to adapt an electronic patient portal for patient work. Appl Clin Inform 2018 Jan 14;9(1):174-184 [FREE Full text] [doi: 10.1055/s-0038-1632396] [Medline: 29539648]
38. Ancker J, Mauer E, Kalish R, Vest J, Gossey J. Early adopters of patient-generated health data upload in an electronic patient portal. Appl Clin Inform 2019 Mar 10;10(2):254-260 [FREE Full text] [doi: 10.1055/s-0039-1683987] [Medline: 30970383]
39. Arsoniadis EG, Tambyraja R, Khairat S, Jahansouz C, Scheppmann D, Kwaan MR, et al. Characterizing patient-generated clinical data and associated implications for electronic health records. Stud Health Technol Inform 2015;216:158-162. [Medline: 26262030]
40. Bajracharya AS, Crotty BH, Kowoloff HB, Safran C, Slack WV. Patient experience with family history tool: analysis of patients' experience sharing their family health history through patient-computer dialogue in a patient portal. J Am Med Inform Assoc 2019 Jul 01;26(7):603-609 [FREE Full text] [doi: 10.1093/jamia/ocz008] [Medline: 30946464]
41. Bryce CL, Zickmund S, Hess R, McTigue KM, Olshansky E, Fitzgerald K, et al. Value versus user fees: perspectives of patients before and after using a web-based portal for management of diabetes. Telemed J E Health 2008 Dec;14(10):1035-1043. [doi: 10.1089/tmj.2008.0005] [Medline: 19119825]

42. Chrischilles EA, Hourcade JP, Doucette W, Eichmann D, Gryzlak B, Lorentzen R, et al. Personal health records: a randomized trial of effects on elder medication safety. J Am Med Inform Assoc 2014;21(4):679-686 [FREE Full text] [doi: 10.1136/amiajnl-2013-002284] [Medline: 24326536]
43. Cohn W, Ropka M, Pelletier S, Barrett J, Kinzie M, Harrison M, et al. Health Heritage© a web-based tool for the collection and assessment of family health history: initial user experience and analytic validity. Public Health Genomics 2010;13(7-8):477-491 [FREE Full text] [doi: 10.1159/000294415] [Medline: 20424421]
44. Dullabh PM, Sondheimer NK, Katsh E, Evans MA. How patients can improve the accuracy of their medical records. EGEMS (Wash DC) 2014;2(3):1080 [FREE Full text] [doi: 10.13063/2327-9214.1080] [Medline: 25848614]
45. Eschler J, Meas PL, Lozano P, McClure JB, Ralston JD, Pratt W. Integrating the patient portal into the health management work ecosystem: user acceptance of a novel prototype. AMIA Annu Symp Proc 2016;2016:541-550 [FREE Full text] [Medline: 28269850]
46. Hanauer DA, Preib R, Zheng K, Choi SW. Patient-initiated electronic health record amendment requests. J Am Med Inform Assoc 2014;21(6):992-1000 [FREE Full text] [doi: 10.1136/amiajnl-2013-002574] [Medline: 24863430]
47. Heyworth L, Paquin AM, Clark J, Kamenker V, Stewart M, Martin T, et al. Engaging patients in medication reconciliation via a patient portal following hospital discharge. J Am Med Inform Assoc 2014 Feb;21(e1):e157-e162 [FREE Full text] [doi: 10.1136/amiajnl-2013-001995] [Medline: 24036155]
48. Hill JN, Smith BM, Weaver FM, Nazi KM, Thomas FP, Goldstein B, et al. Potential of personal health record portals in the care of individuals with spinal cord injuries and disorders: provider perspectives. J Spinal Cord Med 2018 May 21;41(3):298-308 [FREE Full text] [doi: 10.1080/10790268.2017.1293760] [Medline: 28325112]
49. Laranjo L, Rodolfo I, Pereira AM, de Sá AB. Characteristics of innovators adopting a national personal health record in Portugal: cross-sectional study. JMIR Med Inform 2017 Oct 11;5(4):e37 [FREE Full text] [doi: 10.2196/medinform.7887] [Medline: 29021125]
50. Lemke AA, Thompson J, Hulick PJ, Sereika AW, Johnson C, Oshman L, et al. Primary care physician experiences utilizing a family health history tool with electronic health record-integrated clinical decision support: an implementation process assessment. J Community Genet 2020 Jul 04;11(3):339-350 [FREE Full text] [doi: 10.1007/s12687-020-00454-8] [Medline: 32020508]
51. Lesselroth B, Adams S, Felder R, Dorr DA, Cauthers P, Church V, et al. Using consumer-based kiosk technology to improve and standardize medication reconciliation in a specialty care setting. Jt Comm J Qual Patient Saf 2009 May;35(5):264-70, AP1. [doi: 10.1016/s1553-7250(09)35037-0]
52. Murray MF, Giovanni MA, Klinger E, George E, Marinacci L, Getty G, et al. Comparing electronic health record portals to obtain patient-entered family health history in primary care. J Gen Intern Med 2013 Dec 16;28(12):1558-1564 [FREE Full text] [doi: 10.1007/s11606-013-2442-0] [Medline: 23588670]
53. Nagykaldi Z, Aspy CB, Chou A, Mold JW. Impact of a wellness portal on the delivery of patient-centered preventive care. J Am Board Fam Med 2012 Mar 07;25(2):158-167 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110130] [Medline: 22403196]
54. Nazi KM, Hogan TP, McInnes DK, Woods SS, Graham G. Evaluating patient access to electronic health records: results from a survey of veterans. Med Care 2013 Mar;51(3 Suppl 1):S52-S56. [doi: 10.1097/MLR.0b013e31827808db] [Medline: 23407012]
55. Park YR, Lee Y, Kim JY, Kim J, Kim HR, Kim Y, et al. Managing patient-generated health data through mobile personal health records: analysis of usage data. JMIR Mhealth Uhealth 2018 Apr 09;6(4):e89 [FREE Full text] [doi: 10.2196/mhealth.9620] [Medline: 29631989]
56. Powell KR, Deroche C. Predictors and patterns of portal use in patients with multiple chronic conditions. Chronic Illn 2020 Dec 04;16(4):275-283. [doi: 10.1177/1742395318803663] [Medline: 30284917]
57. Prey JE, Polubriaginof F, Grossman LV, Masterson Creber R, Tsapepas D, Perotte R, et al. Engaging hospital patients in the medication reconciliation process using tablet computers. J Am Med Inform Assoc 2018 Nov 01;25(11):1460-1469 [FREE Full text] [doi: 10.1093/jamia/ocy115] [Medline: 30189000]
58. Raghu TS, Frey K, Chang Y, Cheng M, Freimund S, Patel A. Using secure messaging to update medications list in ambulatory care setting. Int J Med Inform 2015 Oct;84(10):754-762. [doi: 10.1016/j.ijmedinf.2015.06.003] [Medline: 26113460]
59. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]
60. Seeber L, Conrad T, Hoppe C, Obermeier P, Chen X, Karsch K, et al. Educating parents about the vaccination status of their children: a user-centered mobile application. Prev Med Rep 2017 Mar;5:241-250 [FREE Full text] [doi: 10.1016/j.pmedr.2017.01.002] [Medline: 28127527]
61. Sun R, Burke LE, Saul MI, Korytkowski MT, Li D, Sereika SM. Use of a patient portal for engaging patients with type 2 diabetes: patterns and prediction. Diabetes Technol Ther 2019 Oct 01;21(10):546-556. [doi: 10.1089/dia.2019.0074] [Medline: 31335206]

62. Tsai R, Bell E, Woo H, Baldwin K, Pfeffer M. How patients use a patient portal: an institutional case study of demographics and usage patterns. Appl Clin Inform 2019 Jan 06;10(1):96-102 [FREE Full text] [doi: 10.1055/s-0038-1677528] [Medline: 30727003]
63. Wald JS, Businger A, Gandhi TK, Grant RW, Poon EG, Schnipper JL, et al. Implementing practice-linked pre-visit electronic journals in primary care: patient and physician use and satisfaction. J Am Med Inform Assoc 2010 Sep 01;17(5):502-506 [FREE Full text] [doi: 10.1136/jamia.2009.001362] [Medline: 20819852]
64. Yu DX, Parmanto B, Dicianno BE, Watzlaf VJ, Seelman KD. Accessibility needs and challenges of a mHealth system for patients with dexterity impairments. Disabil Rehabil Assist Technol 2017 Jan 07;12(1):56-64. [doi: 10.3109/17483107.2015.1063171] [Medline: 26153097]
65. Zettel-Watson L, Tsukerman D. Adoption of online health management tools among healthy older adults: an exploratory study. Health Informatics J 2016 Jun 22;22(2):171-183 [FREE Full text] [doi: 10.1177/1460458214544047] [Medline: 25149210]
66. Siek KA, Khan DU, Ross SE, Haverhals LM, Meyers J, Cali SR. Designing a personal health application for older adults to manage medications: a comprehensive case study. J Med Syst 2011 Oct 12;35(5):1099-1121. [doi: 10.1007/s10916-011-9719-9] [Medline: 21562730]
67. Lober WB, Zierler B, Herbaugh A, Shinstrom SE, Stolyar A, Kim EH, et al. Barriers to the use of a personal health record by an elderly population. AMIA Annu Symp Proc 2006:514-518 [FREE Full text] [Medline: 17238394]
68. Arar N, Seo J, Abboud HE, Parchman M, Noel P. Veterans' experience in using the online Surgeon General's family health history tool. Per Med 2011 Sep 01;8(5):523-532 [FREE Full text] [doi: 10.2217/pme.11.53] [Medline: 22076122]
69. Wu RR, Himmel TL, Buchanan AH, Powell KP, Hauser ER, Ginsburg GS, et al. Quality of family history collection with use of a patient facing family history assessment tool. BMC Fam Pract 2014 Feb 13;15(1):31 [FREE Full text] [doi: 10.1186/1471-2296-15-31] [Medline: 24520818]
70. Cimino JJ, Patel VL, Kushniruk AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Informatics 2002 Dec;68(1-3):113-127. [doi: 10.1016/s1386-5056(02)00070-9]
71. Witry MJ, Doucette WR, Daly JM, Levy BT, Chrischilles EA. Family physician perceptions of personal health records. Perspect Health Inf Manag 2010 Jan 01;7:1d [FREE Full text] [Medline: 20697465]
72. Kim MI, Johnson KB. Patient entry of information: evaluation of user interfaces. J Med Internet Res 2004 May 14;6(2):e13 [FREE Full text] [doi: 10.2196/jmir.6.2.e13] [Medline: 15249262]
73. Esmaeilzadeh P, Mirzaei T, Dharanikota S. The impact of data entry structures on perceptions of individuals with chronic mental disorders and physical diseases towards health information sharing. Int J Med Inform 2020 Sep;141:104157. [doi: 10.1016/j.ijmedinf.2020.104157] [Medline: 32559725]
74. Fareed N, MacEwan SR, Vink S, Jonnalagadda P, McAlearney AS. Relationships between patient portal activation and patient satisfaction scores among CG-CAHPS and HCAHPS respondents. Am J Manag Care 2022 Jan 13;28(1):25-31 [FREE Full text] [doi: 10.37765/ajmc.2022.88813] [Medline: 35049258]
75. Di Tosto G, McAlearney AS, Fareed N, Huerta TR. Metrics for outpatient portal use based on log file analysis: algorithm development. J Med Internet Res 2020 Jun 12;22(6):e16849 [FREE Full text] [doi: 10.2196/16849] [Medline: 32530435]
76. Park J, Liang M, Alpert JM, Brown RF, Zhong X. The causal relationship between portal usage and self-efficacious health information-seeking behaviors: secondary analysis of the health information national trends survey data. J Med Internet Res 2021 Jan 27;23(1):e17782 [FREE Full text] [doi: 10.2196/17782] [Medline: 33502334]
77. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3]
78. Moumjid N, Gafni A, Brémond A, Carrère M. Shared decision making in the medical encounter: are we all talking about the same thing? Med Decis Making 2007 Sep 14;27(5):539-546. [doi: 10.1177/0272989x07306779]
79. Arbuthnott A, Sharpe D. The effect of physician-patient collaboration on patient adherence in non-psychiatric medicine. Patient Educ Couns 2009 Oct;77(1):60-67. [doi: 10.1016/j.pec.2009.03.022] [Medline: 19395222]
80. Hamelinck VC, Bastiaannet E, Pieterse AH, van de Velde CJ, Liefers G, Stiggelbout AM. Preferred and perceived participation of younger and older patients in decision making about treatment for early breast cancer: a prospective study. Clin Breast Cancer 2018 Apr;18(2):e245-e253 [FREE Full text] [doi: 10.1016/j.clbc.2017.11.013] [Medline: 29275105]
81. Ende J, Kazis L, Ash A, Moskowitz MA. Measuring patients' desire for autonomy: decision making and information-seeking preferences among medical patients. J Gen Intern Med 1989 Jan;4(1):23-30. [doi: 10.1007/BF02596485] [Medline: 2644407]
82. Hamann J, Neuner B, Kasper J, Vodermaier A, Loh A, Deinzer A, et al. Participation preferences of patients with acute and chronic conditions. Health Expect 2007 Dec;10(4):358-363 [FREE Full text] [doi: 10.1111/j.1369-7625.2007.00458.x] [Medline: 17986072]
83. Benbassat J, Pilpel D, Tidhar M. Patients' preferences for participation in clinical decision making: a review of published surveys. Behav Med 1998 Jan;24(2):81-88. [doi: 10.1080/08964289809596384] [Medline: 9695899]
84. Joseph-Williams N, Elwyn G, Edwards A. Knowledge is not power for patients: a systematic review and thematic synthesis of patient-reported barriers and facilitators to shared decision making. Patient Educ Couns 2014 Mar;94(3):291-309. [doi: 10.1016/j.pec.2013.10.031] [Medline: 24305642]

85. Bruera E, Sweeney C, Calder K, Palmer L, Benisch-Tolley S. Patient preferences versus physician perceptions of treatment decisions in cancer care. J Clin Oncol 2001 Jun 01;19(11):2883-2885. [doi: 10.1200/jco.2001.19.11.2883]
86. Hack TF, Degner LF, Watson P, Sinha L. Do patients benefit from participating in medical decision making? Longitudinal follow-up of women with breast cancer. Psychooncology 2006 Jan;15(1):9-19. [doi: 10.1002/pon.907] [Medline: 15669023]
87. 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]

Abbreviations

PEHR: personal electronic health record
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by A Mavragani; submitted 07.03.22; peer-reviewed by J Lacroix, T Risling; comments to author 30.04.22; accepted 31.07.22; published 27.12.22.

Please cite as:
Damen DJ, Schoonman GG, Maat B, Habibović M, Krahmer E, Pauws S
Patients Managing Their Medical Data in Personal Electronic Health Records: Scoping Review
J Med Internet Res 2022;24(12):e37783
URL: https://www.jmir.org/2022/12/e37783
doi:10.2196/37783
PMID:36574275

Original Paper

A Hybrid Architecture (CO-CONNECT) to Facilitate Rapid Discovery and Access to Data Across the United Kingdom in Response to the COVID-19 Pandemic: Development Study

Emily Jefferson[1], BSc, PhD; Christian Cole[1], BSc, PhD; Shahzad Mumtaz[1], BSc, MCS, PhD; Samuel Cox[2], MMath, PhD; Thomas Charles Giles[2], BSc, PhD; Sam Adejumo[2], MSc; Esmond Urwin[2], BEng, MSc, PhD; Daniel Lea[2], BSc, MSc; Calum Macdonald[3], BSc, MPhys, PhD; Joseph Best[2,4]; Erum Masood[1], BE, MSc; Gordon Milligan[1], BSc, MSc; Jenny Johnston[1]; Scott Horban[1], BSc; Ipek Birced[1], BSc, BA, MSc; Christopher Hall[1], BSc; Aaron S Jackson[1], BSc, PhD; Clare Collins[2], BA; Sam Rising[2], BSc; Charlotte Dodsley[2]; Jill Hampton[1], BA; Andrew Hadfield[2], BSc; Roberto Santos[2], PhD; Simon Tarr[2], PhD; Vasiliki Panagi[2], BSc, MSc; Joseph Lavagna[2], BSc, MSc; Tracy Jackson[3], BA, MSc, PhD; Antony Chuter[5]; Jillian Beggs[1], BSc; Magdalena Martinez-Queipo[6], DipNur, RN, BSc, MClinRes; Helen Ward[7], PhD; Julie von Ziegenweidt[8,9]; Frances Burns[10], BA, PhD; Joanne Martin[11], BS, MA, MB, PhD; Neil Sebire[12], MBBS, BClinSci, MD; Carole Morris[13], BSc; Declan Bradley[14,15], PhD; Rob Baxter[16], BSc, MSc, PhD; Anni Ahonen-Bishopp[17], PhD; Paul Smith[17], BSc, MBA; Amelia Shoemark[18], BSc, ClinSci, PhD; Ana M Valdes[19], PhD; Benjamin Ollivere[19], MBBS, MA, MD; Charlotte Manisty[20], MBBS, MA, PhD; David Eyre[21], BM, BCh, DPhil; Stephanie Gallant[18], BN, MA, MSc; George Joy[22], MBBS; Andrew McAuley[23], BA, MSc, PhD; David Connell[24], BM BCh, MA, PhD; Kate Northstone[25], BSc, MSc, PhD; Katie Jeffery[26,27], BCh, BM, MA, PhD; Emanuele Di Angelantonio[28,29,30,31,32], MSc, MD, PhD; Amy McMahon[28,30], BSc, PhD; Mat Walker[28,30], BSc, PhD; Malcolm Gracie Semple[33,34], BSc, BM, BCh, PhD; Jessica Mai Sims[35], BA, MSc; Emma Lawrence[36], BSc, PhD; Bethan Davies[7], MB, BChir, MPH, PhD; John Kenneth Baillie[37], BSc, MBChB, PhD; Ming Tang[38], BSc, MBA; Gary Leeming[39], BA; Linda Power[40]; Thomas Breeze[41], BA (Hon), PGCE, MSc; Duncan Murray[42,43], BM, BCh, MA, PhD; Chris Orton[44], BA, MSc; Iain Pierce[45,46], BSc, MSc, PhD; Ian Hall[47], BA, BM, BCh, DM; Shamez Ladhani[48], PhD; Natalie Gillson[49], BSc; Matthew Whitaker[7], BA, MSc; Laura Shallcross[50], BA, MBBS, MSc, PhD; David Seymour[4], BSc; Susheel Varma[4], BEng, MSc, MBA, PhD; Gerry Reilly[4], BSc, MSc; Andrew Morris[4], MD, PhD; Susan Hopkins[40], BA, MB, MSc, MD; Aziz Sheikh[51], BSc, MBBS, MSc, MD; Philip Quinlan[2,19], BSc, PhD

[1]Health Informatics Centre, Division of Population and Health Genomics, School of Medicine, University of Dundee, Dundee, United Kingdom

[2]Digital Research Service, University of Nottingham, Nottingham, United Kingdom

[3]Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[4]Health Data Research UK, London, United Kingdom

[5]Lay Partnership in Healthcare Research, Lindfield, United Kingdom

[6]National Health Service Digital, London, United Kingdom

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

[8]Department of Haemotology, University of Cambridge, Cambridge, United Kingdom

[9]National Institute for Healthcare Research BioResource, Cambridge University Hospitals NHS Foundation, Cambridge Biomedical Campus, Cambridge, United Kingdom

[10]Centre for Public Health, Belfast Institute of Clinical Science, Queens University Belfast, Belfast, United Kingdom

[11]Blizard Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[12]Institute of Child Health, Great Ormond Street Hospital, London, United Kingdom

[13]Public Health Scotland, Edinburgh, United Kingdom

[14]Centre for Public Health, Institute of Clinical Science, Queen's University Belfast, Belfast, United Kingdom

[15]Public Health Agency, Belfast, United Kingdom

[16]EPCC, University of Edinburgh, Edinburgh, United Kingdom

[17]BC Platforms, Espoo, Finland

[18]Molecular and Clinical Medicine, School of Medicine, University of Dundee, Dundee, United Kingdom

[19]School of Medicine, University of Nottingham, Nottingham, United Kingdom

[20]Institute of Cardiovascular Sciences, University of College London, London, United Kingdom
[21]Big Data Institute, University of Oxford, Oxford, United Kingdom
[22]Barts Heart Centre, London, United Kingdom
[23]Clinical and Protecting Health Directorate, Public Health Scotland, Glasgow, United Kingdom
[24]School of Medicine, University of Dundee, Dundee, United Kingdom
[25]Population Health Sciences, Avon Longitudinal Study of Parents and Children, Bristol, United Kingdom
[26]Radcliffe Department of Medicine, Oxford University, Oxford, United Kingdom
[27]Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, United Kingdom
[28]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom
[29]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, United Kingdom
[30]National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Behaviour, University of Cambridge, Cambridge, United Kingdom
[31]Health Data Research UK Cambridge, Wellcome Genome Campus, University of Cambridge, Cambridge, United Kingdom
[32]Health Data Science Research Centre, Human Technopole, Milan, Italy
[33]Health Protection Research Unit in Emerging and Zoonotic Infections, Institute of Infections, University of Liverpool, Liverpool, United Kingdom
[34]Respiratory Department, Alder Hey Children's Hospital, Liverpool, United Kingdom
[35]University College London, London, United Kingdom
[36]BioIndustry Association, London, United Kingdom
[37]Outbreak Data Analysis Platform, University of Edinburgh, Edinburgh, United Kingdom
[38]NHS England, Worcestershire, United Kingdom
[39]Civic Data Cooperative, Digital Innovation Facility, University of Liverpool, Liverpool, United Kingdom
[40]Public Health England, London, United Kingdom
[41]Avon Longitudinal Study of Parents and Children, Bristol Medical School, University of Bristol, Bristol, United Kingdom
[42]University of Birmingham, Birmingham, United Kingdom
[43]University Hospital Coventry & Warwickshire NHS Trust, Coventry, United Kingdom
[44]Population Data Science, Swansea University Medical School, Swansea, United Kingdom
[45]Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, London, United Kingdom
[46]Institute of Cardiovascular Science, University College London, London, United Kingdom
[47]Nottingham Biomedical Research Centre, School of Medicine, University of Nottingham, Nottingham, United Kingdom
[48]Immunisation and Countermeasures Division, Public Health England Colindale, London, United Kingdom
[49]UK Health Security Agency, London, United Kingdom
[50]Institute of Health Informatics, UCL, London, United Kingdom
[51]Centre for Population Health Sciences, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Emily Jefferson, BSc, PhD
Health Informatics Centre
Division of Population and Health Genomics
School of Medicine, University of Dundee
Ninewells Hospital
Dundee, DD1 9SY
United Kingdom
Phone: 44 01382 383 353
Email: e.r.jefferson@dundee.ac.uk

Abstract

Background: COVID-19 data have been generated across the United Kingdom as a by-product of clinical care and public health provision, as well as numerous bespoke and repurposed research endeavors. Analysis of these data has underpinned the United Kingdom's response to the pandemic, and informed public health policies and clinical guidelines. However, these data are held by different organizations, and this fragmented landscape has presented challenges for public health agencies and researchers as they struggle to find relevant data to access and interrogate the data they need to inform the pandemic response at pace.

Objective: We aimed to transform UK COVID-19 diagnostic data sets to be findable, accessible, interoperable, and reusable (FAIR).

Methods: A federated infrastructure model (COVID - Curated and Open Analysis and Research Platform [CO-CONNECT]) was rapidly built to enable the automated and reproducible mapping of health data partners' pseudonymized data to the Observational Medical Outcomes Partnership Common Data Model without the need for any data to leave the data controllers' secure environments, and to support federated cohort discovery queries and meta-analysis.

Results: A total of 56 data sets from 19 organizations are being connected to the federated network. The data include research cohorts and COVID-19 data collected through routine health care provision linked to longitudinal health care records and demographics. The infrastructure is live, supporting aggregate-level querying of data across the United Kingdom.

Conclusions: CO-CONNECT was developed by a multidisciplinary team. It enables rapid COVID-19 data discovery and instantaneous meta-analysis across data sources, and it is researching streamlined data extraction for use in a Trusted Research Environment for research and public health analysis. CO-CONNECT has the potential to make UK health data more interconnected and better able to answer national-level research questions while maintaining patient confidentiality and local governance procedures.

(J Med Internet Res 2022;24(12):e40035) doi:10.2196/40035

KEYWORDS

COVID-19; clinical care; public health; infrastructure model; health data; meta-analysis; federated network; health care record; data extraction; data privacy; data governance; health care

Introduction

COVID-19 introduced a new set of conditions to existing challenges in health and clinical data collection within the United Kingdom. Regularly updated data were required at pace to inform decision-making and research, but were being generated by heterogenous sources, such as new "Lighthouse" laboratories [1] set up specifically for the pandemic, academic research laboratories, and usual primary and secondary care settings [2]. The diversity of data sources and the lack of awareness of them made it challenging to identify and access these data sources, as was highlighted by the UK Government Chief Scientific Adviser [3]. In our experience, it was often the case that each research or public sector group had to contact each potential data source individually to obtain information about the data they host, making the process complex and lengthy even for high-level questions, such as simply finding out what data are available. Such challenges are described in detail in the Goldacre Review [4] and across many studies [5-8].

Typically, any analysis of patient data or electronic health records (EHRs) requires many steps covering legal (eg, General Data Protection Regulations [GDPR] compliance) [9], operational (eg, data sharing agreements) [10,11], and security aspects (eg, access to unconsented pseudonymized or anonymized data in a secure environment where the data cannot be exported, ie, a Trusted Research Environment [TRE] [12]) [13]. These steps are crucial to ensure appropriate reuse of data but can take many months to complete before any data analysis can take place [14].

The need for more streamlined and efficient methods for discovering and analyzing EHRs is not new [15], but the COVID-19 pandemic has played a catalytic role in highlighting the need for these methods more than ever before. Data are federated when held at different locations and often hosted by different data controllers. The World Economic Forum has recently published a guideline document that focuses on sharing of sensitive health data in a federated consortium model considering the post–COVID-19 world [16]. Large-scale projects, such as the Global Alliance for Genomics and Health [17]; Canadian Distributed Infrastructure for Genomics [18]; Common Infrastructure for National Cohorts in Europe, Canada, and Africa [19]; and European Health Data and Evidence Network [20] projects, have laid out principles and frameworks supporting safe use of patient data [17,21]. While federated academic tooling (software that works on federated data sets) exists [22-25], the commercial sector appears to have more capability than the best in academia [26-28]. However, commercial systems usually come with contracts and licensing terms that may not be suitable for everyone and also focus on finding patients for recruitment to clinical trials rather than cohort discovery and meta-analysis from EHR data. Equally, the commercial nature of the systems means they are usually based on proprietary standards, which results in further fragmentation and lack of accessibility of data sets.

Given the need for more impactful solutions in accessing aggregated health data, accelerated by the pandemic [29], the COVID - Curated and Open Analysis and Research Platform (CO-CONNECT) was established at scale and at pace. The Health Data Research (HDR) Innovation Gateway [30] (Gateway) is a web resource enabling discovery of and accessibility to UK health research data, and supporting health data research in a safe and efficient manner. The Gateway provides detailed metadata descriptions of over 700 data sets held by members of the UK Health Data Research Alliance, including the Health Data Research Hubs [31]. CO-CONNECT enhances the capabilities of the Gateway by providing a query engine (the Cohort Discovery Tool) to support dynamic cohort building and meta-analysis across individual-level data from multiple data partners.

The aim of CO-CONNECT is to transform the way public health organizations and researchers discover and access COVID-19 data and associated longitudinal health care data from across the United Kingdom. This paper describes how CO-CONNECT maintains patient confidentiality and data security while supporting access to data for research at pace, and how a multidisciplinary team tackled the architecture of this platform as an asset for public health in the United Kingdom.

Methods

Project Initiation and Governance

CO-CONNECT was conceived early after the start of the pandemic when both researchers and public sector bodies were frantically trying to find what data existed across different data custodians to answer pressing questions, which would then inform public policy. Many research studies were being rapidly commissioned, and data were being collected via routine health care, but there was no easy way for different funders and research groups to understand what data were being collected. Once data sets had been identified, it took significant time to set up the agreements for data sharing and access.

For example, a key question at the time was whether someone would be immune to COVID-19 after contracting the disease, and if so, for how long. Low-level detailed serology results, rather than simply "positive/negative" results, were required for calibration of assays and to understand antibody responses related to individual levels of immunity. However, it was challenging for researchers to find which data controllers may be capturing low-level data, and if so, how to rapidly access the data for analysis.

These challenges where widely recognized at the time. When answering questions on the lessons to be learnt from the pandemic at the Science and Technology Committee meeting in July 2020 [3], Sir Patrick Vallance stressed the importance of data flows and data systems to support the pandemic response:

> *One lesson that is very important to learn from this pandemic, and for emergencies in general, is that data flows and data systems are incredibly important. You need the information in order to be able to make the decisions. Therefore, for any emergency situation those data systems need to be in place up front to be able to give the information to make the analysis and make the decisions.*

The CO-CONNECT leads reached out to 26 individuals/organizations who were collecting research cohorts of data or collecting data as part of routine health care provision from the 4 devolved UK Nations to join the project as collaborators. The benefits of the platform, how it would protect patient confidentiality, and how individual-level data would not have to leave the control of the data partner needed to be rapidly communicated for each data partner to agree to the collaboration. There were 4 co-leads on CO-CONNECT, who each bought different expertise to the project and could share the duties of leading such a large project delivered within a tight timeframe during the COVID pandemic.

CO-CONNECT partnered with the National Core Studies program [32] and reported to the UK Scientific Advisory Group for Emergencies [33] through this program. The Advisory Steering Committee meets every 3 months with representatives from the 4th Pillar Testing Programme and the UK Joint Biosecurity Centre, a Chief Scientific advisor, an ethics expert, and the funders.

Architecture of CO-CONNECT Infrastructure

Overview

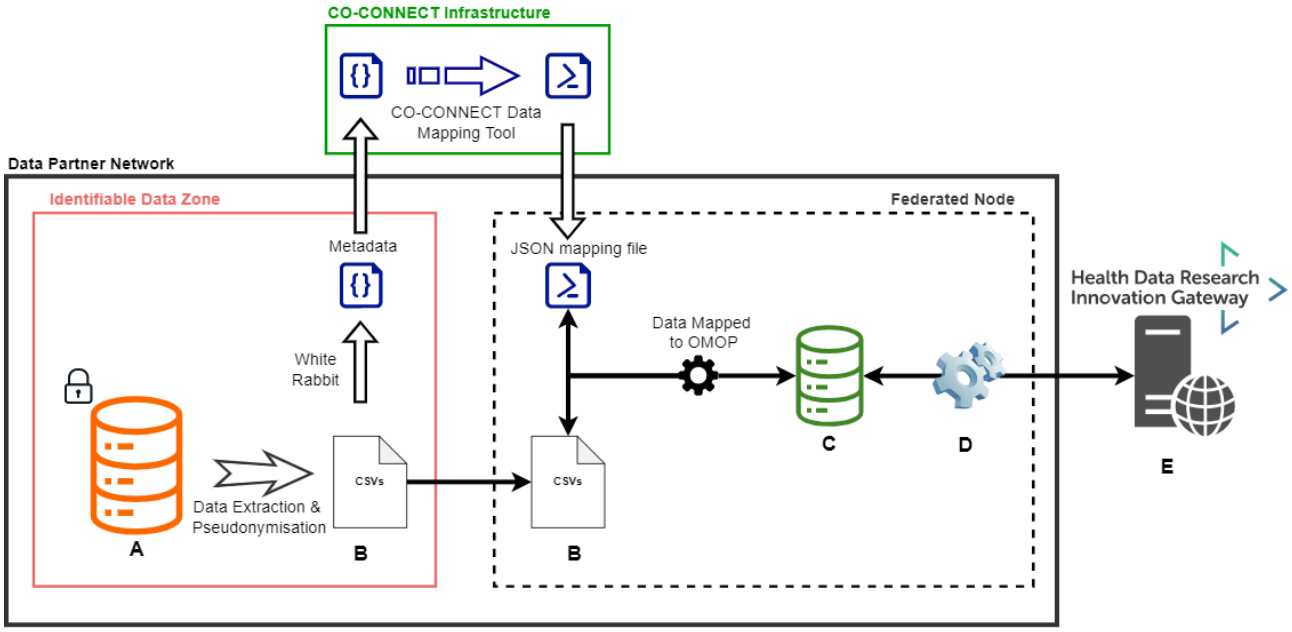
CO-CONNECT delivers a federated capability that enables the discovery of data across multiple sources, referred to as CO-CONNECT data partners, to make them findable, accessible, interoperable, and reusable (FAIR) [34]. The federation has been designed to ensure that data can be processed in line with the GDPR and common law confidentiality requirements.

Figure 1 provides an architecture overview of the components that reside within the secure environment of each data partner's network, with no inbound connectivity, and those that are available externally to researchers and the CO-CONNECT team via a secure login. Throughout the methods section, we reference the components as labeled in Figure 1 (Components A-E) in brackets after the description. Our overview video explains how the system works [35].

In summary, a secure virtual machine (VM) (Federated Node, dashed black box) is set up by the data partner, which is separate from the location where identifiable data are stored (Identifiable Data Zone, red box), but still part of their secure infrastructure. The data partner sends metadata ("Metadata" within the red box) about the data they hold to the CO-CONNECT technical team that determines the rules to map the data into the Observational Medical Outcomes Partnership (OMOP) [36] data standard format (CO-CONNECT Infrastructure, green box). The mapping script (JSON mapping file), developed by the CO-CONNECT technical team, is sent to the data partners who then apply the mapping rules to a pseudonymized version of their data (Data Mapped to OMOP). This generates a database of relevant linked and pseudonymized data sets in OMOP format within their VM (Component C, green database).

Software is installed within the VM, called BC|LINK (Component D), which provides access to the pseudonymized OMOP database (Component C, green database) and is configured to communicate with the Gateway tool (Component E) where approved users can submit queries. The Gateway contains the BC|RQUEST software (Component E) that stores the user-submitted queries and allows the BC|LINK software (Component D) to download these queries and run them against the OMOP database. Only aggregate counts are posted in response and displayed to the user. This is simultaneously repeated across all UK-wide data partners within the federation, which enables users to perform feasibility analysis (to discover relevant data from different sources) and carry out aggregate-level analysis across different UK data partners through one system.

Figure 1. The CO-CONNECT federated architecture. A data partner (dark box) has potentially identifiable data (A) from which an extraction is made and pseudonymized (B). A metadata extraction is performed with WhiteRabbit (within the identifiable Data Zone, red box) and sent to the CO-CONNECT infrastructure (green box). A mapping script to the OMOP CDM is created using the CO-CONNECT data mapping tool (CaRROT-Mapper). The pseudonymized data are securely transferred (B) into a secure virtual machine hosted by the data partner (Federated Node, dashed dark box), mapped to OMOP (CaRROT-CDM), and connected to the federation software (C and D). From there, the data are queryable by the Innovation Gateway (E). Only aggregated fully anonymous data discovery and meta-analysis results are returned to the Gateway (D). CDM: Common Data Model; OMOP: Observational Medical Outcomes Partnership.



Detailed Components of the Architecture

CaRROT Software

All CO-CONNECT developed tools (termed CaRROT [Convenient and Reusable Rapid Ontology Transformer]) are open source and freely available [37,38]. This suite of tools automates the mapping of the data into OMOP and the loading of the data into a database for external querying.

Access to Individual-Level Data

All individual-level data remain under the control of the data partner, and there is no requirement for any direct interaction from the CO-CONNECT pipeline with the data partner's data systems (Database A). The federated node (dashed black box) is established on a VM that is separate from any systems that hold identifiable data.

ID Management and Data Linkage

All patient identifiable data are pseudonymized locally by data controllers (Data Extraction and Pseudonymization) through (1) obfuscation of potentially sensitive information, such as date of birth, and (2) removal of personal identifiable information, such as given names and addresses.

Generating Metadata

WhiteRabbit, from Observational Health Data Science and Informatics (OHDSI) [39], is a software tool to profile data sets to generate metadata that includes descriptions on tables, fields, and the distribution of values within each field [40]. WhiteRabbit resides within the Identifiable Data Zone but is only ever run against a pseudonymized extract of the data in CSV format (Files B), from which the WhiteRabbit report is generated. The data partner always retains control over what data WhiteRabbit can access, the configuration of the parameters, and what is shared to the CO-CONNECT team.

Data Mapping Tool

To ensure consistency of data across the data partners, all of the data sets are on-boarded using OMOP Common Data Model (CDM) version 5.3 [36] developed as part of the OHDSI.

We developed a data mapping tool (CaRROT-Mapper [37]; CO-CONNECT Infrastructure, green box), which ingests WhiteRabbit reports and enables the data team to generate a mapping rule to replace each field or field value to a standard OMOP vocabulary concept ID. From this concept ID, the domain can be established, which in turn confirms which table in the OMOP CDM should be used to store the data. Importantly, rules that were generated previously can be reused by other data partners that have similar data structures or for subsequent updates to the data, rather than starting from scratch. At the time of writing, the CaRROT-Mapper supports transformation to the Person, Observation, Condition Occurrence, Measurement, and Drug Exposure tables.

The conversion and destination tables are captured as "mapping rules" in a single JSON file, which is sent to the data partner.

Extract, Transform, and Load Pipeline

The mapping rules developed are used by the Python Extract, Transform, and Load (ETL) pipeline (CaRROT-CDM) [38,41], to convert the data from its native CSV format into the OMOP CDM. The ETL pipeline can be scheduled to run either on demand or whenever new data are available.

Federated Querying

The BC|RQUEST query portal (Component E) was licensed as a white-labeled instance from BC Platforms [26-28] and integrated into the Gateway as the Cohort Discovery Search Tool [42]. This component provides an interface allowing approved users (bona fide researchers) to create the definition of their cohort (cohort queries) via a drag and drop interface of available OMOP concepts. Cohort queries (also known as study feasibility queries) are created within the query portal and queued to be collected every few seconds by the BC Platforms BC|LINK software installed (Component D) within the data partners' Federated Nodes. BC|LINK executes the queries and returns the aggregated results to the query portal.

A single BC|LINK instance can interact with multiple OMOP data sets held by each data partner, and allows each data partner to independently set all data disclosure rules, including data rounding, low number suppression, and whether metadata analysis can be performed. This allows each data partner to determine risk and set appropriate controls as required for each data set rather than a single setting for all data sets. Although the data are stored in software from BC Platforms, they have no mechanism to access the data. All access to the data remains strictly under the control of the data partner.

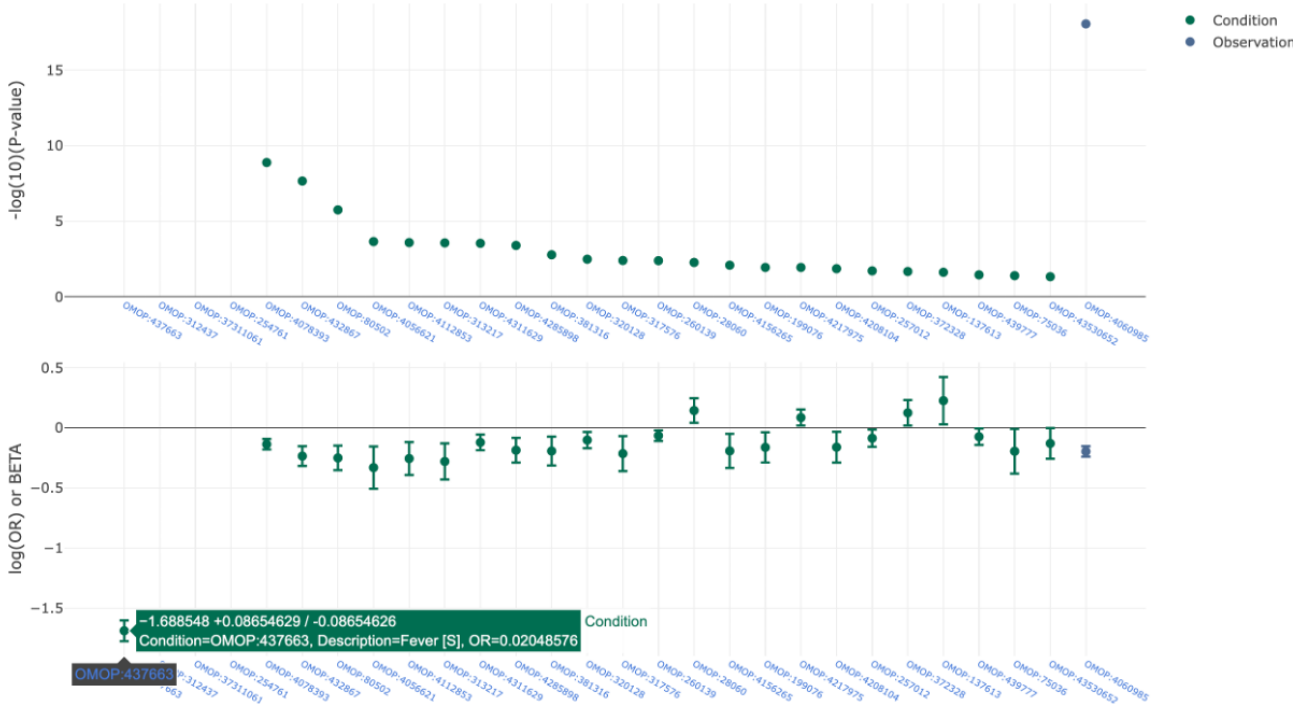
Feasibility Questions

The system allows researchers to dynamically and in real time define the cohorts of interest [42]. They will receive responses from across the network usually within a minute. Such an approach allows the feasibility of potential studies to be understood based on the actual data available and without intervention from data partners. This important feature ensures that researchers understand what is feasible in near real time, while always ensuring the disclosure controls are applied by each data partner.

Meta-Analysis

The capability to perform meta-analysis queries across their data sets is configured by the data partners through an "opt-in" mechanism. Researchers are able to request predefined analyses, through a common user interface, to run across the "opted-in" data sets. An example of a meta-analysis query is to undertake a phenome-wide association study (PheWAS) analysis to understand what phenotypes are linked to different levels of antibody response. In the out-of-the-box capability from BC Platforms, the PheWAS analysis is initially treated as 2 availability queries, one for the case and the other for the control section of the selected cohorts. The subsets of individuals returning within each availability query are then selected from the database, and a PheWAS/Forest analysis is performed across the OMOP CDM search space. This identifies the most overrepresented and underrepresented terms within each cohort. The output is returned to BC|RQUEST as an array of data, which is combined with the information from other cohorts to find the common "META" terms that are overrepresented and underrepresented across all the cohorts. This information is displayed back to the user in the form of a PheWAS plot or a forest plot, or downloaded as a Boolean table of the results. An example is shown in Figure 2.

Figure 2. An example phenome-wide association study plot across 4 test data sets comparing females with pneumonia against a background population of female-only samples. The most overrepresented classes include fever (OMOP:437663), disease caused by 2019-nCoV (OMOP:37311061), dysphenia (OMOP:312437), and cough (OMOP:254761). OMOP: Observational Medical Outcomes Partnership.



Custom meta-analytic modules can also be implemented within the BC Platforms ecosystem. These can be developed in either R or Python. Work to develop more advanced statistical meta-analysis and investigations into potential biases or statistical challenges will form future research.

Data Access Requests

The data discovery and meta-analysis tools only report aggregated-level data. Details of the data sources queried are provided, so that when an appropriate cohort is identified, direct contact with the appropriate data partner can be made to initiate data governance approvals for a specific research study, which requires individual-level data analysis using the cohort identified. The Gateway-standardized governance application process (Five Safes [safe projects, safe people, safe settings, safe outputs, and safe data] [43,44] form) can be used to streamline the effort required to obtain approvals from data partners who have adopted the standard [45].

Engagement With Patients and the Public

We have patient and public representatives co-leading the project, with 2 lay member co-investigators and a public and patient group. Representatives attend our work package, leadership team, and advisory board meetings. Representatives reviewed all the controls developed for CO-CONNECT, ensuring we are protecting patient confidentiality and maintaining trust. We developed a series of public-facing videos: Overview [35], Finding Data [46], and Analyzing Data [47]. We also drafted a lay summary and Frequently Asked Questions page [48].

Ethics Considerations

Research ethics approval was not required for this project as each data partner maintains their own governance and ethics for the original research studies. Anyone requiring access to the platform to perform research needs to apply for their own ethics approval.

Results

Data Coverage

The CO-CONNECT consortium includes 41 leaders from 29 different organizations across the 4 devolved UK Nations and is currently on-boarding over 56 different data sets into the platform. The project was launched in October 2020, with 18 months of funding and extension for another 6 months.

CO-CONNECT is focused on the following 3 different types of data partners: (1) COVID-19 research consented cohorts collecting serology data; (2) routinely collected unconsented data from across the United Kingdom; and (3) research cohorts collected prior and during the current pandemic, which CO-CONNECT is enhancing with the ability to link to COVID-19 data (augmented cohorts).

The sources for each type are shown in Table 1. Approximately half of the COVID-19 research cohorts being collected are from health care workers.

Table 1. List of data sources incorporated into CO-CONNECT.

Cohort type	Source
COVID-19 serology cohorts	
Health care workers	Co-STARS[a] [49], COVIDsortium [50], MATCH [51], Oxford Healthcare Workers [52], PANTHER[b] [53], and SIREN[c] [54]
Blood donors	TRACK-COVID [55]
Care homes	VIVALDI [56]
Hospitalized patients	ISARIC[d] [57]
Schools	sKIDs[e] [58]
Education	ACE[f] [59]
Random sample of the population of adults registered with a general practitioner in England	REACT-2[g] [60]
Hospitalized and community follow-up	FOLLOW-COVID[h] [61]
Augmented cohorts	
Longitudinal cohorts	ATLAS[i] [62] (ALSPAC[j] [63], Generation Scotland [64], GASP[k] [65], NIHR-BioResource [66], TWINS-UK [67]), and Wellcome Longitudinal Population Study [68] (6 cohorts)
Respiratory cohorts	HDR[l] UK BREATHE Hub [69] (17 cohorts)
Routinely collected health data sources/Trusted Research Environments	
England	National Health Service (NHS)–Digital [70] and UK Health and Security Agency (previously Public Health England) [71]
Scotland	Public Health Scotland (PHS) [72]
Northern Ireland	HSC[m] Business Services Organisation [73] and HSC Public Health Agency [74]
Wales	Secure Anonymised Information Linkage (SAIL) service [75]
UK-wide	Office of National Statistics (ONS) [76]

[a]Co-STARS: COVID-19 Staff Testing of Antibody Responses Study.

[b]PANTHER: Pandemic Tracking of Healthcare Workers.

[c]SIREN: SARS-CoV-2 Immunity and Reinfection Evaluation Network.

[d]ISARIC: International Severe Acute Respiratory and emerging Infections Consortium.

[e]sKIDS: COVID-19 Surveillance in School Kids.

[f]ACE: Asymptomatic COVID-19 in Education.

[g]REACT-2: Real-time Assessment of Community Transmission 2.

[h]FOLLOW-COVID: Focused Longitudinal Observational Study to Improve Knowledge of COVID-19.

[i]ATLAS: Access Points to Tissue, Longitudinal Data, Archives, and Samples.

[j]ALSPAC: Avon Longitudinal Study of Parents And Children.

[k]GASP: Genetics of Asthma Severity and Phenotypes.

[l]HDR: Health Data Research.

[m]HSC: Health and Safety Commission.

Data Sets Onboarded

The HDR UK Cohort Discovery Service was first launched in April 2021. At the time of writing, the following data partners are live within the HDR Cohort Discovery Tool: ALSPAC (Avon Longitudinal Study of Parents And Children), PANTHER (Pandemic Tracking of Healthcare Workers), GASP (Genetics of Asthma Severity and Phenotypes), ACE (Asymptomatic COVID-19 in Education) Cohort, MATCH, Generation Scotland, NIHR Bioresource, FOLLOW-COVID (Focused Longitudinal Observational Study to Improve Knowledge of COVID-19), Co-STARS (COVID-19 Staff Testing of Antibody Responses Study), TRACK-COVID, and COVIDSortium. The following data partners have governance approvals in place and are in the process of being on-boarded: ISARIC4C (International Severe Acute Respiratory and emerging Infections Consortium), UKHSA (SIREN [SARS-CoV-2 Immunity and Reinfection Evaluation Network] and sKids [COVID-19 Surveillance in School Kids]), REACT-1 (Real-time Assessment of Community Transmission 1), REACT-2 (Real-time Assessment of Community Transmission 2), Oxford Healthcare Workers, TWINS-UK, Wales/SAIL (COVID Vaccination Dataset [CVVD] and COVID Test Results [PATD]), Public Health Scotland (13 different data sets), and Northern Ireland (COVID

antigen testing pillar 1 and 2, COVID-19 Vaccination, Admissions, and Discharges, Emergency Department). CO-CONNECT is currently working with the remaining data partners to obtain relevant governance approvals for their data sets to be incorporated into the platform.

This is an innovative infrastructure project to support research at scale across the United Kingdom. The unique nature of the project made it challenging to onboard data sets from different organizations in terms of (1) different data governance processes with varying information required, (2) different levels of understanding of governance requirements and the technical solution, and (3) delays in governance due to capacity during a pandemic. To overcome these challenges, approaches, such as one-to-one sessions, technical guidance workshops, and sharing a governance guidance pack [77] with data partners, were used. We also commissioned explainer videos to explain the system and how it protects patient confidentiality for both data partners and the general public [35,46,47]. We plan on describing these challenges and lessons learnt elsewhere.

User Feedback

HDR UK undertook market research in December 2021 and January 2022 led by an external agency. The research included audience mapping, analysis, and 30 interviews with health data users from a range of sectors, including industry, academia, and the National Health Service (NHS). Overall, Cohort Discovery was very positively received, and a short-term goal now for HDR is to "build on perceived successes in search functionality, that is, the Cohort Discovery Tool." The feedback from users was that the Cohort Discovery Tool could help address some of the needs around metadata and that the approach reflected the way in which many want to understand, assess, and access data. The users recognized the value of standardization across data collection/data terms to vastly increase the options for linking and comparing data and wanted to see the tool developed further. There are currently 150 active users. We expect this to increase with additional data sets live on the system and promotion of the resource.

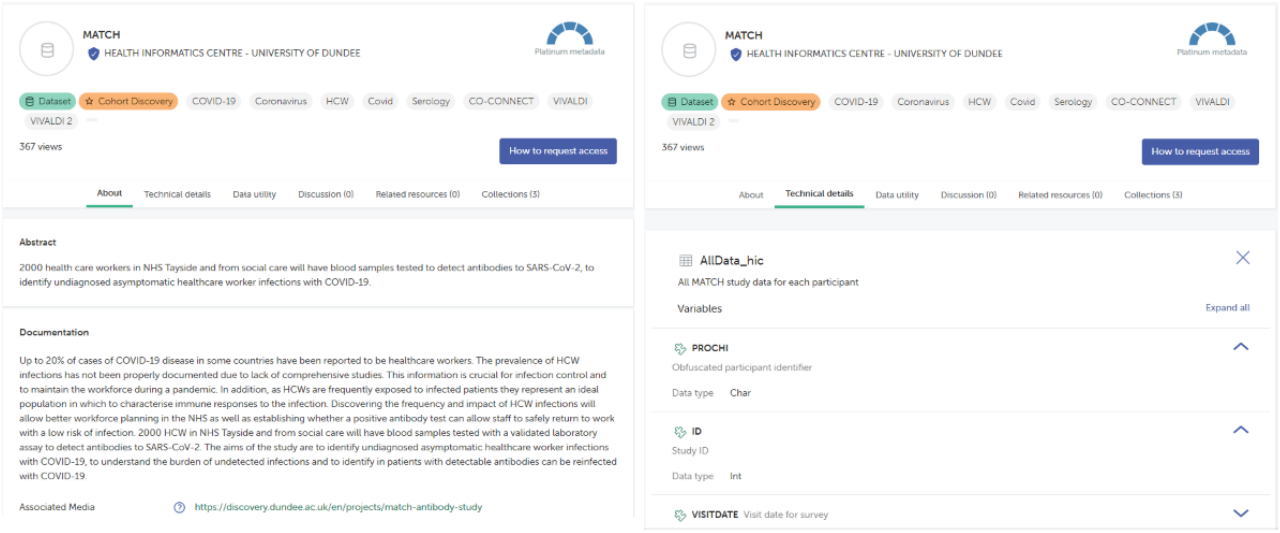
Key Outcome of the CO-CONNECT Infrastructure

CO-CONNECT is enabling rapid data discovery of data sets available from each data partner via near instantaneous aggregate-level cohort building queries. Figure 3 shows the Cohort Discovery Tool, available from the Gateway, with an example query of "all females with asthma" against all available data sets. The aggregated results presented in the Figure 3 example include overall counts, and age and gender distributions across all data partners down to the individual data set level, enabling researchers to rapidly refine their cohorts of interest.

Prior to the Cohort Discovery Tool being embedded within the Gateway, the only information a researcher could access was a static metadata catalogue of data sets/cohorts, such as overall population size, table names, and field names with their data types and descriptions, as shown in Figure 4. In contrast, the Cohort Discovery Tool enables researchers to dynamically define a cohort search query and get aggregate counts matching the cohort search criteria for the data sets.

Figure 3. The HDR UK Cohort Discovery Tool. The interface enables the user to define their cohort search criteria and displays aggregate results across different data sets. The available cohort search criteria (A) are used to create selected cohort criteria (a drag and drop feature, B). Results matching the cohort search criteria across different data sets are presented in the output once the federated queries are completed (C). HDR: Health Data Research.

Figure 4. An example of the static metadata found in the data catalogue of the HDR Innovation Gateway (MATCH data set). (A) Summary of the MATCH data set. (B) Technical details – a list of tables with their field names and data types. HDR: Health Data Research.



CO-CONNECT allows meta-analysis across the data sets, such as time series or binary comparisons. When researchers and public health groups need access to individual-level pseudonymized data for detailed analysis (over and above aggregate-level analysis available in the tool), the data for the analysis can be moved to a TRE for access by the researchers. The CO-CONNECT architecture is being enhanced to support semiautomated streamlined extracts of standardized linked data from across multiple data partners for access within a TRE [12].

Future Work

We are working with data partners to research mechanisms in which, where practical to do so, global pseudoidentifiers are identical across different data partners. This would be achieved by the use of a common one-way irreversible cryptographic hashing algorithm applied to identifiers, such as NHS and Community Health Index numbers, and would enable data linkage across data partners. These global pseudoidentifiers are never shared outside of the group of data partners. This would enable data linkage across data sets from different data sources (see section on extraction into a TRE below) and would support duplicate detection.

To support duplicate detection for the aggregate-level data discovery and meta-analysis functionality, we have a minimum viable product developed with BC Platforms ensuring that for each query, the global pseudoidentifiers are replaced by query-specific identifiers within the VM. The list of query-specific identifiers is returned along with the aggregate-level counts associated with the query to a secure temporary location, and the IDs from each data partner can then be automatically compared, providing the user who initiated the query with an estimate of the overlap of individuals across different cohorts. For example, 200 people met the search criteria from data partner A, while 350 people met the search criteria form data partner B, and 27 people were the same individuals from data partners A and B. The query-specific identifiers are never made visible to the user and are generated afresh using a new salt (random data that is used as an additional input to a one-way function that hashes data) for each query before being deleted at the query end. CO-CONNECT is working across data partners to assess the feasibility of enabling such functionality.

Extraction Into a TRE

The CO-CONNECT architecture is being enhanced to support the linkage and extraction of individual-level data from the pseudonymized databases within each data partner into a TRE. There are many TREs operating across the United Kingdom, such as the National TREs for England [70], Scotland [72], Northern Ireland [74], and Wales [75]. These example TREs were also data partners of CO-CONNECT. Data partners can choose whether to use the CO-CONNECT semiautomated pipeline or their own in-house methods for data extraction. When extracting research project–specific individual-level data into a TRE, the global pseudoidentifiers will be replaced with new project-specific pseudoidentifiers prior to export. This means that data from different data partners are linkable by the research group within the TRE for the specific research project without the global identifiers being shared. As the pseudoidentifiers are project specific, linkage across different research projects is safeguarded against.

Discussion

Hybrid Infrastructure

We have brought together EHR data of national importance into a federated platform. The data can be queried via the Cohort Discovery Tool in the HDR UK Innovation Gateway. An open-source set of tools were developed to standardize the mapping of data into the OMOP standard without the need to view the individual-level data.

CO-CONNECT evolved from a recognized need across multiple domains for a transformative step in the ability for researchers to discover data across a range of data assets. Centralized data architectures have historically been used when it is possible to set up flow of data to a single location, under a single set of governance approvals (such as national registries) and usually

from a small number of organizations. This has been very effective in the United Kingdom with flow of data from the NHS bodies to respective national data repositories, especially when there is a legal mandate, such as the registration of a disease. Such approaches are successful at supporting certain research activities, such as epidemiology, where evaluating the prevalence of a disease can be undertaken with relative ease.

Centralized infrastructure brings economy of scale and the ability to have a specialized team of technologists that can bring standardization to the process and policy. However, such centralized infrastructure cannot infinitely scale to accommodate all data that might be required to perform analyses. It is also clear that while certain aspects of epidemiological research can be undertaken via a centralized model, such as the prevalence or risk associated with different demographic characteristics, it is likely there will never be enough data held in a single location to help answer questions of causation rather than retrospective observations. There is a need to combine information from multiple sources to increase power and generalizability. Aside from technical constraints, the public are equally uncomfortable with their sensitive data being shared widely or within a central database, and thus, keeping all individual-level data local improves patient trust [78,79].

COVID-19 brought a set of challenges such that data analysis and infrastructure were required across and between the national centralized databases of the 4 nations of the United Kingdom. CO-CONNECT was tasked to deliver an overarching platform across existing centralized infrastructure, as well as cater to academic collection of data. This was not a simple distinction between federated or centralized models, but a hybrid infrastructure to support both federation across national centralized TREs and inclusion of specific research data sets into a single ecosystem of collaboration and co-existence.

Federated Cohort Discovery

CO-CONNECT has been designed to work for the whole population of the United Kingdom. These data come from many databases with thousands of fields held within each of the 4 nations. The technical novelty of the architecture lies in the fact that it supports reproducible and semiautomated processing/tooling for inclusion of new data sets and addition of new fields without significant additional effort compared with OHDSI's tooling available [80]. Therefore, while federated cohort discovery tools do exist, this is the first time such a system has been designed to be deployed at this scale. The CO-CONNECT approach federates cohort discovery from one simple-to-use application. It will enable the querying of data sets from the 4 nations within the United Kingdom without separate data governance applications. Researchers are able to query data sets immediately and interactively as part of their feasibility study without the substantial overhead of contacting each data partner to ask about running multiple bespoke feasibility queries.

Centralized Data Curation

All source data are transformed into the OMOP data model via our teams in Dundee, Nottingham, and Edinburgh. The software developed allows the maps to be created centrally but applied locally by each data partner. This retains a clear separation for data governance and importantly enables data partners to be included with minimum effort for them. This is performed via reproducible code, which ensures transformations to the data from the source to the new model are consistent across projects. The mapping of the data into OMOP is supported by the core data science team across all the data partners, ensuring standardization in mapping. Using a reproducible workflow works in concert with automation to support the regular updating of data across the platform via a consistent ETL mechanism.

Data Extraction

Federated analytics is emerging as a credible alternative, but it was recognized that certain analyses cannot be undertaken using current federated approaches. Therefore, despite putting in place a federated architecture, we are designing the approach to allow subsets of pseudonymized data for answering a specific research project to be extracted into a single TRE. Data curation to a standard will aid this process significantly, as all data have already been curated to the OMOP CDM. The automation of these steps streamlines the process of transitioning to individual-level data from a higher-level query and reduces costs. The data partners who chose to adopt the automated process will require limited resource to release data, and throughput can scale without additional investment. Researchers will receive data in a familiar format, allowing them to reuse existing methodologies. The data in the original format can also be provided to the researchers should this be required.

Comparison With Other International COVID-19 Initiatives

We reviewed other existing COVID data efforts across the world [81-86]. Most projects focus on the analysis of data sets that were already known to the researchers, whereas CO-CONNECT (as well as CODEX [84,85]) also provides the capability to search for specific cohorts of data for feasibility analysis across population-wide data.

Projects, such as 4CE [83], N3C [86], and the COVID-19 Data Exchange Platform [84], took a centralized approach. 4CE [83] transformed data into a common format at each data source and then obfuscated the values. 4CE transferred the files to a shared central location, merging the files from different sources so analysis could take place. N3C [86] supported data in 4 different CDMs: PCORnet [87], OMOP, i2b2/ACT [88,89], and TriNetX [27], bringing the data into a central cloud platform for secure analysis. The COVID-19 Data Exchange Platform supported federated nodes in the i2b2 [23] format and federated queries, and also provided a centralized analysis platform. They encountered challenges with obtaining ethical approval for transferring data onto the centralized platform, and at the time of writing, data from only 350 patients had been transferred.

The COVID-19 SCOR project [81] plans to utilize the MedCo software [82], which uses collective homomorphic encryption and obfuscation across decentralized data sources. MedCo is deployable on top of standardized systems, such as i2b2 [23]/SHRINE [90] and TranSMART [91]. The unCoVer project aims to use the DataSHIELD [25] software to perform federated analytics across 18 countries [92]. As far as we are aware, all

these federated analytics solutions require inbound connections to the data and opening ports on firewalls. In the case of MedCo, encryption of the data reduces the privacy risks associated with inbound connections to the data.

The approach taken really depends on the attitudes of the data partners. In CO-CONNECT, most partners would not accept inbound connections into their secure environment and would not be happy to place sensitive data in an area where an inbound connection could be allowed, regardless of encryption or access controls. For those reasons, CO-CONNECT was built on the assumption of never requiring an inbound connection to the federated data to either curate the data or run a feasibility analysis and meta-analysis. As an additional level of security, on top of not allowing inbound queries, the CO-CONNECT architecture could adopt homomorphic encryption in the future to support more advanced federated queries where researchers need to see the underpinning data.

CO-CONNECT, unlike other COVID-19 solutions, supports data partners to automate the mapping of their data into a CDM without having to see the underpinning data. This is advantageous as most data partners do not have their data mapped into the OMOP CDM or the technical capability to do so.

Current Status and Contributions

Metadata covering the data sources are now available to search openly within the Gateway [30]. National and international researchers can request access to the enhanced dynamic cohort discovery capability within the Gateway. Access to individual-level subsets of data by national and international researchers can also be requested via the streamlined governance application process [45].

We welcome requests to onboard data sets into CO-CONNECT; further details are available via the corresponding author.

The platform has been designed to be disease agnostic. COVID-19 has supported the need for such a platform to provide data at pace. However, the model can be reused to support research at pace for other disease areas. The platform underpins the recently funded HDR UK/MRC Alleviate Hub for Pain Research [93], and the architecture and support for cohort building will be supported and enhanced by HDR after the end of CO-CONNECT funding. Exemplar projects using the architecture are planned for the next phase of HDR funding.

Conclusions

We have introduced the CO-CONNECT federated architecture, which addresses the challenges of fragmentation of data and lack of interoperability and standardization, as well as the challenge of linkage of high value data assets to other data assets providing new scientific insights. The architecture has been designed around the following core principles: (1) maintaining patient confidentiality, trust, and data security; (2) empowering data partners to be interconnected in a sustainable environment; (3) utilization and re-enforcement of TREs to analyze data; (4) a focus on data engineering to ensure technical legacy for wider use; and (5) a standard-based approach to ensure interoperability, repeatability, and connectivity to other initiatives, responding to the most pressing needs of the public health and research communities.

The development of this platform will empower public health organizations, research groups, and industry bodies to answer key questions about the COVID-19 pandemic and its effects on human health in a streamlined timely manner, as has been needed for EHRs for many years [15,21]. The solution enables rapid cohort-building data discovery across data partners. None of the data partners had such capability for researchers prior to CO-CONNECT. CO-CONNECT has simplified the complex task of requesting access to each individual data set, by providing transparency on what data are available and from where, and how to request access if individual-level data analysis is required. CO-CONNECT provides novel real-time functionality compared to static metadata dictionaries and descriptions of cohorts already provided within the Gateway.

The immediate impact of CO-CONNECT is the fast, accessible, and standardized availability of aggregate COVID-19–related data, to inform key public health decisions and help tackle the COVID-19 pandemic at pace. As more data sets are onboarded, this will become more powerful.

Acknowledgments

This work uses data provided by patients and collected by the National Health Service as part of their care and support. We acknowledge the support from CO-CONNECT funded by United Kingdom Research and Innovation (Medical Research Council) and the Department of Health and Social Care (National Institute of Health and Care Research) (MR/V03488X/1) and supported by a consortium of 22 partner organizations whose infrastructure made this research possible.

The work builds on early pilot work supported by the University of Nottingham Impact Acceleration Account (EP/R511730/1) from the Engineering and Physical Sciences Research Council.

This work was supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and the Wellcome Trust.

Authors' Contributions

EJ contributed to funding acquisition, writing–original draft, writing–review and editing, and supervision; C Cole contributed to writing–original draft, writing–review and editing, and supervision; SM contributed to writing–original draft, writing–review and editing, software, methodology, data curation, investigation, and formal analysis; SC contributed to writing–original draft, writing–review and editing, software, methodology, data curation, investigation, and formal analysis; TG contributed to writing–review and editing, software, methodology, data curation, investigation, and formal analysis; SA contributed to writing–review and editing, software, methodology, data curation, investigation, and formal analysis; EU contributed to writing–review and editing, software, methodology, data curation, investigation, and formal analysis; DL contributed to writing–review and editing, and software; C Macdonald contributed to writing–original draft, writing–review and editing, software, methodology, data curation, investigation, and formal analysis; J Best contributed to software, methodology, data curation, investigation, and formal analysis; EM contributed to writing–review and editing, software, methodology, data curation, investigation, and formal analysis; GM contributed to writing–original draft, writing–review and editing, supervision, software, methodology, data curation, investigation, and formal analysis; JJ contributed to writing–review and editing, supervision, and project administration; S Horban contributed to writing–original draft, writing–review and editing, software, methodology, data curation, investigation, and formal analysis; IB contributed to software, methodology, data curation, investigation, and formal analysis; CH contributed to writing–review and editing, software, methodology, data curation, investigation, and formal analysis; ASJ contributed to writing–review and editing, software, methodology, and investigation; C Collins contributed to project administration and supervision; SR contributed to software, methodology, data curation, investigation, and formal analysis; CD contributed to project administration and supervision; JH contributed to project administration; AH contributed to software, methodology, data curation, investigation, and formal analysis; RS contributed to software, methodology, data curation, investigation, and formal analysis; ST contributed to software, methodology, data curation, investigation, and formal analysis; VP contributed to software, methodology, data curation, investigation, and formal analysis; JL contributed to software, methodology, data curation, investigation, and formal analysis; TJ contributed to project administration, supervision, and writing–review and editing; AC contributed to writing–review and editing; J Beggs contributed to writing–review and editing; MMQ contributed to methodology, data curation, and investigation; HW contributed to methodology, data curation, and investigation; JvZ contributed to methodology, data curation, and investigation; FB contributed to methodology, data curation, and investigation; JM contributed to funding acquisition, writing–review and editing, and methodology; NS contributed to funding acquisition, writing–review and editing, and methodology; C Morris contributed to writing–review and editing, methodology, data curation, and investigation; DB contributed to writing–review and editing, methodology, data curation, and investigation; RB contributed to methodology and investigation; AAB contributed to methodology and investigation; PS contributed to methodology and investigation; A Shoemark contributed to methodology, data curation, and investigation; AMV contributed to funding acquisition, and writing–review and editing; BO contributed to methodology, data curation, and investigation; C Manisty contributed to writing–review and editing, funding acquisition, methodology, data curation, and investigation; DE contributed to methodology, data curation, and investigation; SG contributed to methodology, data curation, and investigation; GJ contributed to methodology, data curation, and investigation; AMA contributed to methodology, data curation, and investigation; DWC contributed to funding acquisition, methodology, data curation, and investigation; KN contributed to methodology, data curation, and investigation; KJ contributed to funding acquisition, methodology, data curation, and investigation; EDA contributed to funding acquisition, methodology, data curation, and investigation; AMM contributed to methodology, data curation, and investigation; M Walker contributed to methodology, data curation, and investigation; MGS contributed to methodology, data curation, and investigation; JMS contributed to funding acquisition and methodology; EL contributed to funding acquisition and methodology; BD contributed to methodology, data curation, and investigation; JKB contributed to writing–review and editing, funding acquisition, methodology, data curation, and investigation; MT contributed to writing–review and editing, funding acquisition, methodology, data curation, and investigation; GL contributed to methodology, data curation, and investigation; LP contributed to methodology, data curation, and investigation; TB contributed to methodology, data curation, and investigation; NG contributed to methodology and investigation; DM contributed to methodology, data curation, and investigation; CO contributed to writing–review and editing, methodology, data curation, and investigation; IP contributed to methodology, data curation, and investigation; IH contributed to funding acquisition, methodology, data curation, and investigation; SL contributed to funding acquisition, methodology, data curation, and investigation; M Whitaker contributed to methodology, data curation, and investigation; LS contributed to funding acquisition, methodology, data curation, and investigation; DS contributed to funding acquisition, and writing–review and editing; SV contributed to funding acquisition, and writing–review and editing; GR contributed to funding acquisition, writing–review and editing, and methodology; AM contributed to funding acquisition, writing–review and editing, and methodology; S Hopkins contributed to funding acquisition, writing–review and editing, and methodology; A Sheikh contributed to funding acquisition, writing–review and editing, supervision, and methodology; and PQ contributed to funding acquisition, writing–original draft, writing–review and editing, and supervision.

Conflicts of Interest

A Sheikh is a member of the Scottish Government Chief Medical Officer’s COVID-19 Advisory Group and its Standing Committee on Pandemics. PQ was previously on a paid secondment to BC Platforms and now resides on their Scientific Advisory Board as a paid consultant. AA-B and PS work for BC Platforms, whose solution CO-CONNECT utilized.

References

1. Richter A, Plant T, Kidd M, Bosworth A, Mayhew M, Megram O, et al. How to establish an academic SARS-CoV-2 testing laboratory. Nat Microbiol 2020 Dec 02;5(12):1452-1454. [doi: 10.1038/s41564-020-00818-3] [Medline: 33139885]
2. Park S, Elliott J, Berlin A, Hamer-Hunt J, Haines A. Strengthening the UK primary care response to covid-19. BMJ 2020 Sep 25;370:m3691. [doi: 10.1136/bmj.m3691] [Medline: 32978177]
3. Oral evidence: UK Science, Research and Technology Capability and Influence in Global Disease Outbreaks, HC 136. Science and Technology Committee. 2020. URL: https://committees.parliament.uk/oralevidence/701/html/ [accessed 2021-08-14]
4. Better, broader, safer: using health data for research and analysis. GOV UK. URL: https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis [accessed 2022-05-13]
5. Cavallaro F, Lugg-Widger F, Cannings-John R, Harron K. Reducing barriers to data access for research in the public interest—lessons from covid-19. BMJ Opinion. 2020. URL: https://blogs.bmj.com/bmj/2020/07/06/reducing-barriers-to-data-access-for-research-in-the-public-interest-lessons-from-covid-19/ [accessed 2022-05-13]
6. Taylor JA, Crowe S, Espuny Pujol F, Franklin RC, Feltbower RG, Norman LJ, et al. The road to hell is paved with good intentions: the experience of applying for national data for linkage and suggestions for improvement. BMJ Open 2021 Aug 19;11(8):e047575-e047577 [FREE Full text] [doi: 10.1136/bmjopen-2020-047575] [Medline: 34413101]
7. Macnair A, Love SB, Murray ML, Gilbert DC, Parmar MKB, Denwood T, et al. Accessing routinely collected health data to improve clinical trials: recent experience of access. Trials 2021 May 10;22(1):340 [FREE Full text] [doi: 10.1186/s13063-021-05295-5] [Medline: 33971933]
8. The researchers' experience when attempting to access health data for research. NCRI. 2020. URL: https://www.ncri.org.uk/accessing-health-data-for-research/ [accessed 2022-05-13]
9. Larrucea X, Moffie M, Asaf S, Santamaria I. Towards a GDPR compliant way to secure European cross border Healthcare Industry 4.0. Computer Standards & Interfaces 2020 Mar;69:103408. [doi: 10.1016/j.csi.2019.103408]
10. Data sharing agreement template. NHS. 2020. URL: https://www.nhsx.nhs.uk/information-governance/guidance/data-sharing-agreement-template/ [accessed 2021-08-24]
11. Lin C, Stephens KA, Baldwin L, Keppel GA, Whitener RJ, Echo-Hawk A, et al. Developing Governance for Federated Community-based EHR Data Sharing. AMIA Jt Summits Transl Sci Proc 2014;2014:71-76 [FREE Full text] [Medline: 25717404]
12. Hubbard T, Reilly G, Varma S, Seymour D. Trusted Research Environments (TRE) Green Paper. Zenodo. 2020. URL: https://zenodo.org/record/4594704#.Y2UDxeRBzDc [accessed 2022-11-04]
13. Kruse CS, Smith B, Vanderlinden H, Nealand A. Security Techniques for the Electronic Health Records. J Med Syst 2017 Aug;41(8):127 [FREE Full text] [doi: 10.1007/s10916-017-0778-4] [Medline: 28733949]
14. Sloan P. The Compliance Case for Information Governance. Richmond Journal of Law & Technology 2014;20(2):Article 2 [FREE Full text]
15. Trifan A, Oliveira JL. Patient data discovery platforms as enablers of biomedical and translational research: A systematic review. J Biomed Inform 2019 May;93:103154 [FREE Full text] [doi: 10.1016/j.jbi.2019.103154] [Medline: 30922867]
16. Sharing Sensitive Health Data in a Federated Data Consortium Model: An Eight-Step Guide. World Economic Forum. 2020. URL: https://www.weforum.org/reports/sharing-sensitive-health-data-in-a-federated-data-consortium-model-an-eight-step-guide [accessed 2021-08-20]
17. Knoppers BM. Framework for responsible sharing of genomic and health-related data. Hugo J 2014 Dec;8(1):3 [FREE Full text] [doi: 10.1186/s11568-014-0003-1] [Medline: 27090251]
18. Dursi LJ, Bozoky Z, de Borja R, Li H, Bujold D, Lipski A, et al. CanDIG: Federated network across Canada for multi-omic and health data discovery and analysis. Cell Genomics 2021 Nov;1(2):100033. [doi: 10.1016/j.xgen.2021.100033]
19. CINECA - Common Infrastructure for National Cohorts in Europe, Canada, and Africa. URL: https://www.cineca-project.eu [accessed 2022-03-11]
20. European Health Data Evidence Network. URL: https://www.ehden.eu/ [accessed 2021-08-11]
21. Rahimzadeh V, Dyke SO, Knoppers BM. An International Framework for Data Sharing: Moving Forward with the Global Alliance for Genomics and Health. Biopreserv Biobank 2016 Jun;14(3):256-259. [doi: 10.1089/bio.2016.0005] [Medline: 27082668]
22. Dobbins NJ, Spital CH, Black RA, Morrison JM, de Veer B, Zampino E, et al. Leaf: an open-source, model-agnostic, data-driven web application for cohort discovery and translational biomedical research. J Am Med Inform Assoc 2020 Jan 01;27(1):109-118 [FREE Full text] [doi: 10.1093/jamia/ocz165] [Medline: 31592524]
23. 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;17(2):124-130 [FREE Full text] [doi: 10.1136/jamia.2009.000893] [Medline: 20190053]
24. Schüttler C, Prokosch H, Hummel M, Lablans M, Kroll B, Engels C, German Biobank Alliance IT Development Team. The journey to establishing an IT-infrastructure within the German Biobank Alliance. PLoS One 2021 Sep 22;16(9):e0257632 [FREE Full text] [doi: 10.1371/journal.pone.0257632] [Medline: 34551019]

25. Wolfson M, Wallace SE, Masca N, Rowe G, Sheehan NA, Ferretti V, et al. DataSHIELD: resolving a conflict in contemporary bioscience--performing a pooled analysis of individual-level data without sharing the data. Int J Epidemiol 2010 Oct 14;39(5):1372-1382 [FREE Full text] [doi: 10.1093/ije/dyq111] [Medline: 20630989]
26. BC Platforms. URL: https://www.bcplatforms.com [accessed 2021-07-23]
27. Stacey J, Mehta M. Using EHR Data Extraction to Streamline the Clinical Trial Process. Clinical Researcher. 2017. URL: https://acrpnet.org/2017/04/01/using-ehr-data-extraction-streamline-clinical-trial-process/ [accessed 2022-11-04]
28. Clinerion. URL: https://www.clinerion.com/index.html [accessed 2021-11-30]
29. Data science and AI in the age of COVID-19 – report. The Alan Turing Institute. URL: https://www.turing.ac.uk/research/publications/data-science-and-ai-age-covid-19-report [accessed 2021-08-11]
30. HDR UK Innovation Gateway. URL: https://www.healthdatagateway.org/ [accessed 2021-07-28]
31. Sebire NJ, Cake C, Morris AD. HDR UK supporting mobilising computable biomedical knowledge in the UK. BMJ Health Care Inform 2020 Jul 28;27(2):e100122 [FREE Full text] [doi: 10.1136/bmjhci-2019-100122] [Medline: 32723851]
32. COVID-19 National Core Studies. HDR UK. URL: https://www.hdruk.ac.uk/covid-19/covid-19-national-core-studies/ [accessed 2021-07-23]
33. Scientific Advisory Group for Emergencies. GOV UK. URL: https://www.gov.uk/government/organisations/scientific-advisory-group-for-emergencies [accessed 2022-10-04]
34. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
35. CO-CONNECT - Overview. YouTube. URL: https://www.youtube.com/watch?v=dqZtWi6VGGo [accessed 2022-09-30]
36. OMOP CDM v5.3. OHDSI GitHub. URL: https://ohdsi.github.io/CommonDataModel/cdm53.html#Clinical_Data_Tables [accessed 2022-05-13]
37. Cox S, Macdonald C, Lea D, Adejumo S, Panagi V, Tarr S, et al. HDRUK/CaRROT-Mapper: 2.0.1. Zenodo. 2022. URL: https://zenodo.org/record/6587677#.Y2UHT-RBzDc [accessed 2022-11-04]
38. Macdonald C, Panagi V, Tarr S, Santos R, Schlessinger D, Mumtaz S. HDRUK/CaRROT-CDM: CaRROT CDM Builder Version 0.6.0. Zenodo. 2022. URL: https://zenodo.org/record/6593954#.Y2UHl-RBzDc [accessed 2022-11-04]
39. OHDSI - Observational Health Data Sciences and Informatics. URL: https://www.ohdsi.org/ [accessed 2022-10-04]
40. White Rabbit. OHDSI GitHub. URL: http://ohdsi.github.io/WhiteRabbit/WhiteRabbit.html [accessed 2021-07-23]
41. Cox S, Macdonald C, Mumtaz S, Panagi V, Tarr S, Quinlan P. CaRROT-Docs. HDRUK GitHub. URL: https://hdruk.github.io/CaRROT-Docs/ [accessed 2022-03-10]
42. Cohort Discovery. HDR UK Innovation Gateway. URL: https://www.healthdatagateway.org/about/cohort-discovery [accessed 2022-03-10]
43. Ritchie F. The 'Five Safes': a framework for planning, designing and evaluating data access solutions. Zenodo. 2017. URL: https://zenodo.org/record/897821#.Y2UI3-RBzDc [accessed 2022-11-04]
44. The ‘Five Safes’ – Data Privacy at ONS. Office for National Statistics. URL: https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/ [accessed 2022-09-23]
45. Data Access Request Process Overview. HDR UK Innovation Gateway. URL: https://www.healthdatagateway.org/about/data-access-request-process [accessed 2022-02-03]
46. CO-CONNECT - Finding Data. YouTube. URL: https://www.youtube.com/watch?v=HAI3NhJsCKE [accessed 2022-09-30]
47. CO-CONNECT - Accessing and Analysing Data. YouTube. URL: https://www.youtube.com/watch?v=oc25k8WW440 [accessed 2022-09-30]
48. CO-CONNECT | Unleashing the power of data through discovery. URL: https://co-connect.ac.uk/ [accessed 2022-05-31]
49. Grandjean L, Saso A, Torres A, Lam T, Hatcher J, Thistlethwayte R, Co-Stars Study Team. Humoral Response Dynamics Following Infection with SARS-CoV-2. medRxiv. 2020. URL: https://www.medrxiv.org/content/10.1101/2020.07.16.20155663v2 [accessed 2022-11-04]
50. Augusto JB, Menacho K, Andiapen M, Bowles R, Burton M, Welch S, McKnight, et al. Healthcare Workers Bioresource: Study outline and baseline characteristics of a prospective healthcare worker cohort to study immune protection and pathogenesis in COVID-19. Wellcome Open Res 2020 Oct 12;5:179 [FREE Full text] [doi: 10.12688/wellcomeopenres.16051.2] [Medline: 33537459]
51. Abo-Leyah H, Gallant S, Cassidy D, Giam Y, Killick J, Marshall B, et al. The protective effect of SARS-CoV-2 antibodies in Scottish healthcare workers. ERJ Open Res 2021 Apr;7(2):5641 [FREE Full text] [doi: 10.1183/23120541.00080-2021] [Medline: 34104643]
52. Lumley SF, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, Hatch SB, Oxford University Hospitals Staff Testing Group. Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers. N Engl J Med 2021 Feb 11;384(6):533-540 [FREE Full text] [doi: 10.1056/NEJMoa2034545] [Medline: 33369366]
53. Valdes AM, Moon JC, Vijay A, Chaturvedi N, Norrish A, Ikram A, et al. Longitudinal assessment of symptoms and risk of SARS-CoV-2 infection in healthcare workers across 5 hospitals to understand ethnic differences in infection risk. EClinicalMedicine 2021 Apr;34:100835 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100835] [Medline: 33880438]

54. Hall VJ, Foulkes S, Charlett A, Atti A, Monk EJM, Simmons R, SIREN Study Group. SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: a large, multicentre, prospective cohort study (SIREN). Lancet 2021 Apr 17;397(10283):1459-1469 [FREE Full text] [doi: 10.1016/S0140-6736(21)00675-9] [Medline: 33844963]
55. The TRACK-COVID Study. TRACK-COVID. URL: https://www.trackcovid.org.uk/ [accessed 2021-07-23]
56. VIVALDI Study. UCL. URL: https://www.ucl.ac.uk/health-informatics/research/vivaldi-study [accessed 2021-07-23]
57. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, ISARIC4C investigators. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 2020 May 22;369:m1985 [FREE Full text] [doi: 10.1136/bmj.m1985] [Medline: 32444460]
58. COVID-19 surveillance in school KIDs (sKIDs): pre and primary schools. GOV UK. URL: https://www.gov.uk/government/publications/covid-19-surveillance-in-school-kids-skids-pre-and-primary-schools [accessed 2021-08-18]
59. Asymptomatic COVID-19 in Education (ACE) Cohort. HDR UK Innovation Gateway. URL: https://web.www.healthdatagateway.org/dataset/48adf432-9b66-4d03-a25e-008fbbb24f56 [accessed 2021-08-18]
60. Real-time Assessment of Community Transmission (REACT) Study. Imperial College London. URL: http://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/ [accessed 2021-07-18]
61. FOLLOW-COVID. NHS. URL: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/follow-covid/ [accessed 2022-11-16]
62. ATLAS: Advanced data search tool for researchers. UKCRC Tissue Directory and Coordination Centre. URL: https://biobankinguk.org/atlas/ [accessed 2021-07-23]
63. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol 2013 Feb;42(1):111-127 [FREE Full text] [doi: 10.1093/ije/dys064] [Medline: 22507743]
64. Smith B, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr S, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol 2013 Jun;42(3):689-700. [doi: 10.1093/ije/dys084] [Medline: 22786799]
65. Shrine N, Portelli MA, John C, Soler Artigas M, Bennett N, Hall R, et al. Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. Lancet Respir Med 2019 Jan;7(1):20-34 [FREE Full text] [doi: 10.1016/S2213-2600(18)30389-8] [Medline: 30552067]
66. NIHR BioResource. URL: https://bioresource.nihr.ac.uk/ [accessed 2021-09-03]
67. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. Int J Epidemiol 2013 Feb;42(1):76-85 [FREE Full text] [doi: 10.1093/ije/dyr207] [Medline: 22253318]
68. Cohorts - COVID-19 Longitudinal Health and Wellbeing National Core Study. UCL. URL: https://www.ucl.ac.uk/covid-19-longitudinal-health-wellbeing/cohorts [accessed 2021-09-03]
69. BREATHE - Health Data Research Hub. URL: https://www.breathedatahub.com/ [accessed 2021-07-23]
70. Data dashboards. NHS. URL: https://digital.nhs.uk/dashboards [accessed 2021-07-18]
71. Public Health England. GOV UK. URL: https://www.gov.uk/government/organisations/public-health-england [accessed 2021-07-18]
72. COVID-19. Public Health Scotland. URL: https://www.publichealthscotland.scot/our-areas-of-work/covid-19/ [accessed 2021-07-18]
73. Business Services Organisation. URL: https://hscbusiness.hscni.net/ [accessed 2021-08-25]
74. Public Health Agency. URL: https://www.publichealth.hscni.net/ [accessed 2021-08-25]
75. SAIL Databank - The Secure Anonymised Information Linkage Databank. URL: https://saildatabank.com/ [accessed 2021-07-18]
76. Coronavirus (COVID-19). Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases [accessed 2021-07-18]
77. Milligan G, Johnston J, Horban S, Masood E, Cox S, Urwin E, et al. COVID - Curated and OpeN aNalysis and rEsearCh plaTform (CO-CONNECT) Implementation Guide. Zenodo. 2022. URL: https://zenodo.org/record/7150536#.Y2U34uRBzDc [accessed 2022-11-04]
78. Vezyridis P, Timmons S. Understanding the care.data conundrum: New information flows for economic growth. Big Data & Society 2017 Jan 01;4(1):205395171668849. [doi: 10.1177/2053951716688490]
79. Church J. GP Data for Planning and Research: Letter from Parliamentary Under Secretary of State for Health and Social Care to general practices in England. NHS Digital. 2021. URL: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research/secretary-of-state-letter-to-general-practice [accessed 2021-11-10]
80. Chapter 6 Extract Transform Load - The Book of OHDSI. OHDSI GitHub. URL: https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html [accessed 2022-03-11]
81. Raisaro JL, Marino F, Troncoso-Pastoriza J, Beau-Lejdstrom R, Bellazzi R, Murphy R, et al. SCOR: A secure international informatics infrastructure to investigate COVID-19. J Am Med Inform Assoc 2020 Nov 01;27(11):1721-1726 [FREE Full text] [doi: 10.1093/jamia/ocaa172] [Medline: 32918447]

82. Raisaro JL, Troncoso-Pastoriza JR, Misbach M, Sousa JS, Pradervand S, Missiaglia E, et al. MedCo: Enabling Secure and Privacy-Preserving Exploration of Distributed Clinical and Genomic Data. IEEE/ACM Trans Comput Biol Bioinform 2019;16(4):1328-1341. [doi: 10.1109/TCBB.2018.2854776] [Medline: 30010584]
83. Brat GA, Weber GM, Gehlenborg N, Avillach P, Palmer NP, Chiovato L, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. NPJ Digit Med 2020;3:109 [FREE Full text] [doi: 10.1038/s41746-020-00308-0] [Medline: 32864472]
84. Prokosch H, Bahls T, Bialke M, Eils J, Fegeler C, Gruendner J, et al. The COVID-19 Data Exchange Platform of the German University Medicine. Stud Health Technol Inform 2022 May 25;294:674-678. [doi: 10.3233/SHTI220554] [Medline: 35612174]
85. Sedlmayr B, Sedlmayr M, Kroll B, Prokosch H, Gruendner J, Schüttler C. Improving COVID-19 Research of University Hospitals in Germany: Formative Usability Evaluation of the CODEX Feasibility Portal. Appl Clin Inform 2022 Mar;13(2):400-409 [FREE Full text] [doi: 10.1055/s-0042-1744549] [Medline: 35445386]
86. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, N3C Consortium. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc 2021 Mar 01;28(3):427-443 [FREE Full text] [doi: 10.1093/jamia/ocaa196] [Medline: 32805036]
87. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014;21(4):578-582 [FREE Full text] [doi: 10.1136/amiajnl-2014-002747] [Medline: 24821743]
88. Visweswaran S, Becich M, D'Itri VS, Sendro E, MacFadden D, Anderson N, et al. Accrual to Clinical Trials (ACT): A Clinical and Translational Science Award Consortium Network. JAMIA Open 2018 Oct;1(2):147-152 [FREE Full text] [doi: 10.1093/jamiaopen/ooy033] [Medline: 30474072]
89. Visweswaran S, Samayamuthu MJ, Morris M, Weber GM, MacFadden D, Trevvett P, et al. Development of a COVID-19 Application Ontology for the ACT Network. medRxiv 2021 Apr 14:2021.03.15.21253596 [FREE Full text] [doi: 10.1101/2021.03.15.21253596] [Medline: 33791734]
90. Weber GM, Murphy SN, McMurry AJ, MacFadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): A Prototype Federated Query Tool for Clinical Data Repositories. Journal of the American Medical Informatics Association 2009 Sep 01;16(5):624-630. [doi: 10.1197/jamia.m3191]
91. Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. AMIA Jt Summits Transl Sci Proc 2013;2013:6-8 [FREE Full text] [Medline: 24303286]
92. Peñalvo JL, Mertens E, Ademović E, Akgun S, Baltazar AL, Buonfrate D, unCoVer Network. Unravelling data for rapid evidence-based response to COVID-19: a summary of the unCoVer protocol. BMJ Open 2021 Nov 18;11(11):e055630 [FREE Full text] [doi: 10.1136/bmjopen-2021-055630] [Medline: 34794999]
93. Alleviate – our Advanced Pain Discovery Platform (APDP) Hub. HDR UK. URL: https://www.hdruk.ac.uk/helping-with-health-data/our-hubs-across-the-uk/alleviate/ [accessed 2021-07-29]

Abbreviations

CaRROT: Convenient and Reusable Rapid Ontology Transformer
CDM: common data model
CO-CONNECT: COVID - Curated and Open Analysis and Research Platform
EHR: electronic health record
ETL: Extract, Transform, and Load
GDPR: General Data Protection Regulations
HDR: Health Data Research
NHS: National Health Service
OHDSI: Observational Health Data Science and Informatics
OMOP: Observational Medical Outcomes Partnership
PheWAS: phenome-wide association study
TRE: Trusted Research Environment
VM: virtual machine

Edited by G Eysenbach; submitted 07.06.22; peer-reviewed by Z Zrubka, HU Prokosch; comments to author 11.09.22; revised version received 12.10.22; accepted 01.11.22; published 27.12.22.

Please cite as:
Jefferson E, Cole C, Mumtaz S, Cox S, Giles TC, Adejumo S, Urwin E, Lea D, Macdonald C, Best J, Masood E, Milligan G, Johnston J, Horban S, Birced I, Hall C, Jackson AS, Collins C, Rising S, Dodsley C, Hampton J, Hadfield A, Santos R, Tarr S, Panagi V, Lavagna J, Jackson T, Chuter A, Beggs J, Martinez-Queipo M, Ward H, von Ziegenweidt J, Burns F, Martin J, Sebire N, Morris C, Bradley D, Baxter R, Ahonen-Bishopp A, Smith P, Shoemark A, Valdes AM, Ollivere B, Manisty C, Eyre D, Gallant S, Joy G, McAuley A, Connell D, Northstone K, Jeffery K, Di Angelantonio E, McMahon A, Walker M, Semple MG, Sims JM, Lawrence E, Davies B, Baillie JK, Tang M, Leeming G, Power L, Breeze T, Murray D, Orton C, Pierce I, Hall I, Ladhani S, Gillson N, Whitaker M, Shallcross L, Seymour D, Varma S, Reilly G, Morris A, Hopkins S, Sheikh A, Quinlan P
A Hybrid Architecture (CO-CONNECT) to Facilitate Rapid Discovery and Access to Data Across the United Kingdom in Response to the COVID-19 Pandemic: Development Study
J Med Internet Res 2022;24(12):e40035
URL: https://www.jmir.org/2022/12/e40035
doi:10.2196/40035
PMID:36322788

Viewpoint

Social Media for Public Health: Framework for Social Media–Based Public Health Campaigns

Isabella de Vere Hunt[1,2], BMBCh; Eleni Linos[1], MD, MPH, DrPH

[1]Program for Clinical Research and Technology, Stanford University, Palo Alto, CA, United States

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Isabella de Vere Hunt, BMBCh
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building
Radcliffe Observatory Quarter
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865289300
Email: isabella.deverehunt@phc.ox.ac.uk

Abstract

The pervasiveness of social media is irrefutable, with 72% of adults reporting using at least one social media platform and an average daily usage of 2 hours. Social media has been shown to influence health-related behaviors, and it offers a powerful tool through which we can rapidly reach large segments of the population with tailored health messaging. However, despite increasing interest in using social media for dissemination of public health messaging and research exploring the dangers of misinformation on social media, the specifics of how public health practitioners can effectively use social media for health promotion are not well described. In this viewpoint, we propose a novel framework with the following 5 key principles to guide the use of social media for public health campaigns: (1) tailoring messages and targeting them to specific populations—this may include targeting messages to specific populations based on age, sex, or language spoken; interests; or geotargeting messages at state, city, or zip code level; (2) including members of the target population in message development—messages should be designed with and approved by members of the community they are designed to reach, to ensure cultural sensitivity and trust-building; (3) identifying and addressing misinformation—public health practitioners can directly address misinformation through myth-busting messages, in which false claims are highlighted and explained and accurate information reiterated; (4) leveraging information sharing—when designing messages for social media, it is crucial to consider their "shareability," and consider partnering with social media influencers who are trusted messengers among their online followers; and (5) evaluating impact by measuring real-world outcomes, for example measuring foot traffic data. Leveraging social media to deliver public health campaigns enables us to capitalize on sophisticated for-profit advertising techniques to disseminate tailored messaging directly to communities that need it most, with a precision far beyond the reaches of conventional mass media. We call for the Centers for Disease Control and Prevention as well as state and local public health agencies to continue to optimize and rigorously evaluate the use of social media for health promotion.

(J Med Internet Res 2022;24(12):e42179) doi:10.2196/42179

KEYWORDS

social media; digital heath; health communication; campaign; public health; framework; health promotion; public awareness; misinformation; tailored message; tailored messaging; information sharing; information exchange; advertise; advertising

Introduction

The pervasiveness of social media is irrefutable; 72% of adults and 84% of adults aged 18-29 years report use of at least one type of social media platform and an average daily usage of 2 hours [1,2]. Social media has been shown to influence public opinion, political views, and purchasing behaviors, as well as health-related behaviors such as diet and exercise [3-6]. This offers a powerful tool through which we can rapidly reach large segments of the population with tailored health messaging. It has also become a potent arena for widespread dissemination of misinformation and disinformation, posing its own public health threat [7-9]. However, although numerous studies have

explored the dangers of misinformation and have used data from social media to interpret public attitudes [10], the specifics of how public health practitioners can effectively use social media for health promotion are not described. In this viewpoint, we propose a new framework with the following 5 key principles to guide the use of social media for public health campaigns: (1) tailoring messages and targeting them to specific populations; (2) including members of the target population in message development; (3) identifying and addressing misinformation; (4) leveraging information sharing; and (5) evaluating impact by measuring real-world outcomes.

Tailoring Messages and Targeting Them to Specific Populations

In addition to connecting individuals, social media platforms, including Facebook, Instagram, Twitter, and TikTok, have sophisticated advertising platforms that facilitate targeting of advertisements to specified populations with far greater precision than conventional mass media. For example, a multivitamin company can send advertisements to women aged 25-40 years living in Manhattan with a household income in the top 10% in the United States and an interest in organic food. Researchers have used these targeted advertising tools for the purpose of study recruitment [11]. For example, Reiter et al [12] used Facebook's targeted advertising to recruit young gay and bisexual men for a human papillomavirus vaccination intervention by selecting for English-speaking males in the United States aged 18-25 years with any of the following "interest" filters: bisexuality; homosexuality; same-sex relationship; genderqueer; gay pride; lesbian, gay, bisexual, and transgender (LGBT) community; LGBT culture; or rainbow (LGBT movement) [12]. Public health practitioners can leverage ad targeting to send tailored health messages to specific populations based on age, sex, or language spoken; interests, such as "smoking," "aerobic exercise," or "McDonalds"; or geotarget messages at state, city, or zip code level. Targeting messages by language spoken is particularly relevant to immigrant and refugee communities, for whom language barriers may limit understanding of alternative sources of health information; for example, in Germany, social media advertisements in migrants' languages of origin increased COVID-19 vaccine appointment bookings by 133% for Arabic speakers and 76% for Russian speakers [13]. Of note, Facebook continually reviews the available ad targeting options and ad controls with the aim of reducing the possibility of ad discrimination; for example, in January 2022 they removed previously available detailed targeting options that "relate to topics people may perceive as sensitive," which include health causes such as "lung cancer awareness" and "chemotherapy" [14]. We therefore recommend frequent monitoring of targeting options when planning campaign implementation to ascertain what will be available for use.

A key challenge in using social media for public health is that algorithms are designed to present advertisements a person is likely to agree or engage with [15], but in public health, we often seek to reach those who disagree, for example, convincing a smoker to quit or a reluctant parent to vaccinate their child. One way to address this is to tailor messages and separate them into narrower ad sets for specific populations. For instance, although COVID-19 vaccine promotion messages might be unpopular among vaccine-hesitant groups, we can increase message salience by tailoring them to subsets of the target population—a message debunking fertility concerns could be sent to women aged 25-30 years with an interest in "motherhood"; a video by a Spanish-speaking doctor could be delivered to Spanish-speaking adults in a zip code area with low vaccination rates; and a video by a Methodist priest could be sent to people interested in the "Methodist church" [16]. The ability to rapidly pilot-test multiple iterations can identify the most engaging messages for each group. By structuring campaigns into ad sets, we can also allocate more budget to populations who need it most; for example, using indices such as the California Health Place Index, we can preferentially allocate funds to lower health index zip code areas. An et al [11] propose a useful precision public health campaign framework to guide the use of targeted advertising tools on social media to deliver tailored health messages to particular population segments.

Including Members of the Target Population in Message Development

Engaging community partners when designing public health messaging is paramount in building trust and ensuring effectiveness. Messages should be designed with and approved by members of the community they are designed to reach, to ensure cultural sensitivity and trust-building. One way to achieve this is to assemble an advisory board, including members of the target population, and reflecting the demographics of users of the intended platform. For example, approximately 43% of TikTok's audience is 18-24 years of age, and only 3% is aged >55 years [17]; thus, input from younger, Gen Z voices would be crucial for a campaign running on TikTok. Further, any ad targeting strategies should be transparent and sensitive to the potential for discriminatory ad targeting. Indeed, journalists have demonstrated, with historical ad targeting options available on Facebook, how easy it would be to exclude users whom Facebook classifies as a member of a racial or ethnic minority group from target audiences [18]. Although race categories have subsequently been removed from explicit targeting options on Facebook, the ability to direct ads to specific racial groups is still implicitly possible (to varying degrees of accuracy) via proxies such as zip code targeting. Therefore, we call for discussion with and approval by an advisory board of any proposed ad targeting strategies, alongside a clearly documented rationale that aims to benefit the target audience, prior to campaign implementation. When translating messages, using input from native speakers to ensure optimal language choice rather than relying on automated translations is crucial.

Identifying and Addressing Misinformation

Misinformation on social media has been shown to influence health attitudes; in a randomized controlled trial assessing the effect of web-based misinformation on COVID-19 vaccine intentions, recent exposure to misinformation decreased vaccine

intent by 6.4% among participants who previously stated they would definitely accept a vaccine [19]. The extent to which social media facilitates dissemination of misinformation was exemplified by infodemics—defined by the World Health Organization as an overabundance of both inaccurate and accurate information—during the COVID-19 pandemic [20,21]. Vosoughi et al [8] hypothesize that false news reaches more people than the truth does because it has a higher degree of novelty and provokes stronger emotional reactions of recipients, making it more likely to be passed on. Public health practitioners can directly address misinformation through myth-busting messages, in which false claims are highlighted and explained and accurate information reiterated. This should be an iterative process, beginning with message design and continuing through active comment moderation, including direct responses to false comments during a live campaign. In a randomized controlled trial of messages debunking highly prevalent health information in Sierra Leone, direct and detailed debunking was most effective [22]. Live interactions are a key part of how information is disseminated on social media, yet traditional mass media communication models do not account for this interactivity. Parackal et al [23] propose the dynamic transactional model of communication as a suitable framework for modelling the "two-way communication" in which both the sender and the receiver actively participate in the communication process that takes place on social media [23].

Leveraging Information Sharing by the Target Population

Trusted messengers, including healthcare providers, religious leaders, and celebrities, can play an important role in public health messaging; for example, basketball player Magic Johnson's announcement of his HIV-positive status in 1991 was correlated with increased condom use among Black and Hispanic individuals [24]. Yet social media differs from traditional broadcast media in the rapidity at which messages disseminated from an original source can be publicly reshared by the target population. In a survey experiment on Facebook of 1489 adults, 51% reported that a health article on diabetes was well reported and trustworthy when it was shared by a public figure they trusted, whereas only 34% thought the same article was trustworthy when it was shared by someone they did not trust [25]. When designing messages for social media, it is crucial to consider their "shareability"—can the message be designed in a way that encourages users to share it with their friends? Partnering with social media influencers—users of social media with established credibility among their followers—is a useful approach for leveraging trusted messengers.

Evaluating Impact by Measuring Real-world Outcomes

Evaluation of social media–based public health campaigns should include measurement of the health-related behavior of interest in the target population. Breza et al [26] provide an excellent example, as follows: in a cluster randomized controlled trial, investigators used distance travelled in treatment regions, measured using mobile phone location data of Facebook users, as well as COVID-19 infections recorded at the zip code level, as the outcome measures to assess the impact of a social media advertising campaign asking participants to avoid holiday travel to reduce COVID-19 infections. In our own work, we have piloted an approach using analysis of foot traffic data to tanning salons as an outcome measure to assess the impact of a social media campaign aiming to reduce indoor tanning. Social media platforms also record web-based evaluation metrics, including number of people reached, average duration of videos viewed, reactions, shares, unique link clicks, and cost per individual reached. How these web-based metrics map onto real-world behaviors is unclear; reporting of web-based outcome measures alongside real-world measures can improve our understanding of how these metrics correlate with health-related behavioral change.

Conclusions

Leveraging social media to deliver public health campaigns enables us to capitalize on sophisticated for-profit advertising techniques to disseminate tailored messaging directly to communities that need it most, with a precision far beyond the reaches of conventional mass media. We do not present social media as a public health panacea; grave concerns about cyberbullying, privacy breaches, and misinformation on social media must be addressed in parallel [27]. Further, collaboration between public health scientists and technology companies will be vital to support widespread and potentially expensive ad campaigns, with the success of such partnerships highlighted by extensive COVID-19 vaccine promotion efforts supported by Facebook ad credits [16]. However, in a nation in which three-quarters of adults use social media, for some of whom social media will be the only source of health information, the Centers for Disease Control and Prevention as well as state and local public health agencies must optimize and rigorously evaluate its use for health promotion.

Acknowledgments

IDVH is supported by the National Institute of Health Research (Academic Clinical Fellowship), and EL is supported by the National Institutes of Health (grants DP2CA225433 and K24AR075060).

Authors' Contributions

IDVH wrote the original draft of this viewpoint, under the supervision of EL. Both authors contributed equally to multiple rounds of redrafting and edits.

Conflicts of Interest

None declared.

References

1. Social media fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2022-01-26]
2. Henderson G. How much time does the average person spend on social media? Digital Marketing. 2020 Aug 24. URL: https://www.digitalmarketing.org/blog/how-much-time-does-the-average-person-spend-on-social-media [accessed 2022-02-08]
3. Zhuravskaya E, Petrova M, Enikolopov R. Political effects of the internet and social media. Annu Rev Econ 2020 Aug 02;12(1):415-438. [doi: 10.1146/annurev-economics-081919-050239]
4. Yom-Tov E, Shembekar J, Barclay S, Muennig P. Erratum: author correction: the effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants. NPJ Digit Med 2018 Aug 16;1(1):38 [FREE Full text] [doi: 10.1038/s41746-018-0047-z] [Medline: 31305589]
5. Is social media changing the doctor-patient relationship? AAFP. URL: https://www.aafp.org/news/practice-professional-issues/20180608commsurveys.html [accessed 2022-02-24]
6. Sarkar U, Le GM, Lyles CR, Ramo D, Linos E, Bibbins-Domingo K. Using social media to target cancer prevention in young adults: viewpoint. J Med Internet Res 2018 Jun 05;20(6):e203 [FREE Full text] [doi: 10.2196/jmir.8882] [Medline: 29871850]
7. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine 2018 Jun 14;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040] [Medline: 29773322]
8. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Dec 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
9. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
10. 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 Digital Health 2021 Mar;3(3):e175-e194. [doi: 10.1016/s2589-7500(20)30315-0]
11. An J, Kwak H, Qureshi HM, Weber I. Precision public health campaign: delivering persuasive messages to relevant segments through targeted advertisements on social media. JMIR Form Res 2021 Sep 24;5(9):e22313 [FREE Full text] [doi: 10.2196/22313] [Medline: 34559055]
12. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Recruiting young gay and bisexual men for a human papillomavirus vaccination intervention through social media: the effects of advertisement content. JMIR Public Health Surveill 2017 Jun 02;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]
13. New study highlights social media's key role in COVID-19 vaccine uptake for migrants. IOM Global Migration Data Analysis Centre. 2022 Feb 11. URL: https://gmdac.iom.int/news/new-study-highlights-social-medias-key-role-covid-19-vaccine-uptake-migrants [accessed 2022-02-16]
14. Removing certain ad targeting options and expanding our ad controls. Meta. 2021 Nov 09. URL: https://www.facebook.com/business/news/removing-certain-ad-targeting-options-and-expanding-our-ad-controls [accessed 2022-11-20]
15. Oremus W, Alcantara C, Merrill JB, Galocha A. How Facebook shapes your feed. The Washington Post. 2021 Oct 26. URL: https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works/ [accessed 2022-02-23]
16. de Vere Hunt I, Dunn T, Mahoney M, Chen M, Nava V, Linos E. A social media based public health campaign encouraging COVID-19 vaccination across the United States. Am J Public Health 2022 Sep;112(9):1253-1256. [doi: 10.2105/ajph.2022.306934]
17. Omnicore. 2022 Mar 13. URL: https://www.omnicoreagency.com/tiktok-statistics/ [accessed 2022-05-03]
18. Angwin J, Parris Jr T. Facebook lets advertisers exclude users by Race. ProPublica. 2016 Oct 28. URL: https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race [accessed 2022-11-24]
19. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Feb 05:337-348. [doi: 10.1038/s41562-021-01056-1]
20. Novel Coronavirus (2019-nCov) situation report - 13. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf [accessed 2022-03-05]
21. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nat Hum Behav 2020 Dec 29;4(12):1285-1293. [doi: 10.1038/s41562-020-00994-6] [Medline: 33122812]
22. Winters M, Oppenheim B, Sengeh P, Jalloh MB, Webber N, Pratt SA, et al. Debunking highly prevalent health misinformation using audio dramas delivered by WhatsApp: evidence from a randomised controlled trial in Sierra Leone. BMJ Glob Health 2021 Nov 10;6(11):e006954 [FREE Full text] [doi: 10.1136/bmjgh-2021-006954] [Medline: 34758970]
23. Parackal M, Parackal S, Mather D, Eusebius S. Dynamic transactional model: a framework for communicating public health messages via social media. PPH 2021 Sep 14;141(5):279-286. [doi: 10.1177/1757913920935910] [Medline: 32660353]
24. Moskowitz JT, Binson D, Catania JA. The association between magic Johnson's HIV serostatus disclosure and condom use in at - risk respondents. JSR 1997 Jan;34(2):154-160. [doi: 10.1080/00224499709551879]

25. Who shared it? How Americans decide what news to trust on social media. AP-NORC at the University of Chicago. URL: https://apnorc.org/projects/who-shared-it-how-americans-decide-what-news-to-trust-on-social-media/ [accessed 2022-02-24]
26. Breza E, Stanford FC, Alsan M, Alsan B, Banerjee A, Chandrasekhar AG, et al. Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial. Nat Med 2021 Aug 19;27(9):1622-1628 [FREE Full text] [doi: 10.1038/s41591-021-01487-3] [Medline: 34413518]
27. Anderson M. A Majority of teens have experienced some form of cyberbullying. Pew Research Center. 2018 Sep 27. URL: https://www.pewresearch.org/internet/2018/09/27/a-majority-of-teens-have-experienced-some-form-of-cyberbullying/ [accessed 2022-03-03]

Abbreviations

LGBT: lesbian, gay, bisexual, and transgender

Edited by T Leung; submitted 25.08.22; peer-reviewed by S Beasley, H Cho, Y Mejova, P Shafer; comments to author 17.11.22; revised version received 29.11.22; accepted 30.11.22; published 14.12.22.

Please cite as:
de Vere Hunt I, Linos E
Social Media for Public Health: Framework for Social Media–Based Public Health Campaigns
J Med Internet Res 2022;24(12):e42179
URL: https://www.jmir.org/2022/12/e42179
doi:10.2196/42179
PMID:36515995

Tutorial

Avoiding Under- and Overrecruitment in Behavioral Intervention Trials Using Bayesian Sequential Designs: Tutorial

Marcus Bendtsen[1], PhD

Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Marcus Bendtsen, PhD
Department of Health, Medicine and Caring Sciences
Linköping University
Linköping, 581 83
Sweden
Phone: 46 13 28 10 00
Email: marcus.bendtsen@liu.se

Abstract

Reducing research waste and protecting research participants from unnecessary harm should be top priorities for researchers studying interventions. However, the traditional use of fixed sample sizes exposes trials to risks of under- and overrecruitment by requiring that effect sizes be determined a priori. One mitigating approach is to adopt a Bayesian sequential design, which enables evaluation of the available evidence continuously over the trial period to decide when to stop recruitment. Target criteria are defined, which encode researchers' intentions for what is considered findings of interest, and the trial is stopped once the scientific question is sufficiently addressed. In this tutorial, we revisit a trial of a digital alcohol intervention that used a fixed sample size of 2129 participants. We show that had a Bayesian sequential design been used, the trial could have ended after collecting data from approximately 300 participants. This would have meant exposing far fewer individuals to trial procedures, including being allocated to the waiting list control condition, and the evidence from the trial could have been made public sooner.

(J Med Internet Res 2022;24(12):e40730) doi:10.2196/40730

KEYWORDS

digital alcohol intervention; Bayesian sequential design; sample size; randomized controlled trial; trial recruitment; behavioural intervention; participant recruitment; research participants; research methods; effect size; trial procedure

Introduction

Overview

Substantial effort is often expended on recruiting and collecting data from participants in behavioral intervention trials. Delivering interventions to participants often incur additional costs that need to be considered in restricted budgets. These efforts and costs need to be balanced with study objectives, as increasing the number of participants leads to reduced uncertainty in effect estimates. It is, therefore, not surprising that sample size considerations are given serious attention during the planning of trials, mixed in with feelings of despair, disbelief, and above all, hope.

With misguided faith in null hypothesis testing delivering certainty about effects in otherwise uncertain circumstances [1], power calculations to determine sample sizes have become a staple in study protocols as well as in ethics approval and grant applications. Fixating the risks of false negatives and false positives (power and significance) at widely adopted rates, researchers conducting power calculations tend to focus on the magnitude of effects they wish to detect as the variable dictating sample sizes. However, effects of interventions are uncertain, which is precisely why trials are conducted in the first place, and so deciding on the magnitude of effect a priori is in practice impossible. What sometimes then happens is that researchers, in fear of underrecruiting and not having enough power to detect statistically significant effects, pick the smallest effect size that they would not want to miss [2,3]. This results in unnecessary costs and efforts to recruit, intervene, and collect data from participants if effect sizes turn out to be greater than this minimal effect size. Other times, the effect size is assumed to be unreasonably large to reduce the required sample size and convince ethics and grant boards that the trial is feasible [2]. This leads to underrecruiting, and it leads to the null hypothesis not being rejected, and as is often the case, misinterpreted to be evidence of no effect, despite the existence of an observable difference between groups [4,5].

Over- and underrecruiting participants is both costly and unethical [3]. It leads to subjecting more than necessary

participants to unnecessary effects from study procedures [6], potentially harmful or noneffective interventions and control conditions [7-9], or ending a trial with ambiguous findings when recruiting more participants could deliver less uncertain evidence [10]. One solution is to abandon a priori fixed sample sizes altogether, letting the data collected during the trial dictate when recruitment should end. Bayesian sequential designs are examples of this approach [11-13], where data are continuously analyzed and decisions are made throughout the trial period on whether or not recruitment should end.

Objective

The objective of this study is to demonstrate how a recently completed trial of a digital alcohol intervention would have played out had a Bayesian sequential design been used, rather than following a traditional fixed sample size based on a priori power calculations. We will show that participants were excessively overrecruited, resulting in costs and efforts wasted when the evidence was already at hand.

Bayesian Statistics and Sequential Designs

The literature on Bayesian statistics and sequential designs is substantial [1,10-13], and readers should have no problem finding in-depth descriptions. Therefore, we will introduce both, while at the same time assuring readers that they should feel comfortable moving on to the real-world examples even if not all details in this section are understood.

Bayesian Statistics

To understand Bayesian sequential designs, one needs to have at least a general understanding of Bayesian statistics. Within the Bayesian paradigm, one is interested in estimating the *posterior probability distribution* of parameters. In trials, the parameter that is given the most attention is the one that represents the effect of the intervention. The posterior probability distribution tells us how likely different parameter estimates are relative to one another. For instance, in a trial of a smoking cessation intervention, we could report the probability that the odds ratio (OR) of successful smoking cessation is greater than 1, that it is greater than 1.5, or that it lies between 0.9 and 1.1, and so on. As a concrete example, Figure 1 show two posterior probability distributions over OR estimated from a trial of a digital smoking cessation intervention among high school students [14-16]. The posterior distributions in Figure 1 show us that the effect of the intervention on 8-week prolonged abstinence from cigarettes 3 months post baseline (left plot) was approximately 1.2, and that 73.8% of the posterior probability distribution was above an OR of 1—leaving some uncertainty about the effectiveness of the intervention on this outcome measure. The right plot in Figure 1 shows that the OR for 4-week point prevalence of abstinence from smoking was approximately 1.8, and that 98.4% of the posterior probability distribution was above an OR of 1, suggesting strong evidence that there was a difference between groups with respect to this outcome measure.

The posterior probability distribution is calculated by combining the information available through the data collected, with what is known as the *prior probability distribution*—known simply as the *prior*. The prior represents our belief regarding the parameters before we collect data (ie, in a trial, the prior represents our belief about the effects before the trial commenced). The prior can be used to take a skeptical stance regarding effects by centering the prior around the null or to incorporate findings from previous studies by centering the prior around effect sizes estimated in previous trials. When data are scarce, the prior will influence the posterior distribution to a larger extent but will fade away as more data are collected. When using skeptical priors, this means that effect estimates are pulled toward the null when data are scarce, which is a powerful method of ensuring that conclusions of effects are not drawn prematurely using small sample sizes.

To illustrate this, Figure 2 shows three prior distributions. In Figure 2A, a prior distribution in the form of a normal distribution with a mean of 2 (SD 1) is shown. This prior says that, before we see any data, we believe that the effect of the intervention being studied is most likely to be around 2 but we also are uncertain about this, encoded by the width of the distribution. In Figure 2B, a prior distribution in the form of a normal distribution with a mean of 0 (SD 1) is shown. This prior says that, before we see any data, we believe that the effect of the intervention is most likely to be around 0 (ie, taking a skeptical stance); similarly, we are encoding uncertainty in this assumption through the width of the distribution. Finally, in Figure 2C, a prior distribution in the form of a normal distribution is shown, which is centered at 0 but has a SD of 0.1 and is therefore much narrower than the normal distributions in Figure 2A and Figure 2B. This prior encodes a strong belief that the effect is close to 0. In all cases, the priors influence on the final posterior probability distribution will be strongest when there are fewer data points; thus, when the sample size grows, the data speak louder than the prior, and our prior beliefs will be overridden by the data.

Figure 1. Marginal posterior distributions of odds ratios for smoking cessation (prolonged abstinence and point prevalence of smoking abstinence)—comparing study participants who had access to a digital smoking cessation intervention versus waiting list control group participants.

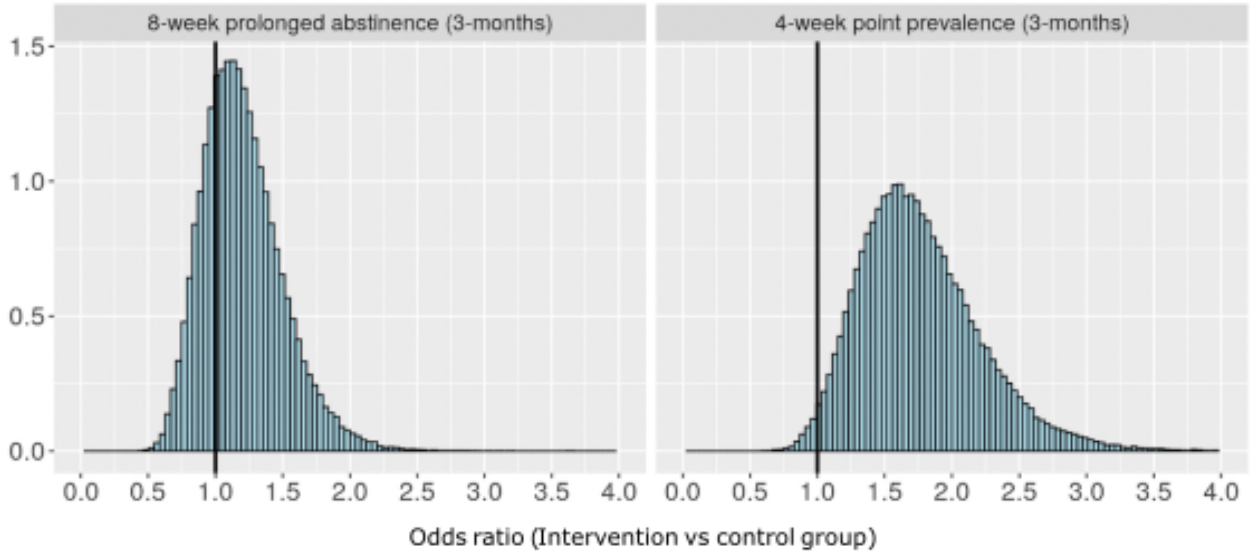
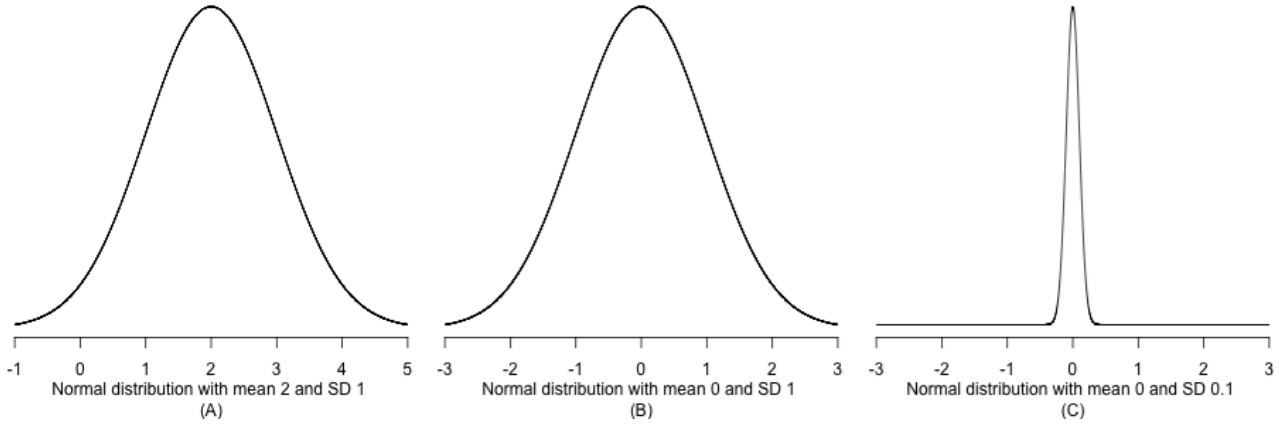


Figure 2. Examples of prior distributions; (A) normal distribution with mean 2 and SD 1; (B) normal distribution with mean 0 and SD 1; (C) normal distribution with mean 0 and SD 0.1.



Bayesian Sequential Designs

Rather than targeting a fixed sample size, a trial adopting a Bayesian sequential design aims to recruit enough participants so that the posterior distribution of the effect estimate is informative relative to the study objectives. For instance, in a trial of a smoking cessation intervention, where our main concern is the OR of abstinence, we may decide that we want to show that the posterior probability of the OR being greater than 1 is at least 89% (or any other probability we find sufficient relative to the study context). Therefore, we collect data and continuously analyze it until we have reduced the uncertainty enough so that we can show that the OR is greater than 1 with at least 89% probability. There is, however, no need to have only one target; rather, it is often reasonable to include at least one more target defining when the intervention seems ineffective and it is futile to continue the trial. An example of this would be if the posterior probability is at least 92% that the OR is greater than 0.9 and less than 1.1 (ie, close to the null). The targets, often referred to as *criteria*, are succinctly expressed using formal notation. Thus, for the smoking cessation intervention trial example given above, the target criteria could be as follows:

- Effect: $p (OR > 1 | D) > 89\%$
- Futility: $p (0.9 < OR < 1.1 | D) > 92\%$
- Harm: $p (OR < 1 | D) > 89\%$

Note that criteria should be defined relative to the study objectives, the context in which they are evaluated, and their potential benefits and harms. If one was evaluating the effects of a surgical procedure, perhaps the 89% probability of effect should be closer to 98% probability, while the probability for harm should perhaps be revised down to 75%.

A Trial of a Digital Alcohol Intervention

Overview

To demonstrate how a trial may develop using a Bayesian sequential design in contrast to a fixed sample size, we revisit a randomized trial of a digital alcohol intervention [17,18]. The effects of the intervention were estimated using a 2-arm parallel group trial, where one group was given access to the intervention for 4 months, while the other group was given information about alcohol and health aimed to motivate them to drink less and given access to the intervention after the trial. The trial was prospectively registered in the ISRCTN registry (48317451).

Ethics Approval

The trial received ethics approval on November 6, 2018, by the regional ethical committee in Linköping, Sweden (DNR 2018/417-31).

Study Procedures

In this tutorial, we will only give a brief overview of the trial procedures; a full description of the trial is available in the study protocol [18]. The target population was Swedish-speaking adults seeking help on the internet to reduce their alcohol consumption. Individuals were required to be at least 18 years of age, have access to a mobile phone, and be classified as risky drinkers according to Swedish guidelines. Participants who showed interest in the study and gave informed consent were asked to respond to a baseline questionnaire (which also assessed eligibility) and were subsequently randomized. Participants were not blind after allocation, as they were aware whether or not they received immediate access to the digital intervention.

The core element of the digital intervention was a text message sent to participants each Sunday afternoon. The text message included a prompt to self-monitor one's current alcohol consumption, with a hyperlink to a web-based tool. Those who decided to click on the link were asked to report their recent drinking and were then given access to personalized support. More information on the intervention is available in the study protocol [18].

Participants allocated to the control group were advised that they would receive information designed to motivate them to think more about reducing their alcohol consumption and that after 4 months they would receive additional support delivered to their mobile phone. Participants in the control group also received a single text message with basic health information regarding short- and long-term effects of alcohol consumption that also included a link to a website with information about alcohol.

Outcomes and Follow-up

There were two primary outcomes in the trial, as follows:

- Frequency of heavy episodic drinking (HED), which was assessed by asking participants how many times they consumed 4 (for women), 5 (for men), or more standard drinks on one occasion the past month.
- Total weekly alcohol consumption (TWC), which was measured using a short-term recall method by asking participants the number of standard drinks consumed the past week.

Outcomes were assessed at 2- and 4-month postrandomization, initiated by sending text messages to participants with hyperlinks to questionnaires. Participants were called to collect responses if there was no response to reminders.

Original Sample Size Calculation

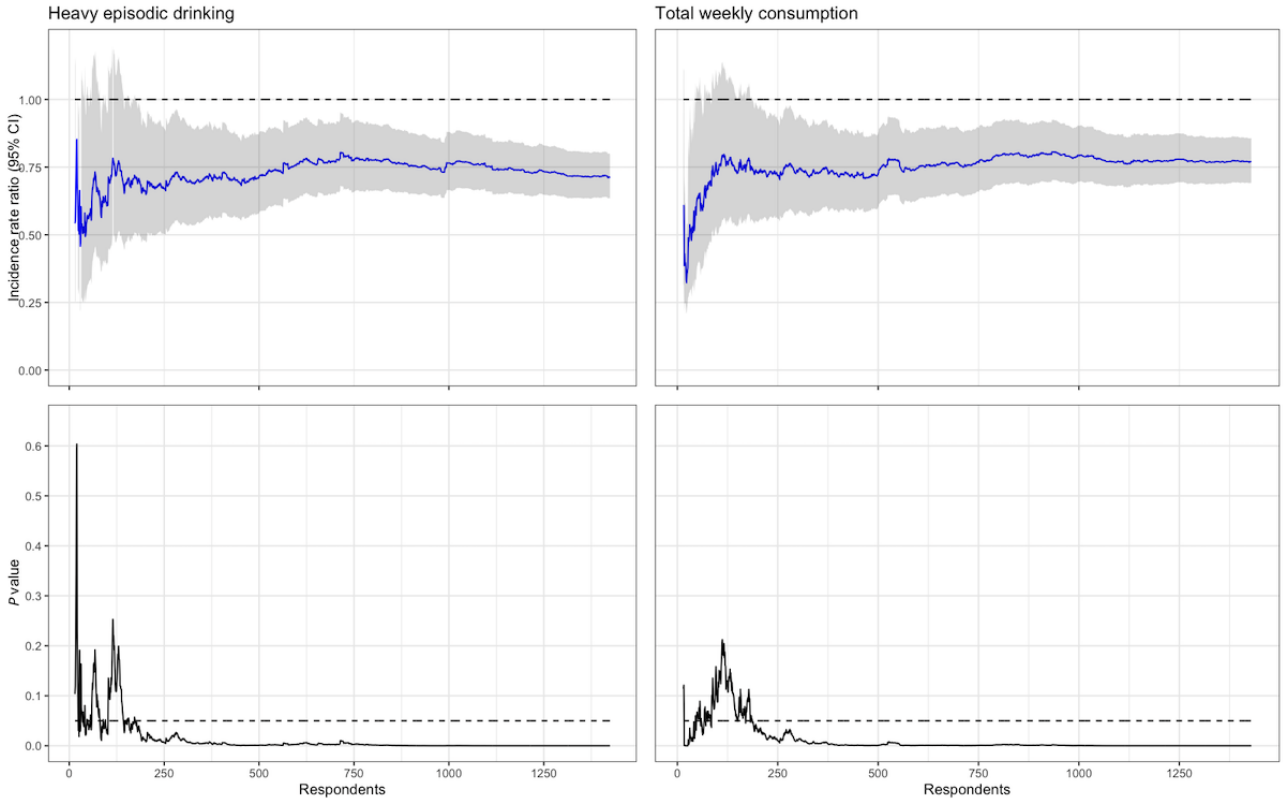
The required sample size was determined using Monte Carlo simulations. A full description of the simulations is available in the study protocol [18]; thus, for succinctness, we restrict the description in this tutorial to the most relevant parts. We believed that a minimal relevant effect for the type of intervention studied, taking into consideration the unguided nature of the intervention and the setting, would be if the intervention group were consuming 15% less alcohol per week at the 4-month follow-up in comparison to the control group. We aimed for an expected power of 80% at the 0.05 significance threshold. Based on our previous studies of digital interventions in Sweden, we expected an attrition rate between 5% and 25%. The simulations suggested an expected sample size of 2126 individuals (interquartile range 2031-2198).

Participants were recruited over a series of 6-month periods. Between each period, we checked if the planned sample size had been achieved. Between April 25, 2019, and November 26, 2020, at which time recruitment was stopped, we randomized 2129 participants. This equated to approximately 19 months of recruitment, having allowed an initial grace period of 1 month for advert placement algorithms to optimize their performance.

Estimates Over Time

Putting aside the required sample size of 2129 participants, what would our null hypothesis-based analyses have looked like if we had stopped the trial after collecting data from only 15 participants? What about after 100 or 200 participants? In Figure 3, two pairs of plots are presented that show our analyses of HED and TWC given a certain number of responders to the 4-month follow-up. Looking at Figure 3, we can see in the plots on the top row the incidence rate ratio (IRR) and 95% CI. The analyses showed an IRR less than 1 (ie, the intervention group was drinking less than the control group) already from the first few responders. In the bottom row, the *P* value can be seen to fluctuate heavily in the beginning, crossing the line of statistical significance on multiple occasions and settling below the .05 line at approximately 200 responders. After 200 responders, the IRR estimates (top row) continue to move around somewhat but staying close to approximately an IRR of 0.75. In our main analyses of the trial [17], which included the full sample size, we concluded that the IRR for TWC was 0.77 (95% compatibility interval 0.69-0.86), and the IRR for HED was 0.71 (95% compatibility interval 0.63-0.79)—findings that were already at hand if we had stopped recruiting after collecting data from approximately 250 participants (ie, 12% of the planned sample size).

Figure 3. Maximum likelihood estimates and *P* values plotted against the number of respondents.



Bayesian Sequential Design

If we had decided to not use a fixed sample size but had rather adopted a Bayesian sequential design, we would have foregone a power calculation and instead defined target criteria for when recruitment should end. These criteria may have been the following:

- Effectiveness: p (IRR < 1 | D) > 97.5% and p (IRR < 0.87 | D) > 50%
- Futility: p (0.87 < IRR < 1.15 | D) > 97.5%

The effectiveness criterion says that we should stop recruitment if the probability that the intervention group is drinking less than the control group is greater than 97.5%; it also says that the probability of the estimated IRR being less than 0.87 should be greater than 50%. An IRR of 0.87 is comparable with our fixed sample size power calculation assumption of 15% less alcohol consumption in the intervention group versus the control group. The futility target criterion says that we will stop recruitment if it is more than 97.5% likely that the estimated IRR is between 0.87 and 1.15, that is, within a range of effect sizes that are considered too small to be of importance considering the context.

Just like we did for the null hypothesis analyses in Figure 3, we can plot the target criteria over time to see what they would look like given a certain number of participants. Since these are Bayesian analyses, we must decide on priors for coefficients before we do inference. In our demonstration, we compare the use of standard normal priors, that is, normal distributions with a mean of 0 (SD 1), with more conservative normal priors, with a mean of 0 (SD 0.1), as in Figure 2B and Figure 2C.

Figure 4 shows, for HED, the evaluated target criteria over number of respondents. In the top left plot, we see the median of the posterior distribution of the IRR using standard normal priors (ie, SD of 1). The analysis shows that the estimated effect was in the direction of the intervention group consuming less alcohol than the control group already early in the trial. In the bottom left plot of Figure 4, the effectiveness criteria are represented by the blue and green lines and the futility criteria by the red line. As it can be seen, after approximately 225 participants, the criteria are fulfilled, and the trial could have ended with evidence of the intervention producing lower HED. However, scrutiny of the bottom left plot shows that the same conclusion could have been made after 175 participants, as the criteria were fulfilled briefly. Generally, we would like to avoid making conclusions with small sample sizes, and the plots show that findings are not stable early on. We can avoid making claims when data are scarce by encoding skepticism using priors. In the two plots on the right in Figure 4, the same analyses are presented but with a normal prior distribution with a SD of 0.1. As it can be seen in the top right plot, effect estimates are strongly pulled toward an IRR=1 at the beginning of the trial, and in the bottom right plot, both effect criteria are below their respective target lines. It is not until after approximately 300 participants that the criteria settle down and show strong evidence that the intervention has a positive effect on alcohol consumption. This *shrinkage* of estimates plays a crucial role in protecting from spurious findings when data are scarce.

In Figure 5, IRR estimates and effect criteria are plotted for TWC. In the left plots, we can see that the effectiveness criteria are above their respective lines (97.5% and 50%) already after 15-20 participants, and then they begin to waver until settling down after approximately 300 participants. The futility criterion

never comes close to crossing the 97.5% line. Although mathematically correct, most researchers should feel uneasy about making claims of effectiveness after only 15-20 participants, and the left-hand side of Figure 4 shows rightly that. Again, skepticism encoded in the prior may help, and as it can be seen in the plots on the right in Figure 5, early claims of effect are protected against. However, scrutiny of the right-hand side of Figure 5 also reveal that there are multiple instances when both effectiveness criteria are fulfilled but later cross below their target lines again, prior to data being available from 300 participants. Even more skeptical priors may protect against these early findings; however, good judgement from researchers when studying the development of their evidence may be more effective.

Figure 4. Posterior probability distributions and target criteria plotted over available data from respondents with respect to total weekly consumption (TWC) using both standard normal priors (left) and skeptical priors (right). IRR: incidence rate ratio.

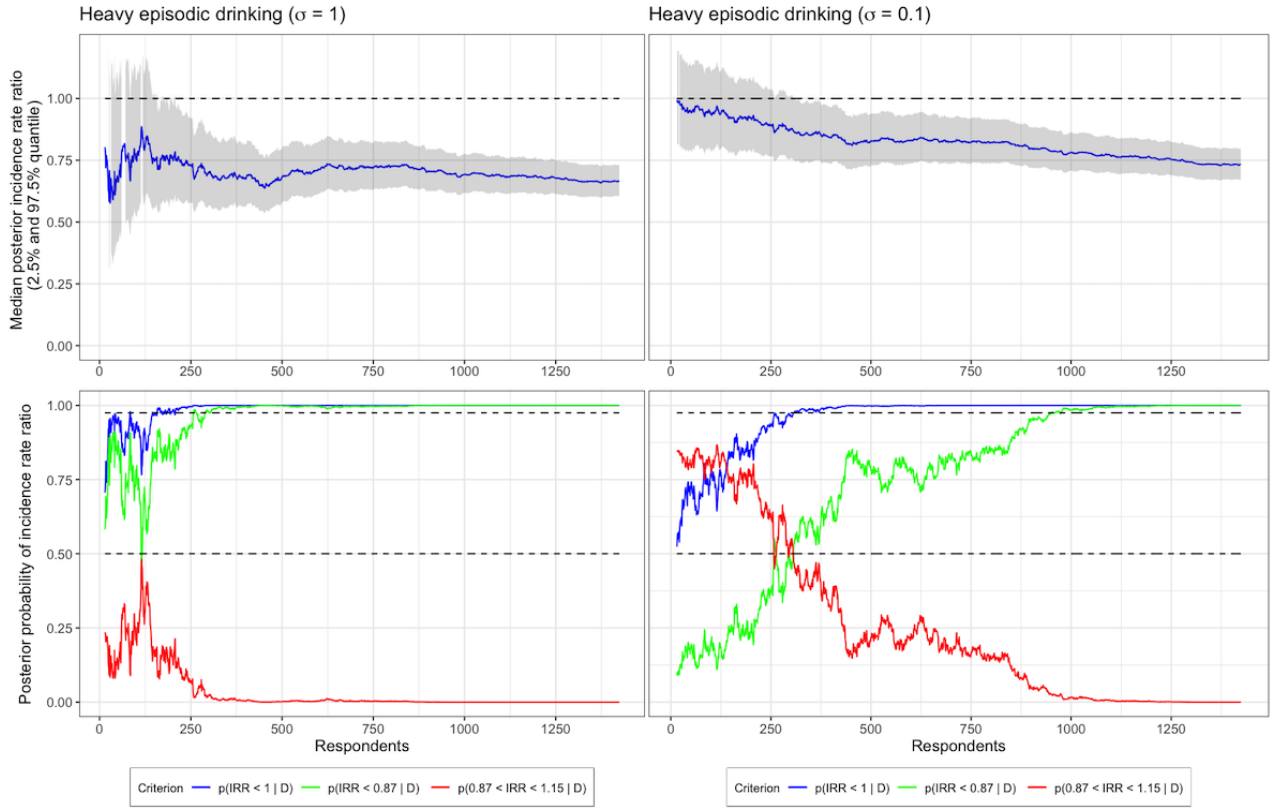
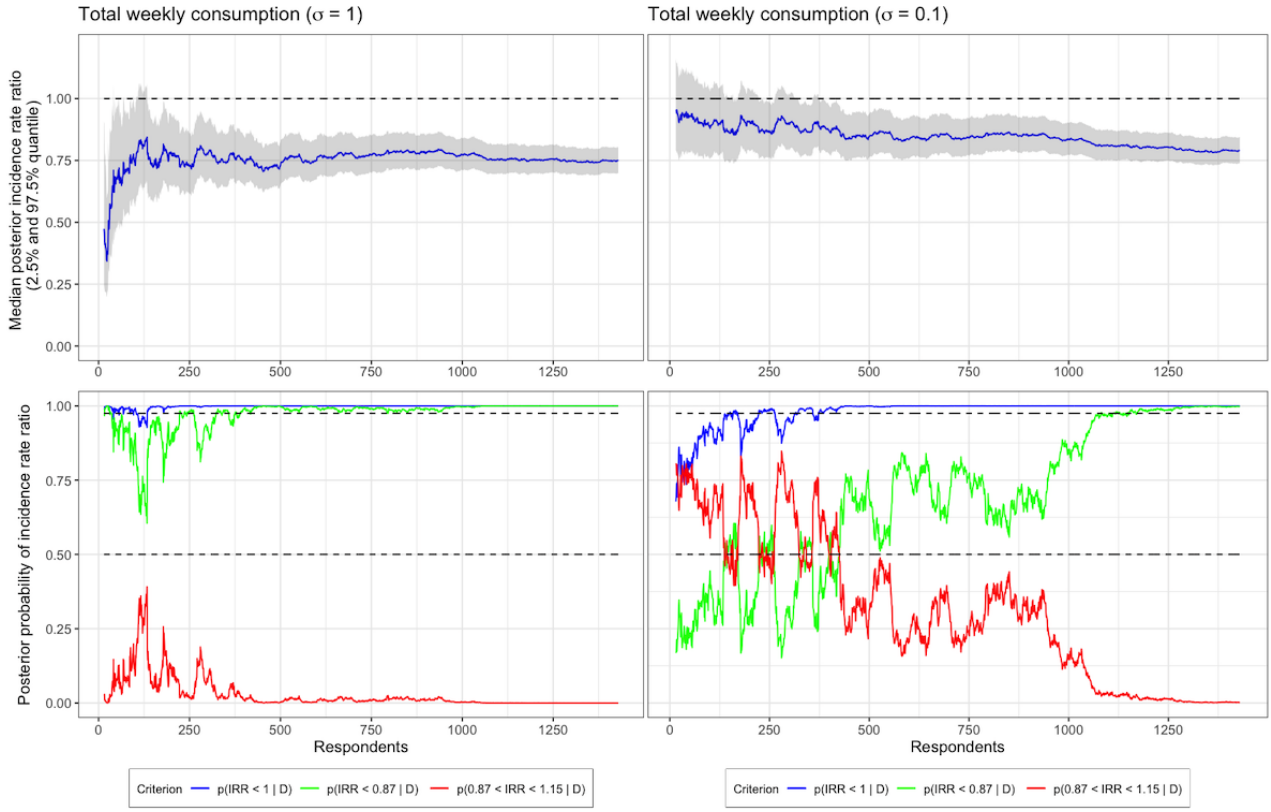


Figure 5. Posterior probability distributions and target criteria plotted over available data from respondents with respect to total weekly consumption (TWC) using both standard normal priors (left) and skeptical priors (right). IRR: incidence rate ratio.



Discussion

A trial of a digital alcohol intervention could have stopped recruitment after approximately 15% of the prespecified sample size had been recruited if a Bayesian sequential design had been used. The consequences would have been fewer participants recruited to a control condition that made them wait for the novel support tool and reduced costs of recruitment; in addition, evidence of the intervention's effectiveness could have been made public sooner. Instead, overrecruitment was the result of anticipating small effects from a public health intervention of this type, while also controlling for the risk of type 1 and 2 errors.

Trials are conducted because effects of interventions are not known; thus, the design of trials should facilitate discovery efficiently. This is not to say that prior knowledge cannot be useful when designing Bayesian sequential designs; on the contrary, both conservative views on the effects and data from previous trials can be incorporated into the priors used during analysis. Priors are ideal in this circumstance since they dominate the analysis when data are scarce, protecting from spurious findings, yet their influence is lessened as more data become available.

Bayesian sequential designs do not rely on an a priori fixed sample size; nevertheless, planning, ethics approval, and grant applications often require one. This can still be achieved by estimating the final sample sizes using simulation [12]. Statistical software can generate synthetic data simulating the planned trial, and analyses can be done using these synthetic data to evaluate the criteria specified in the trial design. By repeating this procedure multiple times, with varying effect sizes, an estimate of how many participants it will require to fulfill the criteria can be produced.

One caveat that should be avoided when using Bayesian sequential designs is to view the target criteria as hard and fast rules—making them shortcuts to going back to dichotomizing evidence into effect and no effect. Instead, the target criteria should be viewed as researchers' intentions for what is considered findings of interest. One may have fulfilled some criteria of the trial but not others and still decide to end the trial. The trial should be stopped when, on the basis of accumulated results, the answer to a scientific question is sufficiently well known that the results can be used in a broader context [12]. The posterior distribution of effect can be estimated and reported, with the probability of a difference between groups indicating the certainty about findings.

In some trials, it will not be possible to access follow-up data continuously throughout the trial period to check the criteria, and so a Bayesian sequential design may not be possible to adopt. This may be the case if data are collected at multiple sites, possibly internationally, and it is time-consuming to collate all data to do analyses. However, it should be noted that the benefits of sequential designs may still be used in cases where it is possible to analyze data at least occasionally, for instance for every 50-100 participants. Analyses do not have to be done for every new data point available but rather for larger sets of participants.

Finally, reducing research waste and protecting research participants from unnecessary harm should be top priorities for researchers studying interventions. To avoid under- and

overrecruitment, which occurs when using fixed sample sizes, is an important mitigation, and Bayesian sequential designs allow for exactly this. Examples of their use in behavioral intervention trials can be found in the literature [19-22], and when appropriate, they should become standard procedure.

Acknowledgments

This project received funding from the Alcohol Research Council of the Swedish Alcohol Retailing Monopoly (DNR 2019-0056 and DNR 2020-0043). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability

Deidentified data sets generated during or analyzed during this study will be made available upon reasonable request to the corresponding author, after approval of a proposal and with a signed data access agreement.

Conflicts of Interest

MB owns a private company (Alexit AB) that maintains and distributes evidence-based lifestyle interventions to be used by the public and in health care settings. Alexit AB played no role in developing the intervention, study design, data analysis, data interpretation, or writing of this report. Services developed and maintained by Alexit AB were used for sending text messages and data collection.

References

1. Bendtsen M. A Gentle Introduction to the Comparison Between Null Hypothesis Testing and Bayesian Analysis: Reanalysis of Two Randomized Controlled Trials. J Med Internet Res 2018 Oct 24;20(10):e10873 [FREE Full text] [doi: 10.2196/10873] [Medline: 30148453]
2. Bland JM. The tyranny of power: is there a better way to calculate sample size? BMJ 2009 Oct 06;339(oct06 3):b3985-b3985. [doi: 10.1136/bmj.b3985] [Medline: 19808754]
3. Berry DA. Bayesian Statistics and the Efficiency and Ethics of Clinical Trials. Statist Sci 2004 Feb 1;19(1):175-187 [FREE Full text] [doi: 10.1214/088342304000000044]
4. Wasserstein RL, Lazar NA. The ASA Statement on P-Values: Context, Process, and Purpose. The American Statistician 2016 Jun 09;70(2):129-133. [doi: 10.1080/00031305.2016.1154108]
5. Amrhein V, Greenland S. Remove, rather than redefine, statistical significance. Nat Hum Behav 2018 Jan 25;2(1):4-4. [doi: 10.1038/s41562-017-0224-0] [Medline: 30980046]
6. McCambridge J, Kypri K, Elbourne D. Research participation effects: a skeleton in the methodological cupboard. J Clin Epidemiol 2014 Aug;67(8):845-849 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.03.002] [Medline: 24766858]
7. Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. BMC Med Res Methodol 2013 Dec 06;13(1):150 [FREE Full text] [doi: 10.1186/1471-2288-13-150] [Medline: 24314204]
8. Müssener U, Linderoth C, Bendtsen M. Exploring the Experiences of Individuals Allocated to a Control Setting: Findings From a Mobile Health Smoking Cessation Trial. JMIR Hum Factors 2019 Apr 02;6(2):e12139 [FREE Full text] [doi: 10.2196/12139] [Medline: 30938682]
9. Bendtsen M, Ulfsdotter Gunnarsson K, McCambridge J. Effects of a waiting list control design on alcohol consumption among online help-seekers: protocol for a randomised controlled trial. BMJ Open 2021 Aug 26;11(8):e049810 [FREE Full text] [doi: 10.1136/bmjopen-2021-049810] [Medline: 34446493]
10. Bendtsen M. The P value line dance: when does the music stop? J Med Internet Res 2020 Aug 27;22(8):e21345 [FREE Full text] [doi: 10.2196/21345] [Medline: 32852275]
11. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006 Jan;5(1):27-36. [doi: 10.1038/nrd1927] [Medline: 16485344]
12. Lee JJ, Chu CT. Bayesian clinical trials in action. Stat Med 2012 Nov 10;31(25):2955-2972 [FREE Full text] [doi: 10.1002/sim.5404] [Medline: 22711340]
13. Gsponer T, Gerber F, Bornkamp B, Ohlssen D, Vandemeulebroecke M, Schmidli H. A practical guide to Bayesian group sequential designs. Pharm Stat 2014 Aug 24;13(1):71-80. [doi: 10.1002/pst.1593] [Medline: 24038922]
14. Müssener U, Linderoth C, Thomas K, Bendtsen M. mHealth smoking cessation intervention among high school students: 3-month primary outcome findings from a randomized controlled trial. PLoS One 2020 Mar 6;15(3):e0229411 [FREE Full text] [doi: 10.1371/journal.pone.0229411] [Medline: 32142514]
15. Bendtsen M, Bendtsen P, Müssener U. Six-month outcomes from the NEXit junior trial of a text messaging smoking cessation intervention for high school students: randomized controlled trial with Bayesian analysis. JMIR Mhealth Uhealth 2021 Oct 21;9(10):e29913 [FREE Full text] [doi: 10.2196/29913] [Medline: 34673532]

16. Thomas K, Bendtsen M, Linderoth C, Müssener U. mHealth smoking cessation intervention among high-school pupils (NEXit Junior): study protocol for a randomized controlled trial. Trials 2018 Nov 22;19(1):648 [FREE Full text] [doi: 10.1186/s13063-018-3028-2] [Medline: 30466484]
17. Bendtsen M, Åsberg K, McCambridge J. Effectiveness of a digital intervention versus alcohol information for online help-seekers in Sweden: a randomised controlled trial. BMC Med 2022 May 17;20(1):176 [FREE Full text] [doi: 10.1186/s12916-022-02374-5] [Medline: 35578276]
18. Bendtsen M, McCambridge J. Reducing Alcohol Consumption Among Risky Drinkers in the General Population of Sweden Using an Interactive Mobile Health Intervention: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Apr 18;8(4):e13119. [doi: 10.2196/13119]
19. Bendtsen M, Thomas K, Linderoth C, Bendtsen P. Effects of a text messaging smoking cessation intervention among online help seekers and primary health care visitors in Sweden: protocol for a randomized controlled trial using a Bayesian group sequential design. JMIR Res Protoc 2020 Dec 03;9(12):e23677 [FREE Full text] [doi: 10.2196/23677] [Medline: 33269703]
20. Åsberg K, Lundgren O, Henriksson H, Henriksson P, Bendtsen P, Löf M, et al. Multiple lifestyle behaviour mHealth intervention targeting Swedish college and university students: protocol for the randomised factorial trial. BMJ Open 2021 Dec 30;11(12):e051044. [doi: 10.1136/bmjopen-2021-051044]
21. Åsberg K, Blomqvist J, Lundgren O, Henriksson H, Henriksson P, Bendtsen P, et al. Digital multiple health behaviour change intervention targeting online help seekers: protocol for the COACH randomised factorial trial. BMJ Open 2022 Jul 26;12(7):e061024 [FREE Full text] [doi: 10.1136/bmjopen-2022-061024] [Medline: 35882466]
22. Bendtsen M, Garnett C, Toner P, Shorter GW. The effect of question order on outcomes in the core outcome set for brief alcohol interventions among online help-seekers: protocol for a factorial randomized trial. JMIR Res Protoc 2020 Nov 26;9(11):e24175 [FREE Full text] [doi: 10.2196/24175] [Medline: 33242024]

Abbreviations

HED: heavy episodic drinking
IRR: incidence rate ratio
OR: odds ratio
TWC: total weekly consumption

Edited by T Leung, G Eysenbach; submitted 03.07.22; peer-reviewed by I Wilson, E Andrikopoulou; comments to author 29.10.22; revised version received 02.11.22; accepted 30.11.22; published 16.12.22.

Please cite as:
Bendtsen M
Avoiding Under- and Overrecruitment in Behavioral Intervention Trials Using Bayesian Sequential Designs: Tutorial
J Med Internet Res 2022;24(12):e40730
URL: https://www.jmir.org/2022/12/e40730
doi:10.2196/40730
PMID:36525297

Original Paper

The Web-Based Advance Care Planning Program "Explore Your Preferences for Treatment and Care": Development, Pilot Study, and Before-and-After Evaluation

Doris van der Smissen[1], MSc; Judith A C Rietjens[1], PhD; Sandra van Dulmen[2,3,4], PhD; Ton Drenthen[5], PhD; F Ragnhild M D Vrijaldenhoven-Haitsma[1], MD; Marijke Wulp[6], MSc, MA; Agnes van der Heide[1], MD, PhD; Ida J Korfage[1], PhD

[1]Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands

[2]Nivel (Netherlands institute for health services research), Utrecht, Netherlands

[3]Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, Netherlands

[4]Faculty of Caring Science, Work Life and Social Welfare, University of Borås, Borås, Sweden

[5]Dutch College of General Practitioners, Utrecht, Netherlands

[6]Agora, Bunnik, Netherlands

Corresponding Author:
Doris van der Smissen, MSc
Department of Public Health
Erasmus MC, University Medical Center Rotterdam
Doctor Molewaterplein 40
Rotterdam, 3015 GD
Netherlands
Phone: 31 107038470
Email: d.vandersmissen@erasmusmc.nl

Abstract

Background: Web-based advance care planning (ACP) programs may support patients in thinking about and discussing their preferences for future treatment and care. However, they are not widely available, and only a limited number of programs are evidence based.

Objective: We aimed to develop and evaluate an evidence-based, interactive web-based ACP program that guides users through the process of thinking about, discussing, and recording of preferences for treatment and care.

Methods: The program "Explore your preferences for treatment and care" was developed, pilot-tested on feasibility, and subsequently evaluated; engagement in ACP was assessed before program completion and 2 months after program completion using the ACP Engagement Survey (score 1-5) among 147 persons with chronic disease. Usability (score 0-100) and user satisfaction (score 1-5) were also assessed.

Results: ACP engagement increased from 2.8 before program completion to 3.0 two months after program completion (P<.001); contemplation about ACP increased from 2.6 to 2.8 (P=.003), and readiness for ACP increased from 2.2 to 2.5 (P<.001). No changes were found for knowledge about ACP (3.0-3.2; P=.07) and self-efficacy for ACP (3.8-3.8; P=.25). The program was perceived as usable (mean 70, SD 13), attractive (mean 3.8, SD 0.7), and comprehensible (mean 4.2, SD 0.6).

Conclusions: We developed an evidence-based, interactive web-based ACP program in cocreation with patients, relatives, and health care professionals. Before-and-after evaluation showed that the program can support people in taking first steps in ACP and in reflecting on preferences for treatment and care, by guiding them through the process of ACP using a stepwise approach. Participants perceived the program as usable and understandable, and they were satisfied with the program and with the amount of information. Health care professionals may use the program as a tool to start ACP discussions with their patients. The program may increase awareness of ACP.

(J Med Internet Res 2022;24(12):e38561) doi:10.2196/38561

KEYWORDS

advance care planning; internet-based intervention; decision aids; patient education; eHealth; health communication; patient-centered care; chronic disease

Introduction

Advance Care Planning

Reflecting on future treatment and care can be relevant for everyone, but especially for older persons and people with chronic diseases [1]. Advance care planning (ACP) enables patients to define their goals and preferences for future medical treatment and care, to discuss these with relatives and health care professionals, and to record these in a document such as an advance directive [1]. ACP facilitates decision-making by patients, relatives, and health care professionals [1,2]. In case patients' health condition worsens and they are unable to express their preferences themselves, it is important that their preferences are known by relatives and health care professionals to facilitate care in concordance with patients' values, goals, and preferences for treatment and care [1,2].

ACP is typically conducted through face-to-face conversations between patients, their relatives, and health care professionals [2]. However, ACP conversations do not take place as often as patients, their relatives, and health care professionals would prefer [3,4]; patients expect health care professionals to initiate such conversations, whereas health care professionals are hesitant to do so and lack time and training [4,5].

Web-Based ACP Programs

Interactive web-based programs may support patients in the first steps of ACP, for instance, in communicating their preferences and in recording these preferences [6]. Web-based programs can be accessed at any preferred time and place and therefore a larger audience can access ACP (information) [6]. Furthermore, as web-based programs can be tailored and delivered in an interactive and stepwise format, this may complement ACP processes as facilitated by health care professionals, for example, by supporting people in preparing discussions with their health care professionals, and consider their preferences for treatment and care, goals, and values [6].

Examples of web-based ACP programs that have been shown to support people in ACP include the *Prepare For Your Care* program [7] and the *Making Your Wishes Known* program [8]. A scoping review showed that web-based ACP programs have the potential to support people in ACP [6]. Furthermore, the scoping review showed that most evidence-based, interactive web-based ACP programs have been developed in the United States and only a few have been thoroughly evaluated [6]. Evaluation of the web-based ACP programs (including their feasibility, usability, and user satisfaction) is important to ensure the program is reliable and feasible to users. The results of such an evaluation can be used to improve web-based ACP programs and may be incorporated in other ACP interventions.

The aim of this study was to develop and evaluate an evidence-based, interactive web-based program for ACP in the Netherlands.

Methods

Developmental Process of the Web-Based ACP Program

The developmental process of the web-based ACP program was based on the main principles of the Medical Research Council guidance for developing and evaluating complex interventions, which consists of four phases: development or identification of the intervention, feasibility, evaluation, and implementation; while in each phase considering context, developing and refining program theory, engaging stakeholders, identifying key uncertainties, refining the intervention, and economic considerations [9]. From the start of the study, we aimed to include several stakeholders in all phases of the study (patients, health care professionals, and patient organizations) and for the program to be inclusive, including persons with low health literacy. First, we gained insight into the needs, preferences, and values of the stakeholders. On the basis of the needs, preferences, and values, we created a prototype of the web-based program in collaboration with the stakeholders. Next, we evaluated (feasibility and effect evaluation) and implemented the program.

The objectives of the program were to develop a web-based program that (1) informs about ACP and its possibilities and impossibilities; (2) invites patients to think about preferences and goals for future treatment and care; (3) invites patients and relatives to share preferences and goals for future treatment and care with each other and with health care professionals; (4) invites patients to record preferences and goals for treatment and care; and (5) invites patients to appoint a health care representative. Users could choose the steps of ACP they are ready to engage in and are not required to complete the entire program.

Input for the program's content and structure came from several sources (Table 1):

1. In a scoping review, we examined the content, feasibility, and effectiveness of evidence-based, interactive web-based ACP programs [6]. We identified effective ACP elements, such as the exploration of values and goals, communication with relatives and health care professionals, and the recording of ACP [6]. Furthermore, we identified important functionalities of web-based programs such as the use of videos, the option to have the program content read aloud, and the option to print a document. Finally, the scoping review helped to select appropriate outcome measures for the evaluation of the web-based ACP program, for instance, engagement in ACP, the program's usability, and the users' level of satisfaction with the program [6].
2. In an interview study, we identified information needs of patients with chronic diseases and their relatives regarding web-based ACP, such as the need for information about ACP and its relevance, the need for reliable information about their disease and (arranging) care, and the need for

peer support, as well as search terms for finding ACP related information [10].

3. A stakeholder group was formed to include the perspectives of different stakeholders during the development, evaluation, and implementation of the web-based ACP program. The stakeholder group included 1 patient (co-author FRMDVH), 2 relatives, and representatives of the Dutch College of General Practitioners (Nederlands Huisartsen Genootschap), the Dutch Association for Kidney Patients (Nierpatiënten Vereniging Nederland), the Dutch Association for Patients (de Nederlandse Patiëntenvereniging), and Agora (organization to promote the palliative approach). Furthermore, the stakeholder group included 1 expert in health communication of the Nivel (Netherlands institute for health services research), 1 expert in eHealth of the University of Twente, 1 representative of Vital Innovators, an organization that conducts Social Return of Investment analyses, and the researchers (DvdS, IJK, JACR, and AvdH) with expertise in shared decision-making, care at the end of life, and eHealth. During the entire 4-year project, the stakeholder group met approximately 2 to 3 times per year for 2-hour meetings. During these meetings, the progress of the project, preliminary results, and planned next steps were discussed. The stakeholders provided their feedback, which was processed by the researchers. The members of the group also assisted in the implementation of the program, for example, by disseminating the program, and participating in media interviews and in seminars about the program. The program was embedded in the general practitioners' platform *Thuisarts.nl* [11] (English version: GPinfo.nl [12]). Representatives of Thuisarts.nl participated in the stakeholder group. Thuisarts.nl provides health-related information for patients and had 1.6 million unique visitors per month in 2016 [13]. It is visited by patients, and 90% of general practitioners reported to at least sometimes use Thuisarts.nl during consultations [13,14].

Table 1. Main content and characteristics of the web-based program and the studies these were based on.

Main elements of the web-based ACP[a] program	Study findings
Content	
Information about ACP, thinking about values and quality of life, communication about preferences with relatives and health care professionals, appointing a health care representative, recording of preferences in an advance directive, reviewing the advance directive.	• Consensus definition of ACP (including ACP elements) [1,15] • Scoping review [6] • Interview study [10] • Stakeholder group meetings
References to other information pages and websites with information about disease, patient organizations, and peer support.	• Interview study [10]
Structure	
Interactive program; people can answer questions, watch videos, and click on additional information.	• Scoping review [6] • Stakeholder group meetings
Option to save and print one's responses to the questions in the program.	• Scoping review [6] • Stakeholder group meetings
Functionalities such as hyperlinks to external websites, videos, and text-to-speech option.	• Scoping review [6] • Stakeholder group meetings
Clear and simple structure, range of topics not too broad, text not too long (taking people with lower health literacy or computer skills into account).	• Stakeholder group meetings
Embedment in well-known and reliable general practitioners' information platform (Thuisarts.nl [11]), possibilities to link to additional health information.	• Stakeholder group meetings
Inclusion of search terms as indicated by users (eg, "What is ACP?" "Recording of preferences").	• Interview study [10] • Stakeholder group meetings

[a]ACP: advance care planning.

Evaluation of the Web-Based ACP Program

Pilot Study

The program's feasibility was evaluated in a pilot study. On the basis of the definition of feasibility of Bowen et al [6,16], we explored how users perceived the acceptability of the program, usability, and understandability of the text. A total of 6 patients with chronic diseases (aged 28-73 years) were included, including multiple sclerosis, cancer and kidney disease, and 3 physicians (1 male and 2 females, aged 47-66 years), 2 general practitioners and 1 surgeon with ACP experience. The three-step test interview method was used [17]. In step 1, we observed how the interviewees completed the program, while they expressed their thoughts out loud (concurrent think aloud) [17]. Step 2 was aimed at clarifying the expressions observed during step 1 [17]. In step 3, interviewees were asked about their experiences and opinions about the program [17]. The interviews were conducted at the participant's home or the study center

and lasted approximately 1 hour. The interviews were video recorded. The researcher watched the videos and made notes of important feedback, verbalizations, or actions by the interviewees.

The pilot study and evaluation study were approved by the Medical Research Ethics Committee of the Erasmus MC, University Medical Center Rotterdam on October 21, 2019 (MEC-2019-0590). Participants provided written informed consent. The data for the pilot study were collected in February 2020, and the data for the evaluation study were collected from April 2020 to June 2020.

Evaluation Study: Before-and-After Evaluation

Study Population and Study Design

Participants were recruited via an internet-based Dutch research portal [18]. In this portal, people can sign up to participate in research; they can collect points per completed survey, which they can exchange for a gift card. Inclusion criteria were as follows:

1. Having a chronic disease, defined as a disease that lasts 3 months or longer, is not (completely) curable, and which reoccurs regularly. Examples are chronic obstructive pulmonary disease, multiple sclerosis, and cancer. Participants with a psychological disorder or dementia were not invited to the study.
2. Participants aged ≥18 years were included.

We used purposive sampling by inviting comparable numbers of men and women, with diverse educational backgrounds, living across the Netherlands.

Members of the research portal were asked to (1) complete the baseline survey on ACP engagement, health literacy, and demographics; (2) complete the web-based ACP program and a survey on usability and user satisfaction; and (3) complete a survey on ACP engagement after 2 months. Reminders were sent if participants had not completed the measurement within 1 to 2 weeks. All participants had completed the baseline survey before the launch of the web-based ACP program.

Participant Characteristics

We assessed participants' age, level of education, country of birth [19], and the type of chronic disease. We assessed participants' level of health literacy by using the Dutch version of the Set of Brief Screening Questions on a 5-point Likert scale (1=not at all or never; 5=completely or always) [20,21]. We also asked the time it took to complete the web-based ACP program and whether they completed the program together with someone else.

ACP Engagement

ACP has evolved from focusing on completing advance directives to an ongoing behavior change process of considering, discussing, and recording goals, values, and preferences for treatment and care [1,15]. The goal of the web-based ACP program is to make users aware of ACP and to provide guidance in the first steps of ACP, such as thinking about preferences and how to discuss and record these preferences. We hypothesized that the program will influence attitudes toward initiating ACP and involvement in ACP. To assess the participants' behavior change and involvement in ACP, we considered the ACP Engagement Survey to be a suitable instrument as it is about the entire behavior change process of ACP, considering that ACP is an ongoing process [22-24]. As the web-based ACP program is also aimed at informing users about ACP, it could be useful for people who are not yet familiar with ACP, including people who may not be ready for ACP. Research has shown that people need to feel some readiness to start engaging in ACP; however, the ACP process itself can have a positive influence on people's readiness [25]. The participants in the study were recruited via a web-based research portal so they may not have had any prior knowledge of ACP; therefore, we expected a change in ACP engagement comparing baseline with a measurement 2 months after completion of the program.

Participants completed the validated Dutch ACP Engagement Survey (34 items) [22-24] before and 2 months after program completion. This survey focused on four ACP domains: (1) surrogate decision makers; (2) values and quality of life; (3) flexibility in surrogate decision-making; and (4) asking doctors questions [24]. ACP behavior change is measured by four subscales, namely knowledge about ACP (2 questions), contemplation about ACP (3 questions), self-efficacy for ACP (12 questions), and readiness for ACP (17 questions) [22-24]. The response options of knowledge, contemplation, and self-efficacy have a 3-point scale in the Dutch survey version, coded as 1=1, 2=3, and 3=5 [22]. The readiness subscale has a 5-point scale, ranging from 1=I have never thought about it to 5=I have already done it; the fifth answer option can be used to analyze specific ACP behaviors [22-24]. The total ACP engagement score is the mean score of all questions in the survey.

Usability and User Satisfaction of the Web-Based ACP Program

We assessed the program's usability with the System Usability Scale (SUS; 5-point scale: 1=completely disagree to 5=completely agree) [26,27]. A total SUS score was computed using the scoring formula (score of 0-100) [26,27]. We assessed users' satisfaction with the attractiveness and clarity of the program (4 questions), its comprehensibility (3 questions), and emotional support (2 questions) [28,29]. We also asked (1) whether participants would recommend the program to others (1=completely disagree; 5=completely agree); (2) how satisfied they were with the program (1=not at all satisfied; 10=very satisfied) with the possibility to add an explanation; and (3) their view about the amount of information the program provided (1=too little, 5=exactly enough, and 10=too much) with the possibility to add an explanation.

Statistical Analysis

We statistically analyzed participants' responses on the ACP Engagement Survey using the software IBM SPSS Statistics. As the data were not entirely normally distributed, we conducted nonparametric testing with Wilcoxon Signed Rank statistical tests to compare participants' responses on the ACP Engagement Survey before and 2 months after program completion. To see whether the readiness items indicated a behavior change since

baseline, for instance, whether participants moved from the precontemplation behavior change stage to contemplation, action, or maintenance (from 1 to 2 points on the Likert scale at baseline to 3, 4, or 5 points after 2 months), we conducted McNemar tests [23]. To assess whether the change in scores was clinically meaningful, we applied the effect sizes as determined in the validation study of the original ACP Engagement Survey of Shi et al [30]. According to Shi et al [30], mean change scores of approximately 0.2 to 0.3 points are considered to indicate small effect sizes (0.20-0.49), 0.4 to 0.5 points are considered to indicate moderate effect sizes (0.50-0.79), and changes of ≥0.6 points are considered to indicate large effect sizes (≥0.80). To assess the association of level of education with ACP engagement, we performed a subgroup analysis by 2 mixed ANOVAs with a post hoc Bonferroni test. To check for selection bias, we compared age, level of education, and outcomes on ACP engagement of participants who completed all measurements with those who only completed the baseline. A power calculation indicated that we needed a sample size of 70 participants.

Ethics Approval and Patient Consent

This study was approved by the Medical Research Ethics Committee of the Erasmus MC, University Medical Center Rotterdam on October 21, 2019 (MEC-2019-0590). All methods were performed in accordance with the relevant guidelines and regulations (Declaration of Helsinki). The study conforms with the International Committee of Medical Journal Editors (ICMJE) recommendations for the conduct, reporting, editing, and publication and for the protection of research participants. Participants were recruited via an internet-based Dutch research portal [18] and provided written informed consent. The authors confirm that all patient or personal identifiers have been removed or disguised so the persons described are not identifiable and cannot be identified through the details of the story.

Results

The Web-Based ACP Program

The web-based ACP program *Explore your preferences for treatment and care* [31] consists of three steps, guiding users through the following processes:

1. Thinking about preferences for future medical treatment and care
2. Discussing preferences for treatment and care with relatives and health care professionals and appointing a personal representative
3. Recording preferences for treatment and care; instructions are provided on how to record preferences in an advance directive and to review preferences (it is not possible to create an advance directive in the program itself)

Users can choose the steps of ACP they are ready for to engage in and are not required to complete the entire program. The program contains videos, questions, and links to information on health and disease. The user can print or save a document in PDF with the responses to the questions in the program. Textbox 1 shows the content of the program; Multimedia Appendix 1 presents screenshots of the program.

Textbox 1. Content and functionalities of the web-based advance care planning (ACP) program "Explore your preferences for treatment and care" [31].

Content of the program

Main topic (every bullet point is described on a separate page):

Home page

- About the program
- About ACP
- For whom and when
- Useful websites
- Disclaimer

Step 1: Thinking about your treatment and care preferences

- Information about what is important in life and thinking about preferences (with video)
- Question: what is important to you?
- Question: what does this mean for your treatment and care preferences?
- Question: which care would you like to receive or not?
- Your preferences when being severely ill and when you will not recover anymore
- Question: what have you learned from previous experiences?
    - Statements:
        - I want to live as long as possible, even when my quality of life is not good.
        - I want to try various treatments, but I want to stop when my quality of life is no longer good.
        - I want to live as comfortable and free from pain as possible, even if this would mean my life would be shorter.

Step 2: Discussing your treatment and care preferences

- The health care representative (with video)
- Question: have you already thought about a health care representative?
- Question: who would you choose as your health care representative?
- The role of your health care representative
- Question: are there additional things you would want your health care representative to address?
- Discussing your preferences with your health care representative
- Question: what do you need to start the conversation with your health care representative?
- Discussing your preferences with your doctor
- Question: what do you want to discuss with your doctor?

Step 3: Recording your treatment and care preferences

- How to make an advance directive (with video)
- Question: have you already recorded your preferences in an advance directive?
- Question: what topics would you want to record in an advance directive?
- Discussing your advance directive and sharing it
- Question: when would you review your advance directive?

End of program

- Your answers as given while completing this program (user sees answers and can save these in PDF or print these)

Functionalities of the program

- Users can generate a document with the questions and their answers, this document can be printed and saved in PDF
- Users can navigate in the program: they can skip steps or can go back and forward in the program

- Users can track their progress in the program
- Users can answer open and closed questions, and statements
- The program is interactive: users can answer questions, click on links for additional information
- Users can watch videos with patient experiences with ACP
- The program refers to useful information about disease, treatment and care, information for relatives, patient associations, and peer support, partly within the website "Thuisarts.nl."
- Audio can be used, text-to-speech option: text can be read out loud
- The program can be accessed by phone, computer, and tablet
- Clear and understandable language
- Clear structure and layout

Evaluation of the Web-Based ACP Program

Pilot Study

In the pilot study, patients mentioned that the program made them think about their treatment and care preferences and they understood the questions in the program well. The participating physicians thought the program would be valuable for patients to support them in ACP. All interviewees were able to complete the program without problems. Some minor suggestions were given to the web design team. For example, sometimes interviewees clicked on a hyperlink to an external website without noticing they left the program website. Subsequently, the web design team inserted a notification, and they also applied small language improvements based on the interviewees' feedback. All participants thought the program was interesting. Some interviewees mentioned that they would recommend the program to others.

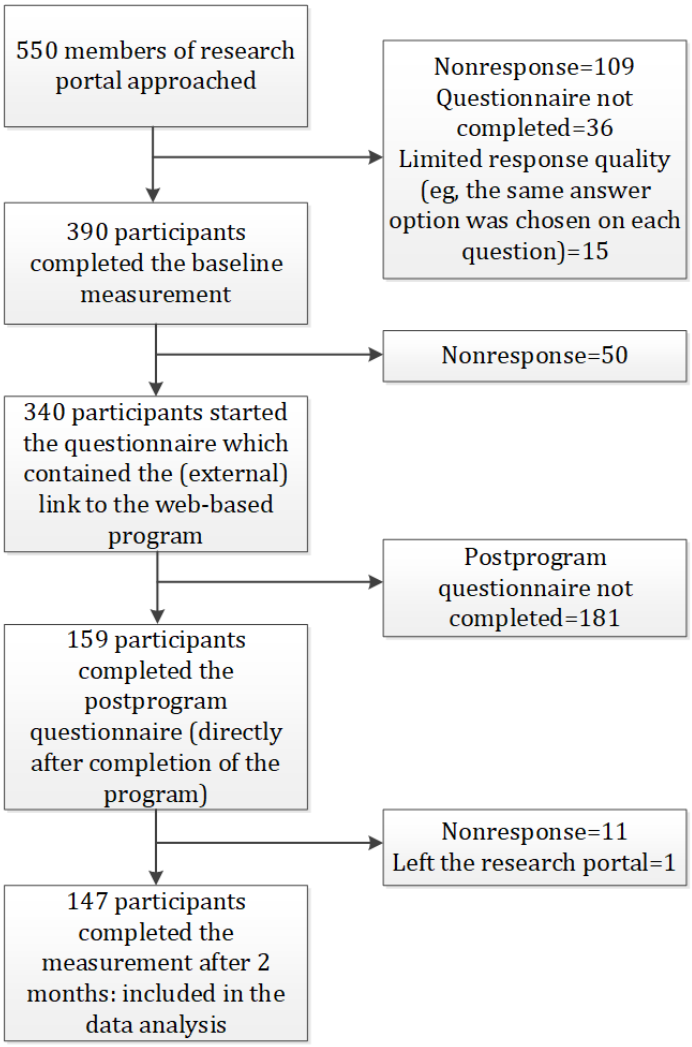
Evaluation Study: Before-and-After Evaluation

Participant Characteristics

The baseline questionnaire was sent to 550 members of the research portal with chronic disease. The baseline measurement was completed by 70.9% (390/550) of the participants, the second measurement (program completion and additional questionnaire) was completed by 40.8% (159/390) of the participants, and the measurement after 2 months was completed by 92.5% (147/159) of the participants (Figure 1).

Participants who completed all questionnaires (n=147) were included in this study. They were 60.5 years of age on average (SD 10.7, range 26-82 years), and of the 147 participants, 82 (55.8%) were male and 65 (44.2%) were female; this was largely representative for the general Dutch population of people aged ≥18 years (with 49% males and 51% females in 2019) [32]. Of the 147 participants, 143 (97.3%) were born in the Netherlands. Levels of education differed: low, 36 (24.5%) participants; medium, 61 (41.5%) participants; and high, 50 (34%) participants; this was largely representative for the general Dutch population of people aged ≥18 years (with 28% low, 43% middle, and 29% highly educated in 2019) [32]. Levels of health literacy were high on average (mean 4.5, SD 0.5; scale 1-5). Participants completed the program within 26.1 minutes on average (SD 22.2 minutes). Two participants completed the program together with a family member or partner.

Figure 1. A flowchart of participants' responses.



ACP Engagement

The total ACP engagement score increased from 2.8 before program completion to 3.0 after 2 months (P<.001). Contemplation about ACP increased from 2.6 to 2.8 (P=.003). Readiness increased from 2.2 to 2.5 (P<.001). No changes were found for knowledge about ACP (3.0-3.2; P=.07) and for self-efficacy for ACP (3.8-3.8; P=.25). Comparing baseline and the measurement after 2 months, we found significant increases for three of four domains: (1) surrogate decision makers increased from 2.8 to 3.0 (P<.001); (2) what matters most in life: health situations increased from 3.0 to 3.2 (P=.003) and care at the end of life increased from 2.9 to 3.0 (P=.02); and (3) flexibility in medical decision-making increased from 2.6 to 2.9 (P<.001). We found no significant differences for domain 4, ask your doctors questions (3.1-3.2; P=.20). Table 2 presents the results. According to the validation study of the original ACP Engagement Survey, the changes in scores indicate clinically meaningful changes [30].

The McNemar test showed no significant difference in participants' stages of behavior change after 2 months compared with baseline (P=.11), except for the domain "flexibility in medical decision-making"; participants moved from precontemplation to higher stages of behavior change (contemplation, action, or maintenance); this difference was significant (P=.04).

The mixed ANOVAs with a post hoc Bonferroni test indicated no significant differences on level of education (low, middle, and high) for the subscales (knowledge: P=.06; contemplation: P=.51; self-efficacy: P=.90; readiness: P=.19; and total ACP engagement score: P=.56) and neither for the domains (surrogate decision makers: P=.49; what matters most in life: health situations: P=.41; and care at the end of life: P=.55; flexibility in medical decision-making: P=.39; and ask your doctors questions: P=.82).

The incomplete responses or dropout in the measurement in which the program needed to be completed was 53.2% (181/340) of participants; they started the questionnaire, which contained the (external) link to the web-based ACP program, but did not complete the postprogram questionnaire. As user data are not recorded in the program, we were not able to see whether these participants completed the program. When comparing the 147 participants who completed all measurements with the 390 participants who completed only the baseline, we found no significant differences among the groups for age (P=.19), education level (P=.29), and levels of ACP engagement (the subscales: knowledge, P=.92; contemplation, P=.34; self-efficacy, P=.40; readiness, P=.61; and the total ACP

engagement score, *P*=.81) and the domains: surrogate decision makers, *P*=.98; what matters most in life: health situations, *P*=.74; and care at the end of life, *P*=.99; flexibility in medical decision-making, *P*=.43; and ask your doctors questions, *P*=.33), suggesting there was no selection bias owing to the dropout.

Table 2. Results of the ACP[a] Engagement Survey (34 items) per subscale and domain (N=147).

Item	Before the ACP program, mean (SD)	2 months after the ACP program, mean (SD)	*P* value
Subscale[b]			
Knowledge about ACP	3.0 (1.4)	3.2 (1.2)	.07
Contemplation about ACP	2.6 (1.1)	2.8 (1.2)	.003
Self-efficacy for ACP	3.8 (0.8)	3.8 (0.9)	.25
Readiness for ACP	2.2 (1.0)	2.5 (1.1)	<.001
Domain[b]			
Surrogate decision makers	2.8 (1.0)	3.0 (1.0)	<.001
What matters most in life			
Health situations	3.0 (0.9)	3.2 (0.9)	.003
Medical care at the end of life	2.9 (0.9)	3.0 (1.0)	.02
Flexibility in medical decision-making	2.6 (1.0)	2.9 (1.0)	<.001
Asking your doctors questions	3.1 (1.0)	3.2 (1.0)	.20
Total of all questions in the ACP Engagement Survey	2.8 (0.8)	3.0 (0.9)	<.001

[a]ACP: advance care planning.

[b]The ACP Engagement Survey evaluates 4 behavior change constructs (the subscales) within 4 the ACP domains—scale 1 to 5.

Usability of the Web-Based ACP Program

Of the 147 participants, 50 (34%) participants indicated they would use the program frequently, 25 (17%) participants would not and 72 (49%) participants were neutral. Of the 147 participants, 115 (78.2%) thought the program was easy to use and 26 (17.7%) participants were neutral. In total, of the 147 participants, 96 (65.3%) participants thought the functions were well-integrated and 115 (78.2%) participants felt they did not need to learn a lot before they could use the program. The mean total SUS score was 70 (SD 13; score 0-100). Table 3 presents the results.

Table 3. Usability of the web-based ACP[a] program according to the participants (N=147).

Usability	Participants, n (%)		
	Disagree[b]	Neutral[c]	Agree[d]
I think I would use this web-based program frequently.	25 (17)	72 (49)	50 (34)
I found the web-based program unnecessarily complex.	104 (70.7)	31 (21.1)	12 (8.2)
I thought the web-based program was easy to use.	6 (4.1)	26 (17.7)	115 (78.2)
I think I would need tech support to be able to use this web-based program.	123 (83.7)	13 (8.8)	11 (7.5)
I found the various functions in this web-based program were well-integrated.	6 (4.1)	45 (30.6)	96 (65.3)
I thought there was too much inconsistency in this web-based program.	116 (78.9)	25 (17)	6 (4.1)
I would imagine that most people would learn to use this web-based program very quickly.	11 (7.5)	41 (27.9)	95 (64.6)
I found the web-based program very cumbersome to use.	109 (74.1)	21 (14.3)	17 (11.6)
I felt very confident using the web-based program.	13 (8.8)	61 (41.5)	73 (49.7)
I need to learn a lot about this web-based program before I could effectively use it.	115 (78.2)	21 (14.3)	11 (7.5)

[a]ACP: advance care planning.

[b]Number of participants who scored 1 to 2 on the Likert scale.

[c]Number of participants who scored 3 on the Likert scale.

[d]Number of participants who scored 4 to 5 on the Likert scale.

Satisfaction With the Web-Based ACP Program

On average, participants rated the attractiveness of the program as 3.8 (SD 0.7; scale1-5), its comprehensibility as 4.2 (SD 0.6; scale 1-5), and the emotional support it provided as 3.4 (SD 0.8; scale 1-5). Of the 147 participants, 96 (65.3%) would recommend the program to others (Table 4). Participants rated their satisfaction with the program with 7.6 on average (SD 1.6; scale 1-10). Of 147 participants, a total of 80 (54.4%) participants added an explanation, of which 70 (88%) were positive, mentioning that the program was clear, easy to use, and important and that it made them think about preferences for treatment and care. Several mentioned that they would like to start with ACP. A few participants mentioned that it was confronting to complete the program or that they already arranged ACP.

Participants thought the amount of information in the program was enough (mean 5.6, SD 1.2; 1=too little, 5=exactly enough, and 10=too much). Of 147 participants total of 54 (36.7%) participants added an explanation, of which 47 (87%) were positive, mentioning the content of the program was enough and the information was clear; 6 (11.1%) participants found the information quite a lot to complete at once.

Table 4. Satisfaction with the web-based ACP[a] program according to the participants (N=147).

User satisfaction	Participants, n (%)		
	Disagree[b]	Neutral[c]	Agree[d]
Satisfaction with attractiveness (mean 3.8, SD 0.7)			
The web-based program is pleasant.	15 (10.2)	34 (23.1)	98 (66.7)
The web-based program is clear.	8 (5.4)	21 (14.3)	118 (80.3)
The web-based program is well-developed.	7 (4.8)	27 (18.4)	113 (76.9)
The web-based program is attractive.	10 (6.8)	50 (34)	87 (59.2)
Satisfaction with comprehensibility (mean 4.2, SD 0.6)			
The web-based program is understandable.	3 (2)	13 (8.8)	131 (89.1)
The texts in the web-based program are understandable.	2 (1.4)	11 (7.5)	134 (91.2)
The web-based program is easy to read.	2 (1.4)	13 (8.8)	132 (89.8)
Satisfaction with emotional support (mean 3.4, SD 0.8)			
The web-based program gives me self-confidence.	13 (8.8)	61 (41.5)	73 (49.7)
The web-based program gives me ease of mind.	20 (13.6)	64 (43.5)	63 (42.9)
I would recommend the web-based program to others.	9 (6.1)	42 (28.6)	96 (65.3)

[a]ACP: advance care planning.

[b]Number of participants who scored 1 to 2 on the Likert scale.

[c]Number of participants who scored 3 on the Likert scale.

[d]Number of participants who scored 4 to 5 on the Likert scale.

Discussion

Principal Findings

The web-based ACP program *Explore your preferences for treatment and care* [31] was considered usable and understandable. The program supported participants to engage in ACP and in thinking about their treatment and care preferences and to feel ready for ACP. The program supported participants to engage in several ACP domains, such as appointment of a health care representative and to think about what matters most in life. Participants were satisfied with the program and with the amount of information. The program gave almost half of the participants ease of mind, 65.3% (96/147) participants would recommend it to others.

Strengths and Limitations

The program was evidence-based and developed in cocreation with patients, relatives, and health care professionals; their input ensured that it would meet the needs of its potential users. We had a varied sample of participants with chronic diseases with different ages and levels of education.

As user data are not recorded in the program, we were not able to see whether participants completed the program and were unable to analyze their responses. Numbers of incomplete responses or dropout were quite high in the measurement immediately following the completion of the ACP program. Filling in the measurement required participants to return to the questionnaire after completing the program on a separate website or web page. It may be the case that this was not clear to participants or, alternatively, that they thought the program was too long or too difficult. However, the response rates in the measurement 2 months after completion of the program were sufficient and rather high according to the research portal: 92.5% (147/159); and we found no significant differences in participant characteristics in our baseline measurement that suggested we had no selection bias owing to the dropout.

Comparison With Prior Work

Most evidence-based, web-based ACP programs have been developed in the United States and only a few have been thoroughly evaluated [6]. We developed an evidence-based, interactive web-based ACP program in cocreation with patients, relatives, and health care professionals and sustainably embedded it in the frequently used and trusted general practitioners' platform *Thuisarts.nl*. We used the ACP Engagement Survey to evaluate the program's effects and found that it could support patients in ACP engagement. This confirmed the findings considering the web-based ACP program *Prepare For Your Care* from the United States [33-35]. We found changes in scores for contemplation about ACP, readiness for ACP, what matters most in life, surrogate decision makers, flexibility in medical decision-making, and total ACP engagement scores, which, according to the validation study of the original ACP Engagement Survey, indicated clinically meaningful changes [30].

The availability of the program on the web may improve access to ACP information at any preferred time and place; this can be important as ACP is considered a process over time. The web-based ACP program may be an addition to the traditional ACP process as facilitated by health care professionals, as it includes information, questions to be answered, and videos. We believe web-based programs should not replace discussions with relatives or health care professionals, but the program may support patients in preparing for ACP discussions [6]. Health care professionals may use the program as a tool to start ACP discussions with their patients. The program can support blended care by a combination of face-to-face conversations and the web-based ACP information; this fits within current developments of self-management and eHealth [6,36,37].

The program was launched in April 2020, and it has been frequently used (>78,000 visits by June 1, 2022).

Recommendations for Future Research

As most participants are born in the Netherlands, we recommend to evaluate the program in persons with other countries of birth as well. In addition, since the participants were members of an internet-based research portal, their level of computer skills may be above the average skill of the Dutch population. As readiness for ACP can differ across patients [25], we recommend to examine how web-based ACP programs affect ACP discussions between patients and health care professionals.

Conclusions

We developed an evidence-based, web-based ACP program *Explore your preferences for treatment and care* in cocreation with patients, relatives, and health care professionals. The before-and-after evaluation showed that the program can support people in taking first steps in ACP and in reflecting on preferences for treatment and care, by guiding them through the process of ACP using a stepwise approach. Participants perceived the program as usable and understandable, and they were satisfied with the program and the amount of information. Health care professionals may use the program as a tool to start ACP discussions with their patients. The program may increase awareness of ACP.

Acknowledgments

The authors would like to thank the members of the stakeholder group for their contribution to the development of the web-based ACP program: the contributing patients and relatives, the Dutch College of General Practitioners (Nederlands Huisartsen Genootschap), the Dutch Association for Kidney Patients (Nierpatiënten Vereniging Nederland), the Dutch Association for Patients (de Nederlandse Patiëntenvereniging), Agora (organization to promote the palliative approach), Nivel (the Netherlands institute for health services research), the University of Twente, and Vital Innovators. The authors would also like to thank all participants for participating in this study and the Netherlands Organization for Health Research and Development for the funding.

This work was supported by the Netherlands Organisation for Health Research and Development (grant 80-84400-98-356). The funding source had no involvement in the study design, collection, analysis, and interpretation of data, in the writing of the report and in the decision to submit the paper for publication.

Data Availability

Data will be available for verification for at least 15 years. All research data will be retained and stored in a study database at the central network server of the Erasmus MC, University Medical Center Rotterdam, and can only be accessed by those specifically granted access to the database involved.

Authors' Contributions

All authors contributed to the concept or design of the work or acquisition, analysis, or interpretation of data; drafted the paper or revised it critically for important intellectual content; approved the version to be published; and have participated sufficiently in the work to take public responsibility for appropriate portions of the content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example screenshots of the web-based advance care planning program "Explore your preferences for treatment and care".
[PDF File (Adobe PDF File), 907 KB - jmir_v24i12e38561_app1.pdf]

References

1. Rietjens JA, Sudore RL, Connolly M, van Delden JJ, Drickamer MA, Droger M, European Association for Palliative Care. Definition and recommendations for advance care planning: an international consensus supported by the European Association for Palliative Care. Lancet Oncol 2017 Sep;18(9):e543-e551. [doi: 10.1016/S1470-2045(17)30582-X] [Medline: 28884703]
2. Brinkman-Stoppelenburg A, Rietjens JA, van der Heide A. The effects of advance care planning on end-of-life care: a systematic review. Palliat Med 2014 Sep;28(8):1000-1025. [doi: 10.1177/0269216314526272] [Medline: 24651708]
3. Raijmakers NJ, Rietjens JA, Kouwenhoven PS, Vezzoni C, van Thiel GJ, van Delden JJ, et al. Involvement of the Dutch general population in advance care planning: a cross-sectional survey. J Palliat Med 2013 Sep;16(9):1055-1061. [doi: 10.1089/jpm.2012.0555] [Medline: 23767949]
4. Jabbarian LJ, Zwakman M, van der Heide A, Kars MC, Janssen DJ, van Delden JJ, et al. Advance care planning for patients with chronic respiratory diseases: a systematic review of preferences and practices. Thorax 2018 Mar;73(3):222-230. [doi: 10.1136/thoraxjnl-2016-209806] [Medline: 29109233]
5. Jimenez G, Tan WS, Virk AK, Low CK, Car J, Ho AH. Overview of systematic reviews of advance care planning: summary of evidence and global lessons. J Pain Symptom Manage 2018 Sep;56(3):436-59.e25 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.05.016] [Medline: 29807158]
6. van der Smissen D, Overbeek A, van Dulmen S, van Gemert-Pijnen L, van der Heide A, Rietjens JA, et al. The feasibility and effectiveness of web-based advance care planning programs: scoping review. J Med Internet Res 2020 Mar 17;22(3):e15578 [FREE Full text] [doi: 10.2196/15578] [Medline: 32181750]
7. Sudore RL, Knight SJ, McMahan RD, Feuz M, Farrell D, Miao Y, et al. A novel website to prepare diverse older adults for decision making and advance care planning: a pilot study. J Pain Symptom Manage 2014 Apr;47(4):674-686 [FREE Full text] [doi: 10.1016/j.jpainsymman.2013.05.023] [Medline: 23972574]
8. Green MJ, Levi BH. Development of an interactive computer program for advance care planning. Health Expect 2009 Mar;12(1):60-69 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00517.x] [Medline: 18823445]
9. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
10. van der Smissen D, Rietjens JA, van Gemert-Pijnen L, van Dulmen S, van der Heide A, Korfage IJ. Information needs of patients with chronic diseases and their relatives for web-based advance care planning: a qualitative interview study. BMC Palliat Care 2021 May 30;20(1):77 [FREE Full text] [doi: 10.1186/s12904-021-00770-x] [Medline: 34053435]
11. Thuisarts.nl. URL: https://www.thuisarts.nl/ [accessed 2022-11-16]
12. GPinfo. URL: https://gpinfo.nl/ [accessed 2022-11-8]
13. Spoelman WA, Bonten TN, de Waal MW, Drenthen T, Smeele IJ, Nielen MM, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Nov 09;6(11):e013166 [FREE Full text] [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]
14. Drenthen T, Beijaert RP, Jansen PW, Korevaar JC, Smeele IJ. [What do you think of Thuisarts.nl? Experiences after 3 years of www.Thuisarts.nl]. Ned Tijdschr Geneeskd 2014;158:A8282. [Medline: 25515391]
15. Sudore RL, Lum HD, You JJ, Hanson LC, Meier DE, Pantilat SZ, et al. Defining advance care planning for adults: a consensus definition from a multidisciplinary Delphi panel (S740). J Pain Symptom Manage 2017 May;53(5):821-32.e1 [FREE Full text] [doi: 10.1016/j.jpainsymman.2016.12.331] [Medline: 28062339]
16. 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]
17. Hak T, Van der Veer K, Jansen H. The Three-Step Test-Interview (TSTI): an observational instrument for pretesting self-completion questionnaires. Survey Res Method 2008 Dec;2(3):143-150 [FREE Full text] [doi: 10.18148/srm/2008.v2i3.1669]
18. Research Agency Flycatcher. Flycatcher Internet Research. URL: https://www.flycatcher.eu/en [accessed 2021-06-28]
19. Gobbens RJ, van Assen MA, Luijkx KG, Wijnen-Sponselee MT, Schols JM. The Tilburg Frailty Indicator: psychometric properties. J Am Med Dir Assoc 2010 Jun;11(5):344-355. [doi: 10.1016/j.jamda.2009.11.003] [Medline: 20511102]
20. Fransen MP, Van Schaik TM, Twickler TB, Essink-Bot ML. Applicability of internationally available health literacy measures in the Netherlands. J Health Commun 2011;16 Suppl 3:134-149. [doi: 10.1080/10810730.2011.604383] [Medline: 21951248]
21. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
22. van der Smissen D, van der Heide A, Sudore RL, Rietjens JA, Korfage IJ. Measuring advance care planning behavior in Dutch adults: translation, cultural adaptation and validation of the Advance Care Planning Engagement Survey. BMC Med Res Methodol 2021 Sep 25;21(1):194 [FREE Full text] [doi: 10.1186/s12874-021-01389-5] [Medline: 34563115]

23. Sudore RL, Heyland DK, Barnes DE, Howard M, Fassbender K, Robinson CA, et al. Measuring advance care planning: optimizing the advance care planning engagement survey. J Pain Symptom Manage 2017 Apr;53(4):669-81.e8 [FREE Full text] [doi: 10.1016/j.jpainsymman.2016.10.367] [Medline: 28042072]
24. Sudore RL, Stewart AL, Knight SJ, McMahan RD, Feuz M, Miao Y, et al. Development and validation of a questionnaire to detect behavior change in multiple advance care planning behaviors. PLoS One 2013 Sep 5;8(9):e72465 [FREE Full text] [doi: 10.1371/journal.pone.0072465] [Medline: 24039772]
25. Zwakman M, Jabbarian L, van Delden J, van der Heide A, Korfage I, Pollock K, et al. Advance care planning: a systematic review about experiences of patients with a life-threatening or life-limiting illness. Palliat Med 2018 Sep;32(8):1305-1321 [FREE Full text] [doi: 10.1177/0269216318784474] [Medline: 29956558]
26. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
27. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, FL, USA: CRC Press; 1996:189-194.
28. van Weert JC, van Noort G, Bol N, van Dijk L, Tates K, Jansen J. Tailored information for cancer patients on the Internet: effects of visual cues and language complexity on information recall and satisfaction. Patient Educ Couns 2011 Sep;84(3):368-378. [doi: 10.1016/j.pec.2011.04.006] [Medline: 21550757]
29. Bol N, van Weert JC, de Haes HC, Loos EF, de Heer S, Sikkel D, et al. Using cognitive and affective illustrations to enhance older adults' website satisfaction and recall of online cancer-related information. Health Commun 2014;29(7):678-688. [doi: 10.1080/10410236.2013.771560] [Medline: 24160560]
30. Shi Y, Barnes DE, Boscardin J, You JJ, Heyland DK, Volow AM, et al. Brief English and Spanish survey detects change in response to advance care planning interventions. J Pain Symptom Manage 2019 Dec;58(6):1068-74.e5 [FREE Full text] [doi: 10.1016/j.jpainsymman.2019.09.004] [Medline: 31539605]
31. Verken uw wensen voor zorg en behandeling. Thuisarts. 2020. URL: https://www.thuisarts.nl/keuzehulp/verken-uw-wensen-voor-zorg-en-behandeling [accessed 2022-04-05]
32. Statline. CBS. 2019. URL: https://opendata.cbs.nl/#/CBS/nl/: [accessed 2022-06-14]
33. Sudore RL, Boscardin J, Feuz MA, McMahan RD, Katen MT, Barnes DE. Effect of the PREPARE website vs an easy-to-read advance directive on advance care planning documentation and engagement among veterans: a randomized clinical trial. JAMA Intern Med 2017 Aug 01;177(8):1102-1109 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1607] [Medline: 28520838]
34. Lum HD, Barnes DE, Katen MT, Shi Y, Boscardin J, Sudore RL. Improving a full range of advance care planning behavior change and action domains: the PREPARE randomized trial. J Pain Symptom Manage 2018 Oct;56(4):575-81.e7 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.06.007] [Medline: 29940209]
35. Sudore RL, Schillinger D, Katen MT, Shi Y, Boscardin WJ, Osua S, et al. Engaging diverse English- and Spanish-speaking older adults in advance care planning: the PREPARE randomized clinical trial. JAMA Intern Med 2018 Dec 01;178(12):1616-1625 [FREE Full text] [doi: 10.1001/jamainternmed.2018.4657] [Medline: 30383086]
36. Stoevelaar R. Self-management and Advance Care Planning at the End of Life. University Medical Center Rotterdam. 2020. URL: https://repub.eur.nl/pub/135098/Proefschrift-Stoevelaar-final.pdf [accessed 2022-11-07]
37. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007 Apr 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]

Abbreviations

ACP: advance care planning
SUS: System Usability Scale

Edited by R Kukafka; submitted 07.04.22; peer-reviewed by J van Delden, M von Wagner; comments to author 26.05.22; revised version received 07.07.22; accepted 11.10.22; published 02.12.22.

Please cite as:
van der Smissen D, Rietjens JAC, van Dulmen S, Drenthen T, Vrijaldenhoven-Haitsma FRMD, Wulp M, van der Heide A, Korfage IJ The Web-Based Advance Care Planning Program "Explore Your Preferences for Treatment and Care": Development, Pilot Study, and Before-and-After Evaluation
J Med Internet Res 2022;24(12):e38561
URL: https://www.jmir.org/2022/12/e38561
doi:10.2196/38561
PMID:36459410

Original Paper

The Development of a Novel mHealth Tool for Obstructive Sleep Apnea: Tracking Continuous Positive Airway Pressure Adherence as a Percentage of Time in Bed

Angela Fidler Pfammatter[1], BA, MS, PhD; Bonnie Olivia Hughes[2], BS; Becky Tucker[2], BA; Harry Whitmore[2], BA, RPSGT; Bonnie Spring[1], PhD; Esra Tasali[2], MD

[1]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Evanston, IL, United States

[2]Department of Medicine, University of Chicago, Chicago, IL, United States

Corresponding Author:
Angela Fidler Pfammatter, BA, MS, PhD
Department of Preventive Medicine
Feinberg School of Medicine
Northwestern University
633 Clark St
Evanston, IL, 60208
United States
Phone: 1 847 491 3741
Email: angela@northwestern.edu

Abstract

Background: Continuous positive airway pressure (CPAP) is the mainstay obstructive sleep apnea (OSA) treatment; however, poor adherence to CPAP is common. Current guidelines specify 4 hours of CPAP use per night as a target to define adequate treatment adherence. However, effective OSA treatment requires CPAP use during the entire time spent in bed to optimally treat respiratory events and prevent adverse health effects associated with the time spent sleeping without wearing a CPAP device. Nightly sleep patterns vary considerably, making it necessary to measure CPAP adherence relative to the time spent in bed. Weight loss is an important goal for patients with OSA. Tools are required to address these clinical challenges in patients with OSA.

Objective: This study aimed to develop a mobile health tool that combined weight loss features with novel CPAP adherence tracking (ie, percentage of CPAP wear time relative to objectively assessed time spent in bed) for patients with OSA.

Methods: We used an iterative, user-centered process to design a new CPAP adherence tracking module that integrated with an existing weight loss app. A total of 37 patients with OSA aged 20 to 65 years were recruited. In phase 1, patients with OSA who were receiving CPAP treatment (n=7) tested the weight loss app to track nutrition, activity, and weight for 10 days. Participants completed a usability and acceptability survey. In phase 2, patients with OSA who were receiving CPAP treatment (n=21) completed a web-based survey about their interpretations and preferences for wireframes of the CPAP tracking module. In phase 3, patients with recently diagnosed OSA who were CPAP naive (n=9) were prescribed a CPAP device (ResMed AirSense10 AutoSet) and tested the integrated app for 3 to 4 weeks. Participants completed a usability survey and provided feedback.

Results: During phase 1, participants found the app to be mostly easy to use, except for some difficulty searching for specific foods. All participants found the connected devices (Fitbit activity tracker and Fitbit Aria scale) easy to use and helpful. During phase 2, participants correctly interpreted CPAP adherence success, expressed as percentage of wear time relative to time spent in bed, and preferred seeing a clearly stated *percentage goal* (“Goal: 100%”). In phase 3, participants found the integrated app easy to use and requested push notification reminders to wear CPAP before bedtime and to sync Fitbit in the morning.

Conclusions: We developed a mobile health tool that integrated a new CPAP adherence tracking module into an existing weight loss app. Novel features included addressing OSA-obesity comorbidity, CPAP adherence tracking via percentage of CPAP wear time relative to objectively assessed time spent in bed, and push notifications to foster adherence. Future research on the effectiveness of this tool in improving OSA treatment adherence is warranted.

(J Med Internet Res 2022;24(12):e39489) doi:10.2196/39489

KEYWORDS

obstructive sleep apnea; continuous positive airway pressure; CPAP adherence; weight loss; lifestyle

Introduction

Approximately one-third of the world's population is considered overweight or obese [1]. Obesity is a major risk factor for obstructive sleep apnea (OSA), a sleep disorder characterized by recurrent complete or partial upper airway obstruction that results in reduced oxygen levels at night, sleep fragmentation, and poor sleep quality [2]. OSA is a global public health and economic burden, estimated to affect one billion people worldwide [3-5]. Untreated OSA is commonly associated with daytime sleepiness and neurocognitive impairment, which increase the risk of motor vehicle accidents [6]. Furthermore, there is strong evidence that beyond the effects of excess weight, OSA is associated with increased cardiometabolic risk and all-cause mortality [7-9]. Currently, continuous positive airway pressure (CPAP) applied at night is considered the treatment of choice for OSA, and there is no US Food and Drug Administration–approved drug treatment for OSA [10]. CPAP works by delivering continuous air pressure and preventing upper airway closure during sleep. It can be easily applied using a variety of masks worn on the face at night and is highly efficacious in treating OSA. However, poor adherence to CPAP therapy is a common problem [11-13]. CPAP adherence is defined as CPAP use for >4 hours per night; evidence from clinical research studies and real-world data suggests that adherence is variable among individuals, with a large proportion of patients being nonadherent to the treatment [13-16]. Several interventions such as educational materials, motivational interviewing, remote monitoring, and mobile health (mHealth) technologies have been used to promote adherence to CPAP therapy but have provided limited clinical translation to routine patient care [11,17-19]. For implementation in clinical practice, interventions aimed at fostering treatment adherence should be cost-effective and scalable to large and diverse patient populations. Therefore, novel approaches are urgently needed to promote adherence to CPAP therapy. In addition, weight loss is often recommended for patients with OSA, but it remains a major challenge in this patient population [20,21]. Thus, there is a critical need to develop new tools to address these important clinical barriers in OSA management.

Effective OSA treatment requires all-night CPAP use, that is, 100% of the time spent in bed, to optimally treat respiratory events, hypoxia, and sleep fragmentation and thus prevent adverse health effects associated with hours slept without wearing a CPAP device. Therefore, an accurate calculation of adherence to CPAP use requires a "denominator," that is, hours spent in bed, which differs from adherence to, for example, medication use. However, current CPAP adherence tracking systems (eg, smartphone apps) simply capture the number of hours the CPAP device is used per night but do not account for hours spent in bed without using the CPAP device [22]. Moreover, in clinical practice, according to Medicare criteria, patients who wear their CPAP device for ≥4 hours per night for 70% of the nights are considered "adherent" to therapy. These policy recommendations also have implications on health equity, given the known racial or ethnic and socioeconomic differences in sleep patterns, particularly sleep duration [23]. The use of 4-hour CPAP wear as a cutoff point defining adequate treatment adherence is arbitrary and could be misleading because sleep patterns can vary considerably from night to night and among individuals, making it necessary to measure CPAP adherence relative to time spent in bed. For example, based on a 4-hour CPAP adherence threshold, a patient who uses a CPAP device for 5 hours but spends 8 hours in bed per night would be considered adherent. However, the patient's true CPAP adherence should be only 63% as a percentage of time spent in bed. Hence, the patient's treatment adherence should be considered suboptimal, and the patient should be advised clinically to increase their CPAP use. Implementing more correct CPAP adherence goals and guidelines requires quantification of CPAP wear time in proportion to the time spent in bed. In this regard, wearable mHealth devices offer a promising tool for at-home monitoring of sleep using accelerometry-based technology [24-26]. Combining such mHealth technology with CPAP use data provides a unique opportunity to revolutionize OSA treatment adherence guidelines and fulfill a critically unmet need for patients and health care providers.

Our goal was to develop a customized mHealth tool to support treatment adherence to both CPAP and weight loss recommendations in patients with OSA. We developed and tested a unique CPAP adherence tracking module that measured CPAP wear in proportion to the time spent in bed. We aimed to integrate this new CPAP adherence tracking module into our previously developed mHealth technology targeting lifestyle behaviors (nutrition and physical activity) to achieve weight loss in the population with OSA. This is a formative work that used an iterative, user-centered process to design a new CPAP adherence tracking module that integrated with an existing weight loss app; thus, the effectiveness of this tool in treatment adherence has not been tested or reported.

Methods

Overview

This study was conducted between January 2020 and March 2021. This study leveraged our existing technology platform for delivering smartphone apps to patients and web-based dashboards to interventionists [27]. The platform was built specifically to support behavioral interventions [27-29] and has the flexibility to display participant- and coach-facing features according to specifications such as the research study to which they belong and the behaviors targeted for change. Existing participant-facing smartphone apps include features for behavior change interventions to foster weight loss, healthier diet quality, physical activity, and smoking cessation.

Through an iterative user-centered design process, we developed and tested a new CPAP adherence tracking module that integrated our existing weight loss app with nutrition, activity, and weight tracking features. Our platform combined information from multiple devices, including Fitbit activity

trackers and Fitbit Aria scale using the Fitbit application programming interface (API) and CPAP devices (ResMed AirSense 10 AutoSet) using the AirView API. In addition, we refined a companion interventionist web-based dashboard [28] to present interventionists with relevant information from the new CPAP adherence tracking module and the integrated diet, activity, and weight tracking features that they use to tailor coaching.

Iterative User-Centered Design Process

A 3-phase iterative user-centered process was implemented to develop a new CPAP adherence tracking module that integrated our existing weight loss app. Figure 1 provides details of the study design methods and participant characteristics for each study phase. Adult men and women were recruited according to inclusion criteria of age (aged 20-65 years) and an OSA diagnosis. There were no exclusion criteria based on BMI, race or ethnicity, or other demographic characteristics.

In phase 1 (mean age 45, SD 8 years), we aimed to collect feedback on the existing weight loss app from patients with known OSA who were receiving CPAP treatment. Patients who returned for a follow-up appointment at the University of Chicago Sleep Disorders Clinic were recruited if they had a prior diagnosis of OSA and a previously prescribed CPAP device. Potential participants were given a study flyer by their treating physician during clinic visits. Interested individuals discussed the study details with the study coordinator and were enrolled in the study after obtaining informed consent. The participants tested the existing weight loss app along with connected devices, including a wrist-worn Fitbit activity tracker (Fitbit Inspire HR) and a Fitbit weight scale (Fitbit Aria), to self-monitor and receive feedback on their dietary intake, physical activity, and weight for 10 days. At the end of the 10-day period, they completed a survey about the app design that included the System Usability Scale [30], a measure of usability on a 0- to 100-point scale where 65 is the threshold for a system to be considered usable. Participants were compensated US $100 for participation.

In phase 2 (mean age 47, SD 9 years), patients with known OSA who were receiving CPAP treatment were studied. Patients who returned for a follow-up appointment at the University of Chicago Sleep Disorders Clinic were recruited after obtaining informed consent if they had a prior diagnosis of OSA and a previously prescribed CPAP device. The participants completed a web-based survey to provide their preferences for graphical displays and interpretations of various wireframe images displaying information from the CPAP tracking module. Figure 2 illustrates examples of the wireframes shown to the participants. Using the various metrics of CPAP use that are available via the AirView API, the phase 2 of the study aimed to develop a customized CPAP adherence tracking module that displayed information deemed most relevant and helpful by patients receiving CPAP treatment. Participants were asked to describe their understanding of the information displayed on each screen (eg, percentage adherence—percentage CPAP wear time relative to objectively assessed time spent in bed; mask leak; and apnea-hypopnea index [AHI], that is, number of respiratory events per hour of sleep) and how they would react or change their CPAP use behavior after seeing that particular screen displayed on the app. This phase was performed iteratively such that participants' survey responses elicited subsequent modifications in the wireframe images, which led to a revised survey design capturing features of the updated wireframe images. A total of 4 consecutive versions of the CPAP tracking module and survey were produced and presented to participants based on the prior participants' feedback. Thus, during phase 2, participants completed a version of this iterative survey. Design decisions were made based on the participants' feedback and survey data, and a new CPAP module was developed for Android smartphones. Participants were compensated US $100 for participation.

In phase 3 (mean age 55, SD 8 years), patients who were newly diagnosed with OSA at the University of Chicago Sleep Disorders Clinic, who were CPAP naive, and who owned an Android smartphone were recruited after obtaining informed consent. This phase involved in-field testing of the CPAP tracking module of the app while participants were using a newly prescribed ResMed autoadjusting CPAP machine (AirSense 10 AutoSet), Fitbit Aria scale, and Fitbit. The app combined the new CPAP module with weight loss tracking features to enable participants to experience and evaluate the integrated app. The home screen of the CPAP module graphically depicted CPAP adherence as CPAP use relative to the time spent in bed, expressed as a percentage. The home screen also displayed CPAP use and time spent in bed separately in hours and minutes. An additional page in the app displayed CPAP adherence over a time frame of weeks or months as well as details of daily adherence and mask leak. Each participant was provided with a ResMed AirSense 10 AutoSet CPAP device, Fitbit Aria scale, and Fitbit activity monitor (Fitbit Inspire HR) to use for 3 to 4 weeks. Fitbit data were used to track the time spent in bed, and ResMed data were used to track CPAP wear time to allow calculation of percentage of CPAP adherence relative to the time spent in bed. The participants also received weekly phone calls to troubleshoot any CPAP-, Fitbit-, or app-related issues. In addition, some participants in the later part of the testing received push notifications within an hour of their self-reported bedtime as a reminder to wear CPAP device in the evening. Moreover, a message was sent upon Fitbit sensing waking to remind the participant to wear and synchronize the Fitbit device. Upon completion of phase 3, participants completed a survey that included the System Usability Scale [30] and questions about their positive and negative feedback on the app's design and burden of use. Participants were compensated US $230 for participation.

Figure 1. Study design and participant characteristics for each study phase. CPAP: continuous positive airway pressure; OSA: obstructive sleep apnea.

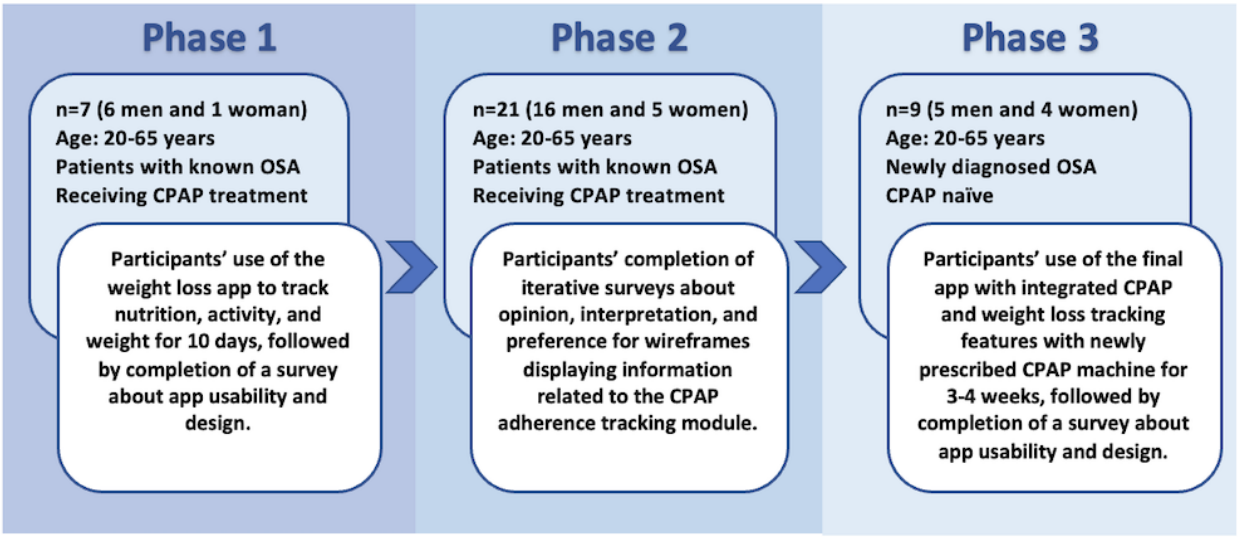
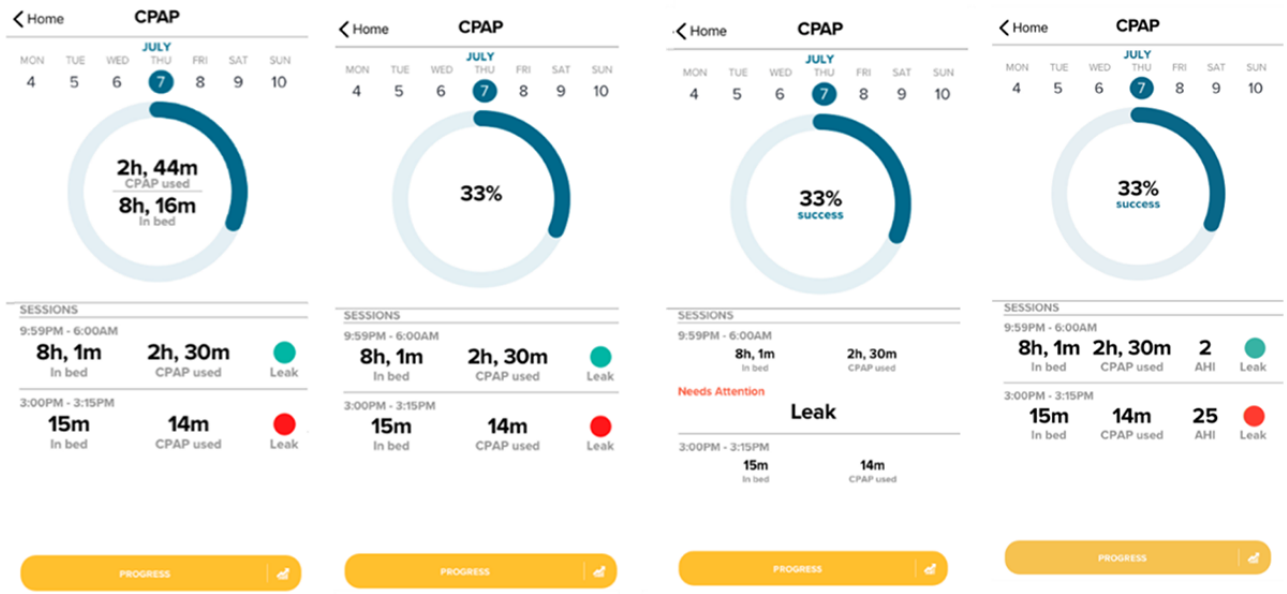


Figure 2. Example wireframes displaying continuous positive airway pressure tracking module shown to participants by web-based surveys.CPAP: continuous positive airway pressure; AHI: apnea-hypopnea index.



Ethics Approval

This study was approved by the University of Chicago Institutional Review Board (#19-1446). All participants provided informed consent before the study after a member of the research team explained all details of the study and the participants received satisfactory answers to all of their questions.

Results

Overview

A total of 37 patients with OSA (27 men and 10 women) participated in this study. Table 1 summarizes the primary questions and main findings of each study phase.

During phase 1, a total of 7 participants (6 men and 1 woman) were enrolled. All participants found the app easy to use, except for some difficulty in searching for specific foods in the database. All participants also endorsed that they felt confident when using the app. Of the 7 participants, 3 (43%) reported that they did not like the overall design of the app, and all 3 of them attributed their dislike to the nutrition tracking. When asked about the feature they liked best, of the 7 participants, 3 (43%) endorsed nutrition tracking, 3 (43%) endorsed weight tracking, and 2 (29%) endorsed physical activity tracking. None of the participants found the app too difficult to use, reported that they would need help, or stated that it was too complex. None of the participants encountered problems with the connected devices, that is, Fitbit activity tracker (Fitbit Inspire HR) or Fitbit weight scale (Fitbit Aria), and the participants commented that both were "easy to use" and that they liked "how the information from it went straight into the app." Overall, the participants rated the app on an average of 83 on the 100-point System Usability Scale, a score considered to be indicative of a highly usable system compared with other similar systems from a recent review [30,31].

During phase 2, a total of 21 patients were enrolled. The first 6 participants completed the first version of the survey and provided initial responses to graphical representations (an example shown in Figure 2) of the percentage of CPAP use relative to time spent in bed as well as mask leak indicator. In addition to seeing percentage of adherence, participants also wanted to see details of the time spent in bed and CPAP wear time. They also preferred to have mask leak indicators that were always displayed rather than only when a problem arose. The next 3 participants were shown the representations preferred by the initial 6 participants and were asked questions about their understanding of the information displayed. Moreover, they were shown a wireframe that included AHI in addition to mask leaks. Participants accurately interpreted the meaning of the information displayed by the preferred wireframes and the follow-up behavior they should do in response. However, they were confused about how to interpret AHI. Given this feedback, the next 5 participants were shown 3 options for displaying information related to CPAP goal attainment and 2 options for showing mask leak information without the AHI. Participants correctly interpreted the percentage of success and matched their percentage to a goal. However, their opinions about the best way to display the percentage of time adherent to CPAP wear relative to the time spent in bed were mixed. Participants interpreted leak information correctly, but some had concerns about having a green light indicator for the absence of leak, particularly when it was displayed, in addition to poor CPAP adherence (ie, low percentage of success), which they found confusing. The final 7 participants responded to visuals depicting adherence with and without the CPAP percentage goal indicated and displayed leak information with and without "no leak" text to represent the absence of a problem. Most preferred having no indicator displayed when there was not a problem with leak. Participants preferred seeing the CPAP percentage goal for which they were striving. The results of this phase characterized the Android version of the app delivered in phase 3 of the study.

The phase 3 study used app interfaces, as shown in Figure 3. A total of 10 patients with newly diagnosed OSA who were CPAP naive consented to participate. A patient discontinued participation after providing consent, and no data were collected. All 9 participants endorsed that they liked the design and that the app was easy to navigate. Of the 6 users who received push notifications, all found them helpful and well timed. When asked about their favorite features, participants reported a mix of nutrition, physical activity, and CPAP adherence, with half of them endorsing physical activity. When asked about their least favorite features, nutrition, activity, and CPAP adherence were each chosen by 20% (2/10) of the participants. Participants rated the app an average of 80 on the 100-point System Usability Scale, a score considered to be indicative of a highly usable system [31]. Some participants reported during calls that the app interface did not always show 100% CPAP adherence even when they wore their CPAP device for the entire time they spent in bed. After carefully examining our data in response to this feedback, we adjusted the percentage of CPAP adherence calculation logic to allow a 15-minute buffer for the time spent in bed captured by the Fitbit activity tracker. This minor adjustment in our calculation logic protected against a margin of measurement error for time spent in bed as captured by Fitbit [32], while not compromising the accuracy of the percentage of CPAP adherence measure.

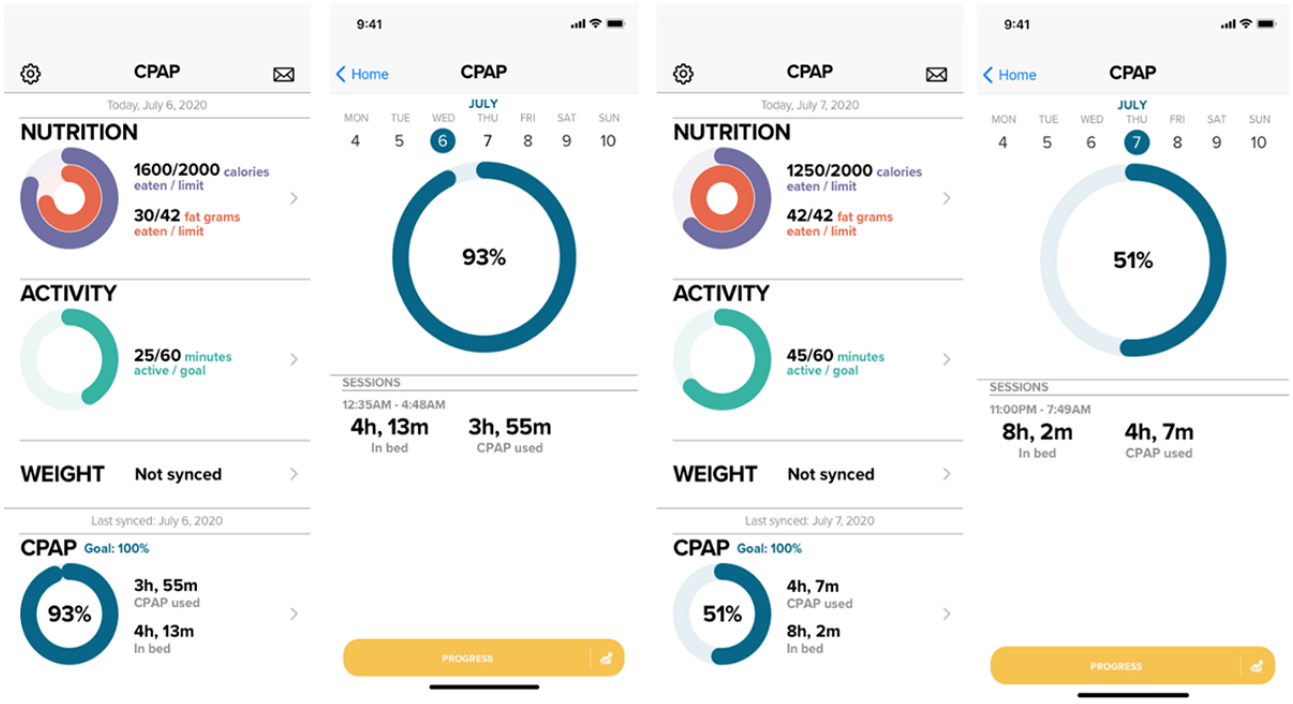
Table 1. Primary questions and main findings.

Phase	Primary questions	Main findings
1	• Are the previously developed weight loss app features and the connected devices usable by the study sample? • Are Fitbit activity tracker and Fitbit Aria scale use helpful for tracking their activity and weight, respectively? • What design aspects of the weight loss app need improvement?	• Rated easy to use by all participants (rated ≥65 on the System Usability Scale); scores ranged 70 to 92.5 out of a possible 100 • Helpful and easy to use devices; no issues noted; positive aspect of devices noted by all participants • Difficulty searching for specific foods in the database by 4 participants
2	• What CPAP[a] adherence tracking information is helpful to participants? • Which graphical displays about CPAP are easy to interpret? • Which CPAP tracking features lead to accurate interpretations by participants?	• Less information preferable (eg, AHI[b] not displayed) • Color to indicate mask leak only when there is a problem is preferred • Clearly displayed 100% success goal is needed • Success depicted as percentage of adherence is easily and accurately interpretable
3	• Is the new CPAP module integrated with the weight loss app usable and acceptable to participants?	• Found to be not burdensome and has an excellent usability rating; all participants rated >65 on the System Usability Scale with a range of 72.5 to 92.5 • Request for reminders at appropriate times of day, that is, reminder 1 hour before bedtime to wear a CPAP device and reminder to synchronize Fitbit in the morning upon waking

[a]CPAP: continuous positive airway pressure.

[b]AHI: apnea-hypopnea index.

Figure 3. Final app with integrated continuous positive airway pressure, nutrition, activity, and weight tracking features. CPAP: continuous positive airway pressure.



Final App Description

The final app design is shown in Figure 3. The app allowed participants to track CPAP adherence, nutrition, physical activity, and weight. The CPAP tracking module showed the percentage of CPAP adherence each night by darkening a portion of a circle's circumference to represent the percentage of time the CPAP device was used (conveyed by the CPAP device via the AirView API) during the time spent in bed (conveyed by the Fitbit wrist activity monitor). "Goal: 100%" was displayed above the circle to remind participants to wear their CPAP the entire time they spent in bed. The section also displayed, in hours and minutes, the duration of CPAP wear time and time spent in bed separately on the right side of the circle. Thus, CPAP tracking in the final app provided participants with meaningful feedback that aimed to encourage and improve their CPAP adherence. If a participant's mask had a high leak at any point, the app marked the night with the label "LEAK" and a red dot to alert the participant to troubleshoot mask issues and contact their health care provider, as required. Individualized push notifications approximately 1 hour before the participant's bedtime reminded participants to use the CPAP device every night and to synchronize their Fitbit every morning upon waking. Notifications were tailored to the participant's self-reported bedtime for the first week and then adjusted to their average bedtime shown by Fitbit, as data become available.

Nutrition intake was tracked when participants searched for and selected foods in the app or added custom foods or recipes by recording calories and fat gram content. Calories and fat gram intake were calculated every day. Participants were shown daily calorie and fat gram goals within the nutrition section, which were calculated based on their weight. Physical activity was tracked in the app by automatically transferring data from the participants' wrist-worn Fitbit activity monitor via Bluetooth. Physical activities could also be manually entered by searching the app database, which included a compendium of physical activities rated by their intensity. An activity could be tracked by selecting the specific activity and its duration. Weight was tracked automatically by synchronizing with the Fitbit Aria scale; weight could also be manually entered into the app if needed. For every element tracked by the app, progress over time could be viewed as a weekly or monthly line graph.

Figure 3 illustrates the final app screenshots for a representative participant on 2 consecutive days. As seen, CPAP wear times are quite similar on Wednesday and Thursday nights (3 hours 55 minutes vs 4 hours 7 minutes). On the basis of the current clinical threshold that defines 4 hours of CPAP wear as adherent, this patient would be considered CPAP "nonadherent" on Wednesday night (3 hours 55 minutes of CPAP wear time) and "adherent" on Thursday night (4 hours 7 minutes of CPAP wear time). However, when time spent in bed was considered using our newly developed CPAP tracking module, the patient's true CPAP adherence relative to the time in spent bed was 93% on Wednesday night and 51% on Thursday night. Thus, our new CPAP adherence metric accounting for time spent in bed provided important and clinically meaningful information about CPAP adherence, which is not captured by current CPAP tracking technologies.

Discussion

Principal Findings

We engaged in a 3-phase iterative, user-centered process to develop and test a smartphone app that aimed to support both CPAP adherence and weight loss behaviors in patients with OSA and overweight or obesity. Our user-centered design process identified key information that participants found useful for tracking their adherence to CPAP as well as user interfaces

that participants found easy to interpret. We found that participants were not only satisfied with less information about their CPAP use but also that their interpretations were more accurate when less information was provided. In addition, there was a need for explicit descriptions of the information provided, which we accomplished by providing descriptors or comparators such as "Goal: 100%." The information collected via the surveys supported our design decisions regarding appropriate features, functions, and user interface for the study. Overall, the participants rated the system's usability as high [31] and reported positive impressions of the app features. In addition, based on participants' feedback during phase 3, we enhanced the usefulness of the app by adding push notifications to remind users to use their CPAP device (ie, 1 hour before bedtime) and wrist activity monitor (ie, synchronize device every morning) at meaningful times of the day.

Comparison With Prior Work

A unique feature of our app was the tracking of CPAP use relative to objectively assessed time spent in bed, represented as percentage of CPAP adherence. To harvest the data needed to calculate percentage of CPAP adherence, we leveraged connected technologies, in this case, a ResMed CPAP device (AirSense 10 AutoSet) to track CPAP wear time and a Fitbit wearable activity sensor to track the time spent in bed. This novel CPAP adherence metric provided markedly different information than the available CPAP tracking technologies, which simply report how many hours a CPAP device was used without accounting for the time spent in bed without using a CPAP device. Although OSA can be effectively treated only when CPAP is used during the entire time spent in bed, current clinical guidelines categorize patients as adherent to treatment based on a cutoff point of 4 hours of CPAP use [33]. This adherence definition is widely accepted, but it is primarily based on expert opinion and remains in common use today despite lack of evidence showing that it is sufficient or has any health benefits compared with other more specific measures of use duration [11]. By leveraging wearable sensor technology that can objectively capture time spent in bed, our new app provided a novel, more informative, and clinically meaningful measure of CPAP adherence that can be implemented into clinical guidelines. First, we defined a new CPAP adherence metric that considered both CPAP wear and time spent in bed. Second, our mHealth tool was the first technology to capture this novel percentage of CPAP adherence metric. Thus, the use of the app and the data it provided filled an important gap in the management of OSA for patients and health care providers. Future rigorous research in diverse populations with OSA is warranted using this new CPAP adherence metric and mHealth technology to investigate the role of percentage of CPAP adherence in a variety of patient-centered and clinical outcomes. Future studies can also provide novel insights into the dose-response effect of CPAP adherence on cognitive, cardiovascular, and metabolic outcomes [34].

To date, the effectiveness of eHealth interventions in improving CPAP adherence remains uncertain, highlighting the need for newly designed technology-supported interventions for OSA [19]. As our app fed back information about CPAP use relative to the time spent in bed, it can support patients in reaching their goal of 100% CPAP adherence, that is, wearing their CPAP during the entire time they spent in bed. Real-time app data on percentage of CPAP adherence can be used not only as a self-management tool for patients but also as a monitoring tool for health care providers. Our existing technology platform also allowed for a web-based dashboard to display patient information from the app so that both the patient and provider can see progress toward goals. In addition, the app integrated lifestyle behavior tracking features (diet, physical activity, and body weight) and thus has the potential to enhance self-management and positive behavior change toward weight loss goals in patients with OSA [19]. Similar to our integrated app design, other emerging mHealth technologies target multiple behavior change interventions in the population with OSA [35,36].

The iterative user-centered design process used to develop our final app integrating CPAP tracking has notable strengths [37,38]. A few prior apps designed for populations with OSA relied primarily on the views expressed by clinical experts in focus groups with minimal feedback from patients [35]. By engaging patients in the design process from the outset and throughout, we increased the likelihood that the final app would be easy to use, helpful, and engaging to the targeted end users [37,39,40]. We used low-fidelity wireframes to represent user interfaces and embedded them in surveys to gather feedback quickly, without requiring time-consuming programming. The iterative nature of the surveys allowed us to respond promptly to end-user feedback about features or graphical displays that were not functional, not interpreted correctly, or not liked, bringing us closer to a feasible and acceptable interface. The process allowed our research team to progress by finding features, functions, and interfaces that accurately represented the required information and that satisfied the intended users, before engaging in extensive and costly programming efforts involved in app design.

Limitations

Our study has several limitations. Although new participants were enrolled in each study phase, resulting in a total sample of 37 that included patients with OSA who were experienced with CPAP use and those who were CPAP naive, the sample was collected from a single center, and a larger, more diverse sample size may provide additional insights and feedback for further improvements and refinements to the app. In this formative work to cocreate an app to support end users, we did not collect demographic characteristics, except for age and gender. Indeed, additional patient characteristics (eg, race or ethnicity, socioeconomic status, and prior CPAP adherence) may affect the usability outcomes of our tool. Thus, as a next step, large studies in diverse patient populations are needed to derive more generalizable assessments of the usability and efficacy of this mHealth tool. In the future, an iOS version of the app will also be needed to increase generalizability.

Our mHealth tool relied on specific manufacturers and used "consumer grade" sleep trackers, which was a pragmatic choice. Although each sleep-tracking device has its own margin of measurement error, the devices and apps are meant to support behavior as part of an intervention rather than for diagnostic

validity. Notably, the research-grade devices (eg, Actigraphy) do not allow people to receive real-time feedback on their behavior from the device, which is critical for self-monitoring and intervention success. Although absolute values may have some margin of error depending on the tracking device, they can still reliably assess trends over time for a given individual. Moreover, in our study, a recent generation Fitbit model (ie, Fitbit Inspire HR) was used, which performs better than the early generation models (owing to the addition of heart rate into the algorithm), especially in differentiating wake from sleep [24]. Nevertheless, over time, improvements and updates will need to be made to our algorithm to keep pace with the rapidly changing technology. Our final CPAP module does not display "AHI" based on feedback from participants who were confused about how to interpret it and preferred less information to be displayed. However, misinterpretation of AHI is possible. In future versions of the app, an AHI metric, that is, apnea burden during the "off-CPAP time," could be displayed to track treatment effectiveness [41]. Such an additional feature could potentially foster patient adherence, which warrants further rigorous testing in larger samples.

In this study, we opted to measure the percentage of CPAP adherence based on the time spent in bed captured by Fitbit and not the actual time spent asleep, which may appear as a potential limitation. CPAP use would ideally be required during the entire sleep period. However, CPAP adherence displayed as a percentage of time spent in bed is more meaningful for patients (ie, end users) in meeting the goal of CPAP use during all sleep periods occurring over time spent in bed. In addition, there is evidence to suggest that Fitbit activity trackers have acceptable levels of measurement accuracy for the time spent in bed compared with research-grade accelerosensors but may either overestimate or underestimate the actual sleep duration depending on the selected sleep-mode setting [24,32]. Finally, it is noteworthy that testing the efficacy of the app was beyond the scope of this study; thus, future trials are necessary.

Conclusions

We developed a new mHealth tool that filled a significant gap in the clinical management of patients with OSA. Our app used a novel CPAP adherence metric, that is, percentage of CPAP adherence that measures CPAP use relative to the time spent in bed and allows tracking of lifestyle behaviors targeting weight loss, such as diet, physical activity, and weight, for use in the population with OSA and comorbid overweight or obesity. The newly developed mHealth technology allowed tracking of both CPAP adherence and lifestyle behaviors, giving it the potential to support multiple behavior changes that optimize care for patients with OSA.

Acknowledgments

The authors would like to acknowledge and thank all the participants who volunteered their time during the study and Ive Arnaoudova for technical assistance in preparing the manuscript. The authors acknowledge ResMed for providing continuous positive airway pressure devices (AirSense 10 AutoSet), supplies for the study, and access to and guidance on the use of the AirView application programming interface. Research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (NIH) (grants R01DK120312 and R01DK125749). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

References

1. The GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 06;377(1):13-27. [doi: 10.1056/nejmoa1614362]
2. Eckert DJ, Malhotra A. Pathophysiology of adult obstructive sleep apnea. Proc Am Thorac Soc 2008 Feb 15;5(2):144-153 [FREE Full text] [doi: 10.1513/pats.200707-114MG] [Medline: 18250206]
3. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MS, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respiratory Med 2019 Aug;7(8):687-698. [doi: 10.1016/s2213-2600(19)30198-5]
4. Lyons MM, Bhatt NY, Pack AI, Magalang UJ. Global burden of sleep-disordered breathing and its implications. Respirology 2020 Jul;25(7):690-702 [FREE Full text] [doi: 10.1111/resp.13838] [Medline: 32436658]
5. Hidden health crisis costing America billions: underdiagnosing and undertreating obstructive sleep apnea draining healthcare system. American Academy of Sleep Medicine. 2016. URL: https://aasm.org/resources/pdf/sleep-apnea-economic-crisis.pdf [accessed 2022-11-18]
6. Patel SR. Obstructive sleep apnea. Ann Intern Med 2019 Dec 03;171(11):ITC81-ITC96. [doi: 10.7326/AITC201912030] [Medline: 31791057]
7. Wang X, Ouyang Y, Wang Z, Zhao G, Liu L, Bi Y. Obstructive sleep apnea and risk of cardiovascular disease and all-cause mortality: a meta-analysis of prospective cohort studies. Int J Cardiol 2013 Nov 05;169(3):207-214. [doi: 10.1016/j.ijcard.2013.08.088] [Medline: 24161531]

8. Yeghiazarians Y, Jneid H, Tietjens JR, Redline S, Brown DL, El-Sherif N, et al. Obstructive sleep apnea and cardiovascular disease: a scientific statement from the American heart association. Circulation 2021 Jul 20;144(3):e56-e67 [FREE Full text] [doi: 10.1161/CIR.0000000000000988] [Medline: 34148375]
9. Pamidi S, Tasali E. Obstructive sleep apnea and type 2 diabetes: is there a link? Front Neurol 2012;3:126 [FREE Full text] [doi: 10.3389/fneur.2012.00126] [Medline: 23015803]
10. Patil SP, Ayappa IA, Caples SM, Kimoff RJ, Patel SR, Harrod CG. Treatment of adult obstructive sleep apnea with positive airway pressure: an American academy of sleep medicine clinical practice guideline. J Clin Sleep Med 2019 Feb 15;15(2):335-343 [FREE Full text] [doi: 10.5664/jcsm.7640] [Medline: 30736887]
11. Bakker JP, Weaver TE, Parthasarathy S, Aloia MS. Adherence to CPAP: what should we be aiming for, and how can we get there? Chest 2019 Jun;155(6):1272-1287. [doi: 10.1016/j.chest.2019.01.012] [Medline: 30684472]
12. Weaver TE. Don't start celebrating--CPAP adherence remains a problem. J Clin Sleep Med 2013 Jun 15;9(6):551-552 [FREE Full text] [doi: 10.5664/jcsm.2746] [Medline: 23772187]
13. Weaver TE, Grunstein RR. Adherence to continuous positive airway pressure therapy: the challenge to effective treatment. Proc Am Thorac Soc 2008 Feb 15;5(2):173-178 [FREE Full text] [doi: 10.1513/pats.200708-119MG] [Medline: 18250209]
14. Weaver TE. Novel aspects of CPAP treatment and interventions to improve CPAP adherence. J Clin Med 2019 Dec 16;8(12):2220 [FREE Full text] [doi: 10.3390/jcm8122220] [Medline: 31888148]
15. Cistulli PA, Armitstead J, Pepin J, Woehrle H, Nunez CM, Benjafield A, et al. Short-term CPAP adherence in obstructive sleep apnea: a big data analysis using real world data. Sleep Med 2019 Jul;59:114-116 [FREE Full text] [doi: 10.1016/j.sleep.2019.01.004] [Medline: 30799254]
16. Patel SR, Bakker JP, Stitt CJ, Aloia MS, Nouraie SM. Age and sex disparities in adherence to CPAP. Chest 2021 Jan;159(1):382-389 [FREE Full text] [doi: 10.1016/j.chest.2020.07.017] [Medline: 32687910]
17. Sawyer AM, Gooneratne NS, Marcus CL, Ofer D, Richards KC, Weaver TE. A systematic review of CPAP adherence across age groups: clinical and empiric insights for developing CPAP adherence interventions. Sleep Med Rev 2011 Dec;15(6):343-356 [FREE Full text] [doi: 10.1016/j.smrv.2011.01.003] [Medline: 21652236]
18. Stepnowsky CJ, Palau JJ, Gifford AL, Ancoli-Israel S. A self-management approach to improving continuous positive airway pressure adherence and outcomes. Behav Sleep Med 2007 Mar 29;5(2):131-146. [doi: 10.1080/15402000701190622] [Medline: 17441783]
19. Aardoom JJ, Loheide-Niesmann L, Ossebaard HC, Riper H. Effectiveness of eHealth interventions in improving treatment adherence for adults with obstructive sleep apnea: meta-analytic review. J Med Internet Res 2020 Feb 18;22(2):e16972 [FREE Full text] [doi: 10.2196/16972] [Medline: 32130137]
20. Billings ME, Krishnan V, Su G, Donovan LM, Patel SR, Hudgel DW, et al. Clinical practice guideline summary for clinicians: the role of weight management in the treatment of adult obstructive sleep apnea. Ann Am Thorac Soc 2019 Apr;16(4):405-408. [doi: 10.1513/AnnalsATS.201810-708CME] [Medline: 30742491]
21. Joosten SA, Hamilton GS, Naughton MT. Impact of weight loss management in OSA. Chest 2017 Jul;152(1):194-203. [doi: 10.1016/j.chest.2017.01.027] [Medline: 28185772]
22. Schwab RJ, Badr SM, Epstein LJ, Gay PC, Gozal D, Kohler M, et al. An official american thoracic society statement: continuous positive airway pressure adherence tracking systems. The optimal monitoring strategies and outcome measures in adults. Am J Respir Crit Care Med 2013 Sep 01;188(5):613-620. [doi: 10.1164/rccm.201307-1282st]
23. Grandner MA, Williams NJ, Knutson KL, Roberts D, Jean-Louis G. Sleep disparity, race/ethnicity, and socioeconomic position. Sleep Med 2016 Feb;18:7-18 [FREE Full text] [doi: 10.1016/j.sleep.2015.01.020] [Medline: 26431755]
24. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273 [FREE Full text] [doi: 10.2196/16273] [Medline: 31778122]
25. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, HUGOPSY Network, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
26. Perez-Pozuelo I, Zhai B, Palotti J, Mall R, Aupetit M, Garcia-Gomez JM, et al. The future of sleep health: a data-driven revolution in sleep science and medicine. NPJ Digit Med 2020 Mar 23;3(1):42 [FREE Full text] [doi: 10.1038/s41746-020-0244-4] [Medline: 32219183]
27. Pfammatter AF, Nahum-Shani I, DeZelar M, Scanlan L, McFadden HG, Siddique J, et al. SMART: study protocol for a sequential multiple assignment randomized controlled trial to optimize weight loss management. Contemp Clin Trials 2019 Jul;82:36-45 [FREE Full text] [doi: 10.1016/j.cct.2019.05.007] [Medline: 31129369]
28. Pfammatter AF, Battalio S, Olvera C, DeZelar M, Moore D, Scanlan L, et al. The EVO study protocol for a randomized controlled evaluation trial of an optimized weight management intervention. Contemp Clin Trials 2022 May;116:106750. [doi: 10.1016/j.cct.2022.106750] [Medline: 35378301]
29. Pfammatter AF, Champion KE, Finch LE, Siddique J, Hedeker D, Spring B. A mHealth intervention to preserve and promote ideal cardiovascular health in college students: design and protocol of a cluster randomized controlled trial. Contemp Clin Trials 2020 Nov;98:106162 [FREE Full text] [doi: 10.1016/j.cct.2020.106162] [Medline: 33038506]

30. System Usability Scale (SUS). usability.gov. URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2021-01-09]
31. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
32. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
33. Epstein LJ, Kristo D, Strollo PJ, Friedman N, Malhotra A, Patil SP, Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. J Clin Sleep Med 2009 Jun 15;5(3):263-276 [FREE Full text] [Medline: 19960649]
34. Weaver TE, Maislin G, Dinges DF, Bloxham T, George CF, Greenberg H, et al. Relationship between hours of CPAP use and achieving normal levels of sleepiness and daily functioning. Sleep 2007 Jun;30(6):711-719 [FREE Full text] [doi: 10.1093/sleep/30.6.711] [Medline: 17580592]
35. Petrov ME, Hasanaj K, Hoffmann CM, Epstein DR, Krahn L, Park JG, et al. Rationale, design, and development of SleepWell24: a smartphone application to promote adherence to positive airway pressure therapy among patients with obstructive sleep apnea. Contemp Clin Trials 2020 Feb;89:105908 [FREE Full text] [doi: 10.1016/j.cct.2019.105908] [Medline: 31843639]
36. Isetta V, Torres M, González K, Ruiz C, Dalmases M, Embid C, et al. A New mHealth application to support treatment of sleep apnoea patients. J Telemed Telecare 2017 Jan;23(1):14-18. [doi: 10.1177/1357633X15621848] [Medline: 26672606]
37. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
38. Farao J, Malila B, Conrad N, Mutsvangwa T, Rangaka MX, Douglas TS. A user-centred design framework for mHealth. PLoS One 2020 Aug 19;15(8):e0237910 [FREE Full text] [doi: 10.1371/journal.pone.0237910] [Medline: 32813711]
39. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown Iii W. Comparison of a user-centered design, self-management app to existing mHealth apps for persons living with HIV. JMIR Mhealth Uhealth 2015 Sep 18;3(3):e91 [FREE Full text] [doi: 10.2196/mhealth.4882] [Medline: 26385783]
40. 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 Sep;Suppl(s2):49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
41. Bianchi MT, Alameddine Y, Mojica J. Apnea burden: efficacy versus effectiveness in patients using positive airway pressure. Sleep Med 2014 Dec;15(12):1579-1581. [doi: 10.1016/j.sleep.2014.07.015] [Medline: 25441743]

Abbreviations

AHI: apnea-hypopnea index
API: application programming interface
CPAP: continuous positive airway pressure
mHealth: mobile health
OSA: obstructive sleep apnea

Edited by G Eysenbach; submitted 12.05.22; peer-reviewed by J Bakker, S Hirshfield; comments to author 11.08.22; revised version received 07.10.22; accepted 24.10.22; published 05.12.22.

Please cite as:
Pfammatter AF, Hughes BO, Tucker B, Whitmore H, Spring B, Tasali E
The Development of a Novel mHealth Tool for Obstructive Sleep Apnea: Tracking Continuous Positive Airway Pressure Adherence as a Percentage of Time in Bed
J Med Internet Res 2022;24(12):e39489
URL: https://www.jmir.org/2022/12/e39489
doi:10.2196/39489
PMID:36469406

Original Paper

Changes in a Digital Type 2 Diabetes Self-management Intervention During National Rollout: Mixed Methods Study of Fidelity

Jack S Benton[1], BSc, MSc, PhD; Sarah Cotterill[2], BA, MSc, PhD; Rhiannon E Hawkes[1], BSc, MSc; Lisa M Miles[1], BSc, MMedSc, DrPH; David P French[1], BSc, MSc, PhD

[1]Manchester Centre for Health Psychology, Division of Psychology and Mental Health, University of Manchester, Manchester, United Kingdom

[2]Centre for Biostatistics, Division of Population Health, Health Services Research & Primary Care, University of Manchester, Manchester, United Kingdom

Corresponding Author:
Jack S Benton, BSc, MSc, PhD
Manchester Centre for Health Psychology
Division of Psychology and Mental Health
University of Manchester
Coupland 1 Building
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 7853986479
Email: jack.benton@manchester.ac.uk

Abstract

Background: "Healthy Living for People with type 2 Diabetes (HeLP-Diabetes)" was a theory-based digital self-management intervention for people with type 2 diabetes mellitus that encouraged behavior change using behavior change techniques (BCTs) and promoted self-management. HeLP-Diabetes was effective in reducing HbA1c levels in a randomized controlled trial (RCT). National Health Service (NHS) England commissioned a national rollout of HeLP-Diabetes in routine care (now called "Healthy Living"). Healthy Living presents a unique opportunity to examine the fidelity of the national rollout of an intervention originally tested in an RCT.

Objective: This research aimed to describe the Healthy Living BCT and self-management content and features of intervention delivery, compare the fidelity of Healthy Living with the original HeLP-Diabetes intervention, and explain the reasons for any fidelity drift during national rollout through qualitative interviews.

Methods: Content analysis of Healthy Living was conducted using 3 coding frameworks (objective 1): the BCT Taxonomy v1, a new coding framework for assessing self-management tasks, and the Template for Intervention Description and Replication. The extent to which BCTs and self-management tasks were included in Healthy Living was compared with published descriptions of HeLP-Diabetes (objective 2). Semistructured interviews were conducted with 9 stakeholders involved in the development of HeLP-Diabetes or Healthy Living to understand the reasons for any changes during national rollout (objective 3). Qualitative data were thematically analyzed using a modified framework approach.

Results: The content analysis identified 43 BCTs in Healthy Living. Healthy Living included all but one of the self-regulatory BCTs ("commitment") in the original HeLP-Diabetes intervention. Healthy Living was found to address all areas of self-management (medical, emotional, and role) in line with the original HeLP-Diabetes intervention. However, 2 important changes were identified. First, facilitated access by a health care professional was not implemented; interviews revealed this was because general practices had fewer resources in comparison with the RCT. Second, Healthy Living included an additional structured web-based learning curriculum that was developed by the HeLP-Diabetes team but was not included in the original RCT; interviews revealed that this was because of changes in NHS policy that encouraged referral to structured education. Interviewees described how the service provider had to reformat the content of the original HeLP-Diabetes website to make it more usable and accessible to meet the multiple digital standards required for implementation in the NHS.

Conclusions: The national rollout of Healthy Living had good fidelity to the BCT and self-management content of HeLP-Diabetes. Important changes were attributable to the challenges of scaling up a digital intervention from an RCT to a nationally implemented

intervention, mainly because of fewer resources available in practice and the length of time since the RCT. This study highlights the importance of considering implementation throughout all phases of intervention development.

(J Med Internet Res 2022;24(12):e39483) doi:10.2196/39483

KEYWORDS

type 2 diabetes; Healthy Living; digital interventions; behavior change; self-management; fidelity; implementation; mixed methods; mobile phone

Introduction

Background

Type 2 diabetes mellitus (T2DM) is one of the most common long-term conditions worldwide [1]. T2DM can lead to a range of health complications, but many of these complications can be prevented if individuals effectively self-manage their condition through healthy eating, physical activity, blood glucose monitoring, medication adherence, problem-solving skills, coping skills, and risk-reduction behaviors [2]. However, performing effective self-management is demanding and influenced by many contextual factors (eg, family, financial status, and community environment) [3], which means it can be difficult to meet the challenges of self-management without support. Self-management interventions can give people the knowledge, skills, and confidence to improve self-management through education, training, and support. Self-management interventions for people with T2DM are recommended by the UK National Institute for Health and Care Excellence for all people diagnosed with T2DM [4].

Self-management interventions for people with T2DM are typically delivered through face-to-face or group-based courses [5-7]. Although these interventions can improve clinical and psychosocial outcomes in people with T2DM [8] and are cost-effective [9], attendance can be extremely low. For example, data from the UK National Diabetes Audit suggest that just 7% of people newly diagnosed with T2DM who were offered structured education were recorded as attending within 1 year of diagnosis [10]. As an alternative, digital interventions (via digital technologies such as websites or smartphones) have the potential to be more convenient for patients as they can be delivered at scale in multiple locations, which also consumes fewer primary care resources. Mounting evidence suggests that digital self-management interventions can improve glycemic control (HbA1c) in people with T2DM [11-13].

A digital self-management intervention that has demonstrated effectiveness is Healthy Living for People With Type 2 Diabetes (HeLP-Diabetes), mainly consisting of a theory- and evidence-based website. In a randomized controlled trial (RCT) in 21 primary care practices in England, HeLP-Diabetes led to a significant, albeit modest, reduction in HbA1c levels of 0.24% (95% CI −0.44% to –0.049%; *P*=.01) at 12 months and was found to be cost-effective [14]. The HeLP-Diabetes website contained information about understanding and treating T2DM, behavior change modules, self-help tools, self-assessment quizzes, videos from people with T2DM, a moderated web-based forum, and an electronic health record. Facilitated access with a practice nurse was provided as part of the HeLP-Diabetes intervention, which consisted of an introductory training session with the practice nurse. Follow-up telephone calls were offered to support patients with using the website.

HeLP-Diabetes was originally designed as an unstructured digital intervention that patients could access without following a linear pathway, and it was this intervention that was tested in the RCT; this study focuses on this intervention. However, in 2013, general practitioners in England were offered incentives to refer people newly diagnosed with diabetes to structured education, and self-management programs were only eligible for accreditation if they followed a structured pathway with a clear curriculum and learning goals. In response to this, the HeLP-Diabetes researchers developed “HeLP-Diabetes: Starting Out”—an additional web-based structured education course based on the content of the original HeLP-Diabetes website. Previous research has tested this structured education course within 5 general practices in London in a small sample of patients (N=791) and found that there were problems with uptake and completion [15]. No studies to date have tested the effectiveness or cost-effectiveness of HeLP-Diabetes: Starting Out in a trial or assessed the fidelity of implementing this intervention in practice.

In 2019, National Health Service (NHS) England commissioned HeLP-Diabetes to be rolled out nationally in routine care under the name “Healthy Living” (Healthy Living for People With Type 2 Diabetes program). NHS England commissioned an external digital service provider to develop and offer Healthy Living as an NHS service. Figure 1 provides an overview of the development of HeLP-Diabetes, HeLP-Diabetes: Starting Out, and Healthy Living.

This study explored the fidelity of the national rollout of Healthy Living to the original HeLP-Diabetes intervention. Intervention fidelity is defined as the extent to which an intervention is delivered as intended [16]. Without good fidelity to the original HeLP-Diabetes intervention, there would be no strong justification for the implementation of Healthy Living, and reasons for intervention effectiveness would be unclear. The fidelity of diabetes self-management interventions remains largely underinvestigated [17], and fidelity evaluations are less common in routine practice than in research studies [18]. Therefore, Healthy Living presents a unique opportunity to assess the fidelity of a real-world national rollout of a digital intervention that has demonstrated effectiveness in an RCT. This study considers the extent to which Healthy Living shows fidelity to the design of the original HeLP-Diabetes intervention in relation to 3 aspects of design.

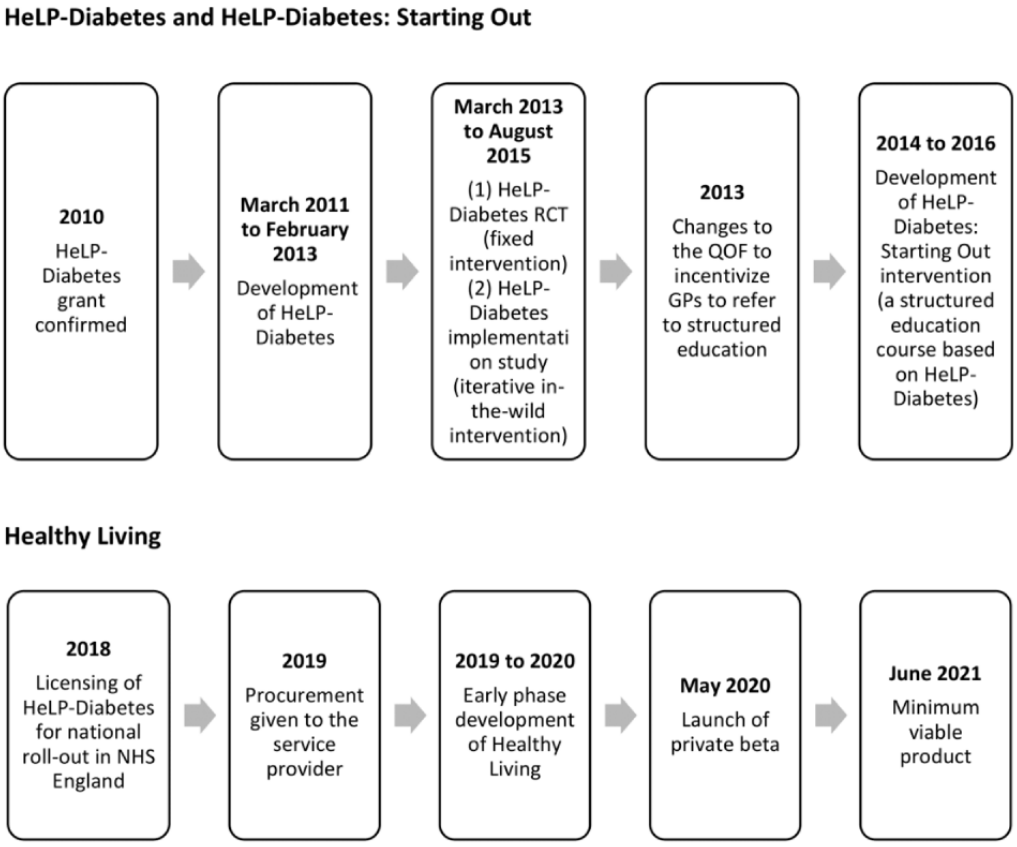
First, HeLP-Diabetes was guided by behavior change techniques (BCTs), which are the “active ingredients” of interventions that are designed to change behavior [19]. HeLP-Diabetes contained

BCTs that were likely to change important health behaviors for people with T2DM, including diet, physical activity, medication adherence, alcohol consumption, and smoking. In the HeLP-Diabetes final report [20], the researchers emphasized the importance of self-regulatory BCTs, which facilitate a negative feedback loop consisting of goal setting, recognizing inconsistencies between goals and current behavior, and developing plans to mitigate these inconsistencies [21].

Second, the self-management content in HeLP-Diabetes was guided by the Corbin and Strauss [22] model for managing a long-term condition. This model states that self-management comprises 3 types of tasks: medical management (eg, adopting healthy behaviors, working with health professionals, and keeping appointments), emotional management (managing the emotions that accompany long-term conditions), and role management (changing, creating, and maintaining new meaningful life roles, such as changes in relationships, work patterns, and day-to-day activities).

Third, the effectiveness of a digital intervention is influenced by features of intervention delivery [23]. This includes all features through which the BCT and self-management content are conveyed, such as the format, materials, intensity, tailoring, and style. Therefore, it is important to assess fidelity to the features of delivery in the original HeLP-Diabetes intervention.

Figure 1. Timeline of the intervention development since 2010. GP: general practitioner; HeLP-Diabetes: Healthy Living for People With Type 2 Diabetes; NHS: National Health Service; QOF: Quality and Outcomes Framework; RCT: randomized controlled trial.



Objectives

Thus, the objectives of this study were to (1) describe Healthy Living in terms of BCTs, self-management tasks, and features of intervention delivery; (2) compare the fidelity of these aspects with the original HeLP-Diabetes intervention; and (3) explain the reasons for any fidelity drift during national rollout.

Methods

Design

This study used a mixed methods design. A content analysis of Healthy Living was conducted using 3 coding frameworks (objective 1). The extent to which BCTs and self-management tasks were included in Healthy Living was derived and compared with published descriptions of HeLP-Diabetes (objective 2). One-to-one semistructured interviews were conducted with the key stakeholders involved in the development of HeLP-Diabetes or Healthy Living to understand the reasons for any changes during national rollout (objective 3).

Content Analysis

Coding Materials

HeLP-Diabetes Intervention

The HeLP-Diabetes website had been deleted at the time of this study, and the intervention had not been previously coded in detail for BCTs or self-management tasks. Therefore, the following publications provided the most comprehensive description of HeLP-Diabetes: (1) the HeLP-Diabetes final report [20]; (2) a journal article describing the theoretical content

of the HeLP-Diabetes intervention [24]; and (3) journal articles relating to the 3 pre-existing behavior change interventions that were integrated into the HeLP-Diabetes intervention—DownYourDrink [25,26], POWeR [27,28], and StopAdvisor [29,30].

Healthy Living

The content analysis assessed all aspects of the Healthy Living service available to users in June 2021. At this point, the website was in the "private beta" phase of service development, where a limited number of people with diabetes were invited to use the service and offer feedback to improve it [31]. At this stage, the website was classified as a "minimum viable product," meaning that all the core features were in place and unlikely to change but it was still undergoing refinement [32].

Healthy Living comprised the components outlined in Textbox 1.

Textbox 1. Components of the Healthy Living intervention.

- A website of 895 web pages containing written articles, videos, self-assessment quizzes, and tools; website content was broken down into 3 main components (refer to Multimedia Appendix 1 for screenshots):
  - "Learn": a structured curriculum where users worked through modules in a linear fashion, based on the Healthy Living for People With Type 2 Diabetes (HeLP-Diabetes): Starting Out website [23]
  - "Find answers": sections dedicated to various topics relating to type 2 diabetes mellitus where users could dip in and out of different pages and sections, based on the HeLP-Diabetes website [21]
  - "Tools": a range of "Goals" and "Tracker" tools based on the HeLP-Diabetes website
- Communication with users via email to encourage engagement

Coding Procedures

Content analysis of Healthy Living was carried out using 3 coding frameworks between June 2021 and October 2021.

BCT Content

The BCT Taxonomy v1 [19] defines 93 distinct BCTs and offers a reliable and valid method for coding the BCT content of behavior change interventions [33]. BCT coding was carried out independently by the first author, who underwent training in the use of the BCT Taxonomy v1 [34]. Coding was performed using data collection forms and coding procedures that have previously been used to code intervention design [35,36] (see Multimedia Appendix 2 for the BCT coding instructions and data collection checklist).

A second author (REH) double-coded 30 web pages of the Healthy Living website to assess the interrater reliability of BCT coding. These 30 web pages were purposively selected to ensure diversity in the type of web page (eg, written article, video, and quiz) and topics (eg, physical activity, working with diabetes, and emotional management).

The following health behaviors were coded as they were the target behaviors in the development of HeLP-Diabetes [20]: diet, physical activity, alcohol consumption, smoking, and medication adherence. Additional health behaviors that were identified in Healthy Living but were not the key target behaviors in HeLP-Diabetes were also coded (eg, sleep-related behaviors and sexual health behaviors). A new instance of a BCT was coded on commencement of a new activity (eg, a new web page or a video on a web page) or if a different health behavior was targeted (eg, diet or smoking). A new instance of a BCT was coded for the technique "Information about health consequences" when a different level of health behavior was targeted (eg, levels of the target behavior "diet" included information about carbohydrates, fats, and sugar). The number of distinct instances of BCTs on each web page was calculated.

Self-management Content

In the absence of a published coding framework for assessing self-management tasks, the authors developed a new set of coding rules. A prespecified list of self-management tasks under each of the 3 types of self-management in the Corbin and Strauss [22] model (medical, emotional, and role management) was created through team discussion (Multimedia Appendix 3). This prespecified list of self-management tasks was informed by the HeLP-Diabetes final report and additional literature on self-management in people with T2DM [37-39].

Self-management tasks were coded for each web page if at least one of the prespecified self-management tasks (Multimedia Appendix 3) was addressed. Coding was intended to assess the extent to which the intervention addressed all aspects of self-management rather than how well. Therefore, self-management tasks were coded if a task was addressed regardless of the nature of the content (eg, basic information provision, advice, and prompting of self-assessment). More than one type of self-management task could be coded on a single web page; for example, both medical and role management were coded if a web page provided information about checking blood sugar levels (medical) before driving (role).

Features of Intervention Delivery

Healthy Living was described using the Template for Intervention Description and Replication (TIDieR) framework for describing complex interventions [40]. TIDieR items (eg, materials, procedures, and modes of delivery) were extracted by the first author from the Healthy Living materials. The TIDieR description was member checked [41] by representatives from NHS England and the service provider for accuracy.

Comparison With HeLP-Diabetes

BCTs specified in the HeLP-Diabetes publications [20,24-30] were extracted into a separate data collection form for comparison with the BCTs identified in Healthy Living.

No comparison was attempted for self-management tasks as the HeLP-Diabetes publications did not provide an exhaustive description of self-management tasks to facilitate a meaningful comparison.

The TIDieR framework has previously been used to describe the HeLP-Diabetes intervention in detail [24], which enabled comparison with Healthy Living in terms of features of delivery.

Analysis

The Cohen κ coefficient [42] was used to assess interrater reliability between the 2 authors who independently coded 30 web pages for BCTs. κ values were calculated for each web page, and the mean of all 30 web pages was calculated. Any coding discrepancies were discussed between the authors until agreement was reached.

The presence and frequency of specific BCTs were compared between Healthy Living and HeLP-Diabetes to assess whether both used similar techniques to achieve behavior change. Particular attention was given to a comparison of the self-regulatory BCTs (eg, goal setting and action planning) as these were identified as important in the development of HeLP-Diabetes. The proportion of additional BCTs that were identified in Healthy Living but were not specified in HeLP-Diabetes was also calculated.

Qualitative Interviews

Sampling and Recruitment

To ensure that the interviews provided an in-depth understanding of the reasons for any changes during national rollout, a purposive sampling strategy was used to select stakeholders who had a high level of involvement in the development of HeLP-Diabetes or Healthy Living. Stakeholders involved in the development of HeLP-Diabetes were identified by emailing members of the original academic research team. Healthy Living stakeholders were sampled through discussions with NHS England and the service provider. Additional stakeholders were identified via snowball sampling. Views were sought from stakeholders in a range of professional roles, including academic researchers, digital content developers, and program managers. As the population who could usefully comment on the intervention development process was small, the sample interviewed was small; hence, the population of interest was exhausted through the sampling strategy used.

Procedure and Materials

Topic guides were used to organize the semistructured interviews, with open-ended questions and additional probes. Topics covered participants' knowledge and understanding of the HeLP-Diabetes intervention content and features of delivery, how Healthy Living had changed from the original intervention, and the reasons for any changes. All interviews were audio recorded using an encrypted audio recording device following full verbal or written consent.

Analysis

Audio recordings were transcribed verbatim by an external transcription company and thematically analyzed using a modified framework approach [43]. The first author read and reread transcripts, noting key ideas, and then independently coded the first 4 interviews, generating a combination of data-driven and a priori thematic codes. The data-driven codes were generated inductively from the data alone without reference to other sources. The a priori codes were based on the author's understanding of what had changed from the original HeLP-Diabetes intervention and participants' explicit rationale for any changes. The codes were summarized into initial themes, which were refined through discussion between 2 authors (JSB and DPF). These themes were then systematically applied to the remaining interviews, with ongoing adaptations until no new themes emerged. Themes were discussed at length between all authors until an agreement was reached on the final themes. The data were coded electronically using NVivo (version 12; QSR International).

Ethics Approval

The wider program of research of which this study is a part was reviewed and approved by the Yorkshire and the Humber – Leeds West NHS Research Ethics Committee (reference 20/YH/0250, September 29, 2020). Full verbal or written consent was obtained from all interview participants. Interview data were anonymized at the point of transcription.

Results

BCT Content

Interrater Reliability

The mean κ value for the coding of BCTs was 0.80 (SD 0.31), thus demonstrating strong agreement [42] between coders before resolving discrepancies (see Multimedia Appendix 4 for all κ values).

Healthy Living

Table 1 shows the number of distinct instances of BCTs identified in Healthy Living. There were 43 BCTs identified in Healthy Living. The most common BCT was information about health consequences (849/2088, 40.7%). Diet was the behavior most commonly targeted by BCTs (targeted by 659/2088, 31.6% of all BCTs), followed by physical activity (471/2088, 22.6%) and medication adherence (454/2088, 21.7%). Multimedia Appendix 5 shows the frequency of BCTs by each health behavior.

Table 1. Instances of behavior change techniques (BCTs) in Healthy Living and how this compares with Healthy Living for People With Type 2 Diabetes (HeLP-Diabetes).

BCTs	Healthy Living components, n (%)				Healthy Living (all 895 pages; n=2088), n (%)	Specified in HeLP-Diabetes?
	"Learn" (273 pages; n=568)	"Find answers" (583 pages; n=1401)	"Tools" (39 pages; n=110)	Email communication (20 messages; n=9)		
Self-regulatory BCTs						
Problem-solving	10 (1.8)	46 (3.3)	14 (12.7)	0 (0)	70 (3.4)	Yes
Self-monitoring of outcomes of behavior	17 (3)	43 (3.1)	6 (5.5)	2 (22.2)	68 (3.3)	Yes
Goal setting (behavior)[a]	32 (5.6)	18 (1.3)	9 (8.2)	1 (11.1)	60 (2.9)	Yes
Review behavior goals	17 (3)	2 (0.1)	23 (20.9)	0 (0)	42 (2)	Yes
Action planning	4 (0.7)	14 (1)	16 (14.5)	0 (0)	34 (1.6)	Yes
Self-monitoring of behavior	1 (0.2)	21 (1.5)	4 (3.6)	2 (22.2)	28 (1.3)	Yes
Goal setting (outcome)[a]	7 (1.2)	4 (0.3)	3 (2.7)	1 (11.1)	15 (0.7)	No
Feedback on behavior	9 (1.6)	3 (0.2)	2 (1.8)	0 (0)	14 (0.7)	Yes
Biofeedback[a]	4 (0.7)	9 (0.6)	0 (0)	0 (0)	13 (0.6)	No
Review outcome goals	3 (0.5)	2 (0.1)	5 (4.5)	1 (11.1)	11 (0.5)	No
Feedback on outcomes of behavior	0 (0)	0 (0)	3 (2.7)	0 (0)	3 (0.1)	Yes
Commitment	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Yes
Other BCTs						
Information about health consequences	232 (40.8)	617 (44)	0 (0)	0 (0)	849 (40.7)	Yes
Social support (unspecified)	73 (12.9)	133 (9.5)	2 (1.8)	2 (22.2)	210 (10.1)	Yes
Information about emotional consequences	41 (7.2)	68 (4.9)	0 (0)	0 (0)	109 (5.2)	Yes
Behavior substitution	33 (5.8)	57 (4.1)	8 (7.3)	0 (0)	98 (4.7)	No
Credible source	22 (3.9)	21 (1.5)	0 (0)	0 (0)	43 (2.1)	No
Instruction on how to perform the behavior	0 (0)	41 (2.9)	0 (0)	0 (0)	41 (2)	Yes
Social support (practical)	12 (2.1)	29 (2.1)	0 (0)	0 (0)	41 (2)	No
Information about social and environmental consequences	8 (1.4)	31 (2.2)	0 (0)	0 (0)	39 (1.9)	No
Information about antecedents	2 (0.4)	36 (2.6)	0 (0)	0 (0)	38 (1.8)	Yes
Behavioral practice or rehearsal	4 (0.7)	32 (2.3)	0 (0)	0 (0)	36 (1.7)	Yes
Demonstration of the behavior	3 (0.5)	32 (2.3)	0 (0)	0 (0)	35 (1.7)	Yes
Adding objects to the environment	4 (0.7)	24 (1.7)	4 (3.6)	0 (0)	32 (1.5)	Yes
Reduce negative emotions	5 (0.9)	16 (1.1)	0 (0)	0 (0)	21 (1)	Yes
Restructuring the physical environment	6 (1.1)	13 (0.9)	2 (1.8)	0 (0)	21 (1)	Yes
Restructuring the social environment	2 (0.4)	13 (0.9)	4 (3.6)	0 (0)	19 (0.9)	No
Social support (emotional)	4 (0.7)	13 (0.9)	0 (0)	0 (0)	17 (0.8)	No
Prompts and cues	3 (0.5)	10 (0.7)	2 (1.8)	0 (0)	15 (0.7)	Yes

BCTs	Healthy Living components, n (%)				Healthy Living (all 895 pages; n=2088), n (%)	Specified in HeLP-Diabetes?
	"Learn" (273 pages; n=568)	"Find answers" (583 pages; n=1401)	"Tools" (39 pages; n=110)	Email communication (20 messages; n=9)		
Increase positive emotions[b]	2 (0.4)	7 (0.5)	0 (0)	0 (0)	9 (0.4)	No
Nonspecific reward	0 (0)	9 (0.6)	0 (0)	0 (0)	9 (0.4)	No
Self-reward	0 (0)	9 (0.6)	0 (0)	0 (0)	9 (0.4)	Yes
Avoidance or reducing exposure to cues for the behavior	0 (0)	5 (0.4)	2 (1.8)	0 (0)	7 (0.3)	Yes
Distraction	0 (0)	7 (0.5)	0 (0)	0 (0)	7 (0.3)	No
Salience of consequences	5 (0.9)	2 (0.1)	0 (0)	0 (0)	7 (0.3)	No
Pros and cons	2 (0.4)	3 (0.2)	0 (0)	0 (0)	5 (0.2)	Yes
Pharmacological support	0 (0)	3 (0.2)	0 (0)	0 (0)	3 (0.1)	Yes
Material incentive (behavior)	0 (0)	2 (0.1)	0 (0)	0 (0)	2 (0.1)	No
Self-incentive	0 (0)	2 (0.1)	0 (0)	0 (0)	2 (0.1)	No
Social incentive	0 (0)	2 (0.1)	0 (0)	0 (0)	2 (0.1)	No
Mental rehearsal of successful performance	0 (0)	1 (0.1)	0 (0)	0 (0)	1 (0)	No
Reattribution	0 (0)	1 (0.1)	0 (0)	0 (0)	1 (0)	No
Salience of behavior[c]	0 (0)	0 (0)	1 (0.9)	0 (0)	1 (0)	No
Social reward	1 (0.2)	0 (0)	0 (0)	0 (0)	1 (0)	Yes
Conserving mental resources	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Yes
Graded tasks	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Yes
Identity associated with changed behavior	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Yes
Identification of self as role model	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Yes
Information about others' approval	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Yes
Self-talk	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Yes
Social comparison	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	Yes

[a]Includes BCTs that were "prompted" rather than directly delivered regardless of whether the BCT Taxonomy v1 (BCTTv1) definition specified that the BCT can be prompted. For example, the definition in the BCTTv1 for goal setting (behavior) was as follows: "Set or agree on a goal defined in terms of the behavior to be achieved." However, this BCT was coded when patients were prompted to set a goal elsewhere as part of the intervention (eg, by clicking on the "Physical activity goal" tool). Refer to Multimedia Appendix 3 for further details on BCT coding procedures.

[b]Increase positive emotions is not listed in the BCTTv1 but was noted by the authors for inclusion in the next version of the taxonomy.

[c]Salience of behaviors is not listed in the BCTTv1 but has been identified as a new BCT by the authors of this paper.

Comparison With HeLP-Diabetes

There were 32 BCTs specified in the HeLP-Diabetes intervention (Table 1). Healthy Living included 75% (24/32) of these BCTs (Table 1). There were an additional 37% (19/51) of BCTs that were identified in Healthy Living but were not specified in HeLP-Diabetes (Table 1).

All but one of the self-regulatory BCTs ("commitment") specified in HeLP-Diabetes were identified in Healthy Living, including problem-solving (70/2088, 3.4%), self-monitoring of outcomes of behavior (68/2088, 3.3%), self-monitoring of behavior (28/2088, 1.3%), goal setting (for behavior; 60/2088, 2.9%), review behavior goals (42/2088, 2%), action planning (34/2088, 1.6%), feedback on behavior (14/2088, 0.7%), and feedback on outcomes of behavior (3/2088, 0.1%; Table 1). There were also other self-regulatory BCTs identified in Healthy Living that were not explicitly specified in the HeLP-Diabetes final report, including goal setting (for outcomes; 15/2088, 0.7%), biofeedback (13/2088, 0.6%), and review outcome goals (11/2088, 0.5%; Table 1).

Self-management Content

The number of distinct instances of self-management tasks that were addressed in Healthy Living is summarized in Table 2. Most of the Healthy Living intervention addressed medical management tasks (821/895, 91.7% of all web pages). Emotional management tasks were addressed in 35.4% (317/895) of all web pages, and role management tasks were addressed in 30.9% (277/895) of all web pages.

Table 2. Instances of self-management tasks addressed in Healthy Living.

Self-management tasks	Healthy Living components, n (%)				Healthy Living (all 895 pages; n=1415), n (%)
	"Learn" (273 pages; n=392)	"Find answers" (583 pages; n=975)	"Tools" (39 pages; n=41)	Email communication (20 messages; n=7)	
Medical	211 (53.8)	565 (57.9)	38 (92.7)	7 (100)	821 (58)
Emotional	103 (26.3)	211 (21.6)	3 (7.3)	0 (0)	317 (22.4)
Role	78 (19.9)	199 (20.4)	0 (0)	0 (0)	277 (19.6)

Features of Intervention Delivery

Healthy Living

Multimedia Appendix 6 [14,15,40,44-51] contains a detailed description of Healthy Living using the TIDieR framework. In brief, Healthy Living was a free digital NHS service for people living with T2DM developed for use on a range of digital devices (ie, smartphones, desktops, and tablets). The website contained 895 web pages, including information about what T2DM is, its causes, and how it can be managed and treated; behavioral advice on diet, physical activity, alcohol, smoking, and medication adherence; and emotional and practical support. There were tools for users to set and review goals, make action plans, and self-monitor. Healthy Living was intended for people diagnosed with T2DM in England, carers, and health care professionals. The service was available by self-referral, although there were ongoing plans to develop primary care referral once beta testing was complete and once health services opened up after the COVID-19 pandemic. Technical support was provided, but there were no health care professionals to support the use of the website or behavior change.

Comparison With HeLP-Diabetes

When comparing Healthy Living with HeLP-Diabetes, there were similarities in relation to the written content and topics (eg, understanding diabetes and preexisting interventions) and types of materials (eg, articles, videos, and tools).

However, there were a number of important differences (summarized in Table 3). The HeLP-Diabetes intervention offered patients facilitated access through a 5- to 10-minute onboarding process with a health care professional in primary care. There were ongoing plans to develop primary care and community hub referral pathways for Healthy Living, but there were no plans to include facilitated access by a health care professional. The Healthy Living website also included an additional structured curriculum, which was based on the HeLP-Diabetes: Starting Out course developed after the RCT by the HeLP-Diabetes team [15]. However, this structured curriculum was not included in the original HeLP-Diabetes intervention tested in the RCT. Features of the HeLP-Diabetes website that were not retained in Healthy Living included a moderated forum (where users could interact with other users and ask health professionals questions) and a health record (where users could record and keep track of appointments and tests with health care professionals).

Table 3. Summary of differences in the features of delivery between Healthy Living for People With Type 2 Diabetes (HeLP-Diabetes), HeLP-Diabetes: Starting Out, and Healthy Living.

Feature	"HeLP-Diabetes" (RCT[a] version)	"HeLP-Diabetes: Starting Out"	"Healthy Living"
Registration	Facilitated access by a practice nurse through a 5- to 10-minute appointment	Self–sign-up, with optional telephone support for those who had difficulty registering or using the website	Self–sign-up (plans to develop referrals through primary care and community hubs but no facilitated access)
Size of website	8 sections, 560 pages	5 sections, with selected content from HeLP-Diabetes; users also had access to the HeLP-Diabetes website via a common home page	3 sections, 895 pages
How the intervention was delivered	Nonlinear—users could access any part of the website and dip in and out as they pleased in any order	Linear and nonlinear—users worked through modules one by one but also had access to the nonlinear component, where they could dip in and out of sections as they pleased in any order	Linear and nonlinear—users worked through modules one by one but also had access to the nonlinear component, where they could dip in and out of sections as they pleased in any order
Curriculum	No curriculum—users could choose which topics to access depending on interest	Structured curriculum—users had access to a series of modules that could be worked through in a spiral fashion	Structured curriculum—users had access to a series of modules that could be worked through in a spiral fashion
Forum and help	There was a moderated web-based forum where users could interact with other users, and there was an "Ask the Expert" option where users could ask health professionals questions; additional resources included local resources tailored to the CCG[b] and a list of frequently asked questions	Users had access to the HeLP-Diabetes nonlinear website, where they could access the web-based forum, "Ask the Expert," and all the additional tailored resources	No moderated web-based forum or tailored support
Health record	Users could record and keep track of appointments with health care professionals and of the results of tests used to monitor diabetes (eg, HbA1c, blood pressure, cholesterol level, and kidney and liver function)	Users had access to the HeLP-Diabetes website, where they could record and keep track of health care appointments and test results	No health record, but an HbA1c tracker was offered
Physical materials	Practice nurses were provided with training leaflets for facilitated access; information leaflets for patients	No physical materials were offered	Information leaflets for patients in primary care
Engagement	Email, SMS text message reminders, and follow-up phone calls	Emails only	Emails only

[a]RCT: randomized controlled trial.

[b]CCG: clinical commissioning group.

Qualitative Interviews

Participants

A total of 9 participants were interviewed, including stakeholders from the HeLP-Diabetes academic team (n=5, 56%), the NHS England diabetes program team (n=1, 11%), and the service provider (n=3, 33%). Most interviews (8/9, 89%) were carried out between May 2021 and June 2021, and 11% (1/9) of the interviews were conducted in September 2020. Interviews lasted between 46 and 102 (mean 70) minutes.

A total of 3 overarching themes were identified: changes because of scalability issues, changes to improve user engagement and outcomes, and digital development challenges.

Theme 1: Scalability Issues

The NHS England interviewee indicated that some features from the original HeLP-Diabetes intervention were not implemented as there were fewer resources (staff and time) within general practices compared with the RCT to deliver these features at scale. Issues of scalability predominantly related to features that required additional support from health care professionals.

Subtheme 1.1: Fewer Resources Available to Provide Facilitated Access

The NHS England interviewee reported that health care professionals would be unable to spend time supporting patients to register and onboard them to the intervention because of capacity issues and time constraints that were not necessarily an issue in the RCT. Interviewees from the HeLP-Diabetes team said that they encountered similar difficulties in their implementation research, which was conducted in parallel to the RCT:

> *We already felt that it wasn't scalable and that we were already very aware that it would have been a big request of general practices to spend time onboarding the user, when we were already aware*

> *of some of the capacity issues and time constraints within an annual review already...We worked with the early engagement areas to sort of validate that understanding and it was a universal confirmation that they would not implement if we maintained that onboarding mechanism.* [Participant 4, NHS England]

When asked about the implications of removing facilitated access to the intervention, the NHS England interviewee believed that it would not necessarily have a negative impact on patients as the program would still involve referral from a health care professional, thus retaining the trust associated with health care professional recommendations. This interviewee also believed that removing facilitated access increased buy-in from health care professionals, which was important to encourage patient referrals to the intervention.

Although acknowledging implementation issues, interviewees from the HeLP-Diabetes team suggested that facilitated access was an important component of the original intervention to support people with lower education or limited computer skills:

> *Our ideal model was facilitated access. Someone from the surgery, it didn't have to a doctor, but maybe a nurse or someone like that could sign you up thereby, signposting that this was a recommended intervention, but also help overcome any initial inertia of a digital device, showing people around, this is how you can log on, this is your password.* [Participant 2, HeLP-Diabetes]

Subtheme 1.2: Omission of the Moderated Web-Based Forum

The service provider was informed by the HeLP-Diabetes team that the uptake of the forum was quite low in the RCT, in part because the RCT did not "reach the critical mass of users to populate the forum" (participant 1, HeLP-Diabetes). As a result, participants generally reported that the moderated forum was not perceived to be an integral feature of the intervention. Therefore, the service provider preferred not to invest substantial staff time in the moderation of a web-based forum, which had been underused in the RCT, and this was supported by NHS England:

> *The decision was made to drop that particular feature of the programme [the moderated forum]. So, it was obviously partly because we had heard from the HeLP team that the uptake had not been great, and then there was also not the resource to support it in the way it would need.* [Participant 9, service provider]

Subtheme 1.3: Encouraging Users to Seek External Support

The moderated forum in the original HeLP-Diabetes website included an "Ask the expert" option, where users could submit questions to a health care professional as "a way of getting help and advice" (participant 3, HeLP-Diabetes). However, the interviewee from NHS England believed that this was not a scalable option for a national program because of limited resources. Furthermore, providing clinical advice would require access to patient medical records and, without this, the service provider would only be able to provide generic advice:

> *['Ask the expert'] wasn't a scalable option for the programme...There's information for a user to self-manage, but there is clear direction throughout the programme that if they need something specific for their self-management, that they need to speak to their own healthcare professional.* [Participant 4, NHS England]

Theme 2: Improving User Engagement and Outcomes

The service provider believed that the original HeLP-Diabetes website needed modifying to improve engagement and outcomes for users.

Subtheme 2.1: Perceived Importance of the Structured Curriculum

On the basis of their experiences from other projects, the service provider believed that the structured curriculum ("Learn") provided the most effective way of improving patient outcomes and monitoring patients' progress. As a result, they perceived the structured curriculum "as the core content" (participant 7, service provider) and wanted most patients to use the structured curriculum, although they acknowledged that some users would prefer to engage in free exploration that was offered in the "Find answers" component of Healthy Living:

> *We are quite geared towards encouraging people to the Learn Journey because it is a structured programme, it quite often will get better results. Like we've seen in other programmes, if someone completes a certain percentage of a learning programme, they're more likely to achieve the outcomes and the goals that they're setting alongside it.* [Participant 9, service provider]

When originally designing the intervention, interviewees from the HeLP-Diabetes team expressed difficulties in grappling with the decision of whether to include a structured curriculum. Even though the evidence base suggested that a structured curriculum was more likely to be effective, participants' main concern was that it was difficult to get users to complete an entire curriculum, especially without additional support or encouragement. Interviewees from the HeLP-Diabetes team explained that patients in their qualitative research said that they would prefer to have access to self-management information as and when they needed it rather than having the burden of completing modules in a prescribed manner; this was a key factor in not including a structured curriculum in the original HeLP-Diabetes intervention:

> *There was definitely a big debate in it, because I think in the literature there was some evidence to say if it's structured it's more likely to be effective, but in all of our qualitative work, people didn't want it like that. But I know we did have a bit of back and forth, but we were mainly led from our work with people that were going to use it, who just sort of said that that would really put them off using it full stop, if it was that sort of more structured, and that they felt that they just wanted to come and be able to dip in and dip out, search for things, and use bits from websites*

that they felt they needed in that moment. [Participant 3, HeLP-Diabetes]

Subtheme 2.2: User Research

An independent user research organization was commissioned by NHS England to conduct user research to inform the development of Healthy Living. This was required as part of the UK Government Digital Service Standard [44] that was originally published in 2016. This user research was perceived as having a positive impact on user engagement, especially as the original HeLP-Diabetes website was perceived as "overwhelming" (participant 4, NHS England) and "visually cluttered" (participant 9, service provider). Interviewees from the service provider reported examples in which this user research led to modifications of the original HeLP-Diabetes website, often because of changes in how people now use digital technology:

> *The medication tracker was not a particularly popular one [from user research]. I think, if I remember rightly, I think in the time that the HeLP programme was live, people have become a lot more dependent on a mobile phone for their reminders and prompts and things like that, and the notion of using a third-party website to support that was perhaps a bit less attractive to people by that point.* [Participant 9, service provider]

Subtheme 2.3: Data-Driven Optimization of the Intervention

Participants recognized the potential of using data analytics to assess and iteratively modify the website to improve user engagement over time. The service provider described how they planned to continually use real-time data to improve the website as they did not perceive the current website as the definitive version:

> *You also have real time analytics, so that you can then start to evolve that programme in real time to optimise retention, completion outcomes.* [Participant 6, service provider]

Theme 3: Digital Development Challenges

The service provider experienced a number of digital challenges when developing the original website.

Subtheme 3.1: Adhering to Digital Standards

The service provider emphasized the challenges of adhering to digital standards, including the UK Government Digital Service Standard [44], Web Content Accessibility Guidelines [45], Digital Technology Assessment Criteria [46], and NHS Digital content [47] and style guidelines [48]. The service provider had to reformat a lot of the original content to abide by these standards, which included "making the content more readable for the average reader" (participant 4, NHS England), "reducing some of the repetition" (participant 9, service provider), and "reformatting the way that the information is provided so that it is in smaller chunks" (participant 4, NHS England). This process of adapting the content to meet digital standards was a new experience for the service provider and was perceived as arduous, especially given the sheer amount of content on the original website. Despite these challenges, interviewees from the service provider were pleased with the end result:

> *We've certainly had quite a steep learning curve around the GDS [Government Digital Service] assessment process that sits alongside the work, that was kind of a new area for us, and I think for many of the NHS team as well...I would say we've learned an awful lot in the process of this project around GDS in itself, but I think we're at a point now where what we have is a good product.* [Participant 9, service provider]

Subtheme 3.2: Lack of Iterative Development Over Time

The service provider said that the HeLP-Diabetes website had not undergone any iterative development for a number of years, which was an issue owing to the significant level of technological evolution in the time since the original website was developed. The service provider felt that they had to significantly update the website in line with current user expectations of a digital service. Interviewees from the HeLP-Diabetes team acknowledged that they had limited internal web development expertise, which was problematic for developing the website and keeping it up-to-date with advances in technology:

> *As a research project that's fundamentally not based in tech, an outside company came in and did the tech for us. But I think that was a real limitation in terms of keeping up to date with the tech and helping the intervention adapt change to the different ways in which people engage with physical content.* [Participant 2, HeLP-Diabetes]

Subtheme 3.3: Reverse Engineering the Original Website

The service provider explained that they did not receive any documents to help them build the original website, such as a site map or a master file of content, which is what would usually happen in other similar digital projects that they had worked on. This meant that they were forced to spend a lot of time working out what the original content was and how it was structured to "reverse engineer the website" (participant 6, service provider). An interviewee from the service provider highlighted the difficulties of translating the underlying theory and BCTs into a digital service:

> *People talk a lot about behavioural frameworks and behaviour change techniques. And then, they talk very little about how they've operationalised them in a digital service. And I think there's a big gap there, because it's easy to write a behavioural framework, it's hard to show how it works in the digital intervention.* [Participant 6, service provider]

Discussion

Summary of Principal Findings

The national rollout of Healthy Living included all but one ("commitment") of the self-regulatory BCTs that were specified in the original HeLP-Diabetes intervention, including goal setting, self-monitoring, and problem-solving. Healthy Living

predominantly addressed medical self-management tasks (821/895, 91.7% of web pages) but also addressed emotional (317/895, 35.4% of web pages) and role (277/895, 30.9% of web pages) self-management tasks. Therefore, the national rollout of Healthy Living had good fidelity to the BCT and self-management content of HeLP-Diabetes.

However, there were a number of changes to features of delivery during the national rollout, two of which are most noteworthy. First, the Healthy Living service included an additional structured learning curriculum that was developed after the RCT by the HeLP-Diabetes team but was not part of the HeLP-Diabetes intervention tested in the RCT. Second, Healthy Living did not implement features that required health care professional support as NHS England believed that they were not scalable. The interviewees described how the service provider had to substantially reformat the content of the original HeLP-Diabetes website to make it more usable and accessible, which was a requirement to meet digital standards to allow the intervention to be scaled up for national implementation in the NHS.

Strengths and Limitations

This fidelity analysis used 3 coding frameworks to assess every page of the Healthy Living website and all email communications offered to users, thus providing a comprehensive fidelity assessment. The authors developed a bespoke framework to assess self-management tasks as existing taxonomies were insufficient to code other aspects of interventions beyond behavior change; to the authors' knowledge, this study is the first to quantify the self-management content of an intervention. Further work is needed to develop this self-management coding framework to ensure applicability for a broad range of interventions and chronic conditions and validate the framework. Although the qualitative sample was small, there were only a limited number of stakeholders with a high level of involvement in the intervention development process (ie, we included all the population of interest). A further strength is that this study was conducted independently of those involved in the development of the intervention, which is rare in previous fidelity assessments [52].

Nevertheless, there are limitations to consider. This study was conducted at a relatively early phase in the development of Healthy Living, so there may be further changes to the website that would potentially alter these findings. This is a common challenge when assessing the fidelity of digital interventions because of the fast-moving pace of digital technology [53]. This evaluation was also conducted before the national implementation of primary care referral pathways into Healthy Living, which was delayed because of disruptions caused by the COVID-19 pandemic. However, as of September 2021, NHS England and the service provider had no plans to provide facilitated access or make major changes to the website content, so it is unlikely that conducting this evaluation when primary care referrals are implemented will alter the conclusions. Finally, the authors could not precisely compare the number of instances of BCTs and self-management tasks between Healthy Living and HeLP-Diabetes as the original HeLP-Diabetes website was not available to us, having been discontinued at the time of evaluation. Not having access to the original intervention website is an example of one of the many challenges of a fidelity evaluation conducted by a research team that is independent of those who developed the original intervention. Instead, we relied on published papers relating to HeLP-Diabetes, which described the BCT and self-management task content in great detail but did not specify where on the website they occurred or how often.

Comparison With Prior Work

To the authors' knowledge, this is the first in-depth fidelity assessment of a nationally implemented digital diabetes self-management intervention. The findings reported in this paper are in line with research assessing the fidelity of design of the NHS Digital Diabetes Prevention Programme (NHS-DDPP), a behavioral intervention for people identified as at high risk of developing T2DM. The researchers found that 85% of the BCTs outlined in the NHS-DDPP specification (which emphasized the importance of self-regulatory BCTs) were included in the service providers' intervention plans [54]. In contrast, a similar evaluation of the face-to-face version of the NHS Diabetes Prevention Program found that only 37% of specified BCTs were delivered during the program [55], and some core self-regulatory BCTs were underdelivered in the observed sessions, such as goal setting, which was delivered in 52.5% of sessions [56]. Healthy Living compares favorably with these other interventions, with 75% (24/32) of the BCTs specified in HeLP-Diabetes identified in Healthy Living, including all but one of the self-regulatory BCTs (Table 1). An explanation for the 25% (8/32) of BCTs missing in Healthy Living may be that information on the BCT content was distributed across multiple documents and was not collated for the service provider; this contrasts with the NHS Diabetes Prevention Programme and NHS-DDPP, where there were prespecified lists of BCTs that were stipulated to be delivered. The high fidelity of BCTs in digital interventions compared with face-to-face interventions is in line with the literature, which suggests that digital interventions may achieve higher fidelity as they do not rely on human delivery [57].

There were 46% (43/93) of possible BCTs offered in the Healthy Living service, which is high compared with other digital self-management interventions. For instance, a systematic review of 8 digital self-management interventions for people with T2DM found that the highest number of BCTs in an intervention was 14 [58]. This is important given that digital behavior change interventions that use more BCTs have been found to have larger effect sizes compared with interventions that use fewer BCTs [59]. Healthy Living also compared favorably with most consumer-facing smartphone apps, which have been found to implement a very limited number of BCTs [60] and often lack firm grounding in theory or evidence [61].

It is important that Healthy Living addressed all aspects of self-management, including emotional management, as previous intervention research suggests that addressing the emotional and psychological aspects of T2DM can reduce diabetes distress and improve HbA1c levels [62-65]. Previous digital interventions for people with T2DM have predominantly focused on providing information and behavior change support but less

so on emotional self-management [66-69]. Hence, there have been calls for more emotional and psychological support to be embedded within routine diabetes care [70,71].

Implications for Healthy Living

Facilitated access was not implemented in Healthy Living as there were fewer resources within the NHS than in the RCT to provide this support on a national scale. However, the HeLP-Diabetes team believed that facilitated access was important to encourage uptake, especially for those who have lower levels of education. This belief may stem from the HeLP-Diabetes implementation research, which found that "the self-sign-up model was associated with users who were better educated and had rated their computer skills as advanced" [20] when compared with the version of HeLP-Diabetes that retained facilitated access. Additional resources may need to be allocated to those Healthy Living users who might benefit from extra assistance to sign up or use the service to ensure that the intervention does not inadvertently contribute to a widening of health inequalities.

NHS policy recommends that all diabetes self-management education interventions contain a structured curriculum with clear learning objectives, which meant that a structured learning curriculum was required for national implementation. However, the stakeholder interviews in this study suggested that there is ambivalence as to whether having a structured curriculum would be more effective than a website with no structured curriculum. The evidence generally indicates that structured education interventions are effective but only if patients sufficiently engage with the intervention [72], something that is difficult to achieve for diabetes self-management education [15]. Given the low uptake previously observed in the HeLP-Diabetes: Starting Out intervention [15], improving engagement with Healthy Living will be critical for intervention effectiveness as the frequency and intensity of digital intervention use are thought to be important in achieving desired outcomes [72].

Implications for Practice

The important changes from the original HeLP-Diabetes intervention were associated with the implementation challenges of going from an RCT to a scaled-up national program. The service specification from NHS England indicated that the service provider should aim to retain fidelity to the original intervention approach that has been evidenced in the HeLP-Diabetes RCT. Although expecting perfect fidelity is unrealistic when moving from controlled to real-world settings [73], there were 2 main problems associated with this approach of aiming to retain fidelity to the intervention from the RCT for the national rollout of Healthy Living.

First, the original RCT had significant dedicated resources that enabled a more intensive intervention through the use of dedicated health care professional support, for example, through facilitated access to support the user registration process. It was clear from the interviews that providing this level of intensity was not feasible in a scaled-up national program. The HeLP-Diabetes team originally identified this as a potential issue during their implementation research but, as this research was conducted in parallel to the RCT, it did not inform the intervention design in the RCT.

The second problem was the sheer length of time from when the HeLP-Diabetes intervention underwent testing in a trial in 2013 to the subsequent procurement of Healthy Living in 2019. This meant that, by the time the website from the RCT was due to be rolled out, it required adaptation to be consistent with the new clinical and technological environment in which it was being implemented, including changes in policy, advances in digital technology, and new standards for providing digital services.

This highlights the importance of considering implementation challenges at earlier phases of intervention development to reduce the level of adaptation necessary for scaling up an intervention to real-world contexts (often called a "scale-up penalty" [73]). Addressing this issue is likely to require a shift in the way that academic health research is funded as funding is often focused on commissioning research on effectiveness rather than on the implementation stages of digital development, both of which are necessary to create effective digital interventions that are implementable outside the context of clinical trials [74]. Health funding needs to accommodate faster and more efficient methods of evaluation that enable the iterative development of digital interventions across their life cycles [75,76], such as the Multiphase Optimization Strategy [77]. There appears to be growing recognition of this. For example, the National Institute of Health and Care Excellence recently published an Evidence Standards Framework for Digital Health Technologies to develop standards that ensure that new digital technologies are not just clinically effective and cost-effective but also enable a more dynamic approach to digital development and delivery [78]. As innovation in digital technology becomes increasingly rapid and as the NHS becomes increasingly digital, a more flexible approach to the way research is funded and conducted will become even more important.

This study highlighted the importance of multidisciplinary teams during the intervention development process and throughout implementation. Although HeLP-Diabetes had multidisciplinary working groups of service users, clinicians, and researchers, the internal HeLP-Diabetes academic team did not have the professional digital development and design knowledge required to iteratively develop a high-quality product that met evolving user expectations and digital standards. Similarly, during national implementation, the service provider reported difficulties in translating the underlying theoretical models into a digital program. Effective operationalization of BCTs to a digital context is currently underdeveloped within academic research [79], so it is unsurprising that the service provider experienced difficulties in translating BCTs into digital content. Addressing the barriers to multisectoral collaborations in digital health intervention research is necessary to ensure that expertise in health and digital software development is integrated across all stages of development, evaluation, and implementation of digital health interventions [74].

Future Research

The authors of this study are conducting other streams of work in relation to the fidelity of the Healthy Living service, which

will involve analyzing use data to assess how much of the intervention content is actually engaged with by users, as well as interviews with Healthy Living users to explore how the intervention is understood and experienced. This future research will help address some of the uncertainties identified in this study, such as the impacts of the 2 important changes in Healthy Living (lack of facilitated access and inclusion of structured education) on user engagement and experiences. This future research will also look at the extent to which Healthy Living users need tailored self-management support to help their own individual needs given that evidence suggests that self-management needs to be orientated to a person's individual needs [80].

Conclusions

This mixed methods study found that the national rollout of Healthy Living had good fidelity to the BCT and self-management content of HeLP-Diabetes. However, this study identified important changes that were attributable to the challenges of scaling up a digital intervention from an RCT to a nationally implemented intervention, mainly because of fewer resources available in practice and the length of time since the RCT. This study demonstrates the importance of considering implementation throughout all phases of intervention development and testing to reduce the level of adaptation necessary for scaling up an intervention to real-world contexts. Greater collaboration between academic researchers and digital development experts is needed to produce evidence- and theory-informed digital health interventions that are usable and accessible enough to meet digital standards.

Acknowledgments

The authors would like to thank National Health Service England and the service provider for facilitating this research at all stages, including providing materials for coding and assisting in the recruitment of stakeholders for the interviews. The authors are grateful to all the participants who took part in the interviews. They would also like to thank Dr Brian McMillan, Professor Rachel Elliott, Dr Lisa Brunton, and Professor Martin Rutter from the Healthy Living Diabetes - Long-term Independent National Evaluation research team, who provided valuable feedback during manuscript preparation. This paper reports independent research funded by the National Institute for Health and Care Research (Policy Research Programme, Healthy Living Diabetes - Long-term Independent National Evaluation, NIHR200933). The views expressed in this presentation are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

Data Availability

Some data sets are available from the corresponding author upon reasonable request, although those making the request will need to obtain the explicit permission of the relevant provider organizations.

Authors' Contributions

DPF and SC secured funding as part of the wider Healthy Living Diabetes - Long-term Independent National Evaluation project. DPF designed and supervised the study. JSB collected and analyzed the data and prepared the manuscript. LMM conducted one of the interviews. REH assisted with coding of behavior change techniques. SC, REH, LMM, and DPF assisted with the qualitative analyses and helped draft the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the 3 main components of the Healthy Living website.
[DOCX File , 441 KB - jmir_v24i12e39483_app1.docx]

Multimedia Appendix 2
Procedures for behavior change technique coding.
[DOCX File , 34 KB - jmir_v24i12e39483_app2.docx]

Multimedia Appendix 3
Prespecified list of self-management tasks.
[DOCX File , 23 KB - jmir_v24i12e39483_app3.docx]

Multimedia Appendix 4
The κ values for assessing the interrater reliability of behavior change technique coding.
[DOCX File , 16 KB - jmir_v24i12e39483_app4.docx]

Multimedia Appendix 5
Instances of behavior change techniques in Healthy Living for each health behavior.
[DOCX File , 15 KB - jmir_v24i12e39483_app5.docx]

Multimedia Appendix 6
Description of Healthy Living using the Template for Intervention Description and Replication framework.
[DOCX File , 25 KB - jmir_v24i12e39483_app6.docx]

References

1. Khan MA, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes - global burden of disease and forecasted trends. J Epidemiol Glob Health 2020 Mar;10(1):107-111 [FREE Full text] [doi: 10.2991/jegh.k.191028.001] [Medline: 32175717]
2. Chatterjee S, Davies MJ, Heller S, Speight J, Snoek FJ, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Feb;6(2):130-142. [doi: 10.1016/S2213-8587(17)30239-5] [Medline: 28970034]
3. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Hess Fischl A, et al. Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American diabetes association, the American association of diabetes educators, and the academy of nutrition and dietetics. Diabetes Care 2015 Jul 05;38(7):1372-1382. [doi: 10.2337/dc15-0730] [Medline: 26048904]
4. Type 2 diabetes in adults: management. National Institute for Health and Care Excellence. 2015 Dec 2. URL: https://www.nice.org.uk/guidance/ng28/resources/type-2-diabetes-in-adults-management-pdf-1837338615493 [accessed 2022-12-01]
5. Davies M, Heller S, Skinner T, Campbell M, Carey M, Cradock S, et al. Effectiveness of the diabetes education and self management for ongoing and newly diagnosed (DESMOND) programme for people with newly diagnosed type 2 diabetes: cluster randomised controlled trial. BMJ 2008 Feb 14;336(7642):491-495. [doi: 10.1136/bmj.39474.922025.be]
6. Deakin T, Cade J, Williams R, Greenwood D. Structured patient education: the diabetes X-PERT Programme makes a difference. Diabet Med 2006 Sep;23(9):944-954. [doi: 10.1111/j.1464-5491.2006.01906.x] [Medline: 16922700]
7. Sturt J, Whitlock S, Fox C, Hearnshaw H, Farmer A, Wakelin M, et al. Effects of the diabetes manual 1:1 structured education in primary care. Diabet Med 2008 Jun;25(6):722-731. [doi: 10.1111/j.1464-5491.2008.02451.x] [Medline: 18435777]
8. Odgers-Jewell K, Ball L, Kelly J, Isenring E, Reidlinger D, Thomas R. Effectiveness of group-based self-management education for individuals with Type 2 diabetes: a systematic review with meta-analyses and meta-regression. Diabet Med 2017 Aug;34(8):1027-1039. [doi: 10.1111/dme.13340] [Medline: 28226200]
9. Teljeur C, Moran PS, Walshe S, Smith SM, Cianci F, Murphy L, et al. Economic evaluation of chronic disease self-management for people with diabetes: a systematic review. Diabet Med 2017 Aug;34(8):1040-1049. [doi: 10.1111/dme.13281] [Medline: 27770591]
10. National Diabetes Audit, 2016-17 Report 1: care processes and treatment targets. NHS Digital. 2018 Mar 14. URL: https://www.hqip.org.uk/wp-content/uploads/2018/03/National-Diabetes-Audit-2016-17-Report-1-Care-Processes-and-Treatment-T...pdf [accessed 2022-12-01]
11. Offringa R, Sheng T, Parks L, Clements M, Kerr D, Greenfield MS. Digital diabetes management application improves glycemic outcomes in people with type 1 and type 2 diabetes. J Diabetes Sci Technol 2018 May 25;12(3):701-708 [FREE Full text] [doi: 10.1177/1932296817747291] [Medline: 29277103]
12. Kebede MM, Zeeb H, Peters M, Heise TL, Pischke CR. Effectiveness of digital interventions for improving glycemic control in persons with poorly controlled type 2 diabetes: a systematic review, meta-analysis, and meta-regression analysis. Diabetes Technol Ther 2018 Nov;20(11):767-782. [doi: 10.1089/dia.2018.0216] [Medline: 30257102]
13. He Q, Zhao X, Wang Y, Xie Q, Cheng L. Effectiveness of smartphone application-based self-management interventions in patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. J Adv Nurs 2022 Feb 29;78(2):348-362. [doi: 10.1111/jan.14993] [Medline: 34324218]
14. Murray E, Sweeting M, Dack C, Pal K, Modrow K, Hudda M, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. BMJ Open 2017 Sep 27;7(9):e016009 [FREE Full text] [doi: 10.1136/bmjopen-2017-016009] [Medline: 28954789]
15. Poduval S, Marston L, Hamilton F, Stevenson F, Murray E. Feasibility, acceptability, and impact of a web-based structured education program for type 2 diabetes: real-world study. JMIR Diabetes 2020 Jan 06;5(1):e15744 [FREE Full text] [doi: 10.2196/15744] [Medline: 31904580]
16. 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]

17. Schinckus L, Van den Broucke S, Housiaux M, Diabetes Literacy Consortium. Assessment of implementation fidelity in diabetes self-management education programs: a systematic review. Patient Educ Couns 2014 Jul;96(1):13-21. [doi: 10.1016/j.pec.2014.04.002] [Medline: 24795074]
18. Aziz Z, Absetz P, Oldroyd J, Pronk NP, Oldenburg B. A systematic review of real-world diabetes prevention programs: learnings from the last 15 years. Implement Sci 2015 Dec 15;10:172 [FREE Full text] [doi: 10.1186/s13012-015-0354-6] [Medline: 26670418]
19. 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]
20. Murray E, Ross J, Pal K, Li J, Dack C, Stevenson F, et al. A Web-Based Self-Management Programme for People With Type 2 Diabetes: The HeLP-Diabetes Research Programme Including RCT. Southampton (UK): NIHR Journals Library; Sep 2018.
21. French DP, Banafa R, Williams S, Taylor C, Brown LJ. How does the understanding, experience, and enactment of self-regulation behaviour change techniques vary with age? A thematic analysis. Appl Psychol Health Well Being 2021 Feb 19;13(1):239-260. [doi: 10.1111/aphw.12243] [Medline: 33075200]
22. Corbin J, Strauss A. Unending Work and Care: Managing Chronic Illness at Home. US: Jossey-Bass; 1988.
23. Dombrowski SU, O'Carroll RE, Williams B. Form of delivery as a key 'active ingredient' in behaviour change interventions. Br J Health Psychol 2016 Nov 06;21(4):733-740. [doi: 10.1111/bjhp.12203] [Medline: 27709824]
24. Dack C, Ross J, Stevenson F, Pal K, Gubert E, Michie S, et al. A digital self-management intervention for adults with type 2 diabetes: combining theory, data and participatory design to develop HeLP-Diabetes. Internet Interv 2019 Sep;17:100241 [FREE Full text] [doi: 10.1016/j.invent.2019.100241] [Medline: 31372349]
25. Linke S, McCambridge J, Khadjesari Z, Wallace P, Murray E. Development of a psychologically enhanced interactive online intervention for hazardous drinking. Alcohol Alcohol 2008 Aug 07;43(6):669-674 [FREE Full text] [doi: 10.1093/alcalc/agn066] [Medline: 18693217]
26. Wallace P, Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, et al. On-line randomized controlled trial of an internet based psychologically enhanced intervention for people with hazardous alcohol consumption. PLoS One 2011 Mar 09;6(3):e14740 [FREE Full text] [doi: 10.1371/journal.pone.0014740] [Medline: 21408060]
27. Yardley L, Williams S, Bradbury K, Garip G, Renouf S, Ware L, et al. Integrating user perspectives into the development of a web-based weight management intervention. Clin Obes 2012 Oct 14;2(5-6):132-141. [doi: 10.1111/cob.12001] [Medline: 25586248]
28. Yardley L, Ware LJ, Smith ER, Williams S, Bradbury KJ, Arden-Close EJ, et al. Randomised controlled feasibility trial of a web-based weight management intervention with nurse support for obese patients in primary care. Int J Behav Nutr Phys Act 2014 May 21;11:67 [FREE Full text] [doi: 10.1186/1479-5868-11-67] [Medline: 24886516]
29. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
30. Brown J, Michie S, Geraghty AW, Yardley L, Gardner B, Shahab L, et al. Internet-based intervention for smoking cessation (StopAdvisor) in people with low and high socioeconomic status: a randomised controlled trial. Lancet Respiratory Med 2014 Dec;2(12):997-1006. [doi: 10.1016/s2213-2600(14)70195-x]
31. How the beta phase works. GOV.UK. 2021. URL: https://www.gov.uk/service-manual/agile-delivery/how-the-beta-phase-works [accessed 2022-12-01]
32. Lenarduzzi V, Taibi D. MVP explained: a systematic mapping study on the definitions of minimal viable product. In: Proceedings of the 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). 2016 Presented at: 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA); Aug 31- Sep 2, 2016; Limassol, Cyprus. [doi: 10.1109/seaa.2016.56]
33. Abraham C, Wood CE, Johnston M, Francis J, Hardeman W, Richardson M, et al. Reliability of identification of behavior change techniques in intervention descriptions. Ann Behav Med 2015 Dec;49(6):885-900. [doi: 10.1007/s12160-015-9727-y] [Medline: 26290002]
34. BCT Taxonomy homepage. BCT Taxonomy. URL: https://www.bct-taxonomy.com/ [accessed 2022-12-01]
35. Hawkes R, Cameron E, Bower P, French D. Does the design of the NHS Diabetes Prevention Programme intervention have fidelity to the programme specification? A document analysis. Diabet Med 2020 Aug;37(8):1357-1366 [FREE Full text] [doi: 10.1111/dme.14201] [Medline: 31808578]
36. Hawkes RE, Cameron E, Miles LM, French DP. The fidelity of training in behaviour change techniques to intervention design in a national diabetes prevention programme. Int J Behav Med 2021 Dec 09;28(6):671-682 [FREE Full text] [doi: 10.1007/s12529-021-09961-5] [Medline: 33559009]
37. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul 01;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
38. Improving supported self-management for people with diabetes. Diabetes UK. 2009 Nov. URL: https://www.diabetes.org.uk/resources-s3/2017-09/Supported_self-management.pdf [accessed 2022-12-01]

39. Decoster VA. The emotions of adults with diabetes. Social Work Health Care 2003 May;36(4):79-99. [doi: 10.1300/j010v36n04_05]
40. 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. BMJ 2014 Mar 07;348(mar07 3):g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
41. Depoy E, Gitlin L. Gathering information in naturalistic inquiry. In: Introduction to Research - Understanding and Applying Multiple Strategies 4th edition. Amsterdam, Netherlands: Elsevier; 2011.
42. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
43. 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;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
44. Government functional standard GovS 005: digital, data and technology. GOV.UK. 2020 Jul 2. URL: https://www.gov.uk/guidance/digital-data-and-technology-functional-standard-version-1 [accessed 2022-12-01]
45. Web Content Accessibility Guidelines (WCAG) 2.1. W3C Recommendation. 2018 Jun 5. URL: https://www.w3.org/TR/WCAG21/ [accessed 2022-12-01]
46. Digital Technology Assessment Criteria (DTAC). NHS England. URL: https://transform.england.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/ [accessed 2022-12-01]
47. Content style guide. NHS. URL: https://service-manual.nhs.uk/content [accessed 2022-12-01]
48. NHS Digital style guidelines. NHS Digital. URL: https://tinyurl.com/3eb6cwe3 [accessed 2022-12-01]
49. Five year forward view. National Health Service. 2014 Oct. URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [accessed 2022-12-02]
50. Pittaway S, Cupitt C, Palmer D, Arowobusoye N, Milne R, Holttum S, et al. Comparative, clinical feasibility study of three tools for delivery of cognitive behavioural therapy for mild to moderate depression and anxiety provided on a self-help basis. Ment Health Fam Med 2009 Sep;6(3):145-154 [FREE Full text] [Medline: 22477905]
51. Herxheimer A, McPherson A, Miller R, Chapple A, Shepperd S, Ziebland S, et al. [DIPEx (Database of Individual Patients Experience of illness): a multimedia proposal to share experiences and information about illnesses between patients and health professionals]. Aten Primaria 2003;31(6):386-388 [FREE Full text] [doi: 10.1016/s0212-6567(03)70704-5] [Medline: 12716575]
52. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dent 2011;71(s1):S52-S63 [FREE Full text] [doi: 10.1111/j.1752-7325.2011.00233.x] [Medline: 21499543]
53. Coorey G, Peiris D, Scaria A, Mulley J, Neubeck L, Hafiz N, et al. An internet-based intervention for cardiovascular disease management integrated with primary care electronic health records: mixed methods evaluation of implementation fidelity and user engagement. J Med Internet Res 2021 Apr 26;23(4):e25333 [FREE Full text] [doi: 10.2196/25333] [Medline: 33900204]
54. Hawkes RE, Miles LM, French DP. Fidelity to program specification of the national health service digital diabetes prevention program behavior change technique content and underpinning theory: document analysis. J Med Internet Res 2022 Apr 27;24(4):e34253 [FREE Full text] [doi: 10.2196/34253] [Medline: 35476035]
55. French D, Hawkes R, Bower P, Cameron E. Is the NHS diabetes prevention programme intervention delivered as planned? An observational study of fidelity of intervention delivery. Ann Behav Med 2021 Oct 27;55(11):1104-1115 [FREE Full text] [doi: 10.1093/abm/kaaa108] [Medline: 33580647]
56. Hawkes RE, Warren L, Cameron E, French DP. An evaluation of goal setting in the NHS England diabetes prevention programme. Psychol Health 2022 Feb 01;37(2):131-150. [doi: 10.1080/08870446.2021.1872790] [Medline: 33517780]
57. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
58. van Vugt M, de Wit M, Cleijne WH, Snoek FJ. Use of behavioral change techniques in web-based self-management programs for type 2 diabetes patients: systematic review. J Med Internet Res 2013 Dec 13;15(12):e279 [FREE Full text] [doi: 10.2196/jmir.2800] [Medline: 24334230]
59. 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]
60. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
61. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb;40(1):85-98 [FREE Full text] [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]

62. Jones A, Vallis M, Pouwer F. If it does not significantly change HbA1c levels why should we waste time on it? A plea for the prioritization of psychological well-being in people with diabetes. Diabet Med 2015 Feb 10;32(2):155-163. [doi: 10.1111/dme.12620] [Medline: 25354315]
63. Speight J, Hendrieckx C, Pouwer F, Skinner TC, Snoek FJ. Back to the future: 25 years of 'guidelines for encouraging psychological well-being' among people affected by diabetes. Diabet Med 2020 Aug 08;37(8):1225-1229. [doi: 10.1111/dme.14165] [Medline: 31661183]
64. Egede LE, Ellis C, Grubaugh AL. The effect of depression on self-care behaviors and quality of care in a national sample of adults with diabetes. Gen Hosp Psychiatry 2009 Sep;31(5):422-427. [doi: 10.1016/j.genhosppsych.2009.06.007] [Medline: 19703635]
65. Schmidt CB, van Loon BJ, Vergouwen AC, Snoek FJ, Honig A. Systematic review and meta-analysis of psychological interventions in people with diabetes and elevated diabetes-distress. Diabet Med 2018 Jun 13;35(9):1157-1172. [doi: 10.1111/dme.13709] [Medline: 29896760]
66. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013 Mar 28(3):CD008776 [FREE Full text] [doi: 10.1002/14651858.CD008776.pub2] [Medline: 23543567]
67. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
68. 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]
69. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
70. Wylie T, Shah C, Connor R, Farmer A, Ismail K, Millar B, et al. Transforming mental well-being for people with diabetes: research recommendations from Diabetes UK's 2019 Diabetes and Mental Well-Being Workshop. Diabet Med 2019 Dec;36(12):1532-1538 [FREE Full text] [doi: 10.1111/dme.14048] [Medline: 31177573]
71. Author archives: all-party parliamentary group for diabetes. APPG for Diabetes. URL: https://diabetesappg.wordpress.com/author/matthewlower/ [accessed 2022-12-01]
72. 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]
73. McCrabb S, Lane C, Hall A, Milat A, Bauman A, Sutherland R, et al. Scaling-up evidence-based obesity interventions: a systematic review assessing intervention adaptations and effectiveness and quantifying the scale-up penalty. Obes Rev 2019 Jul;20(7):964-982. [doi: 10.1111/obr.12845] [Medline: 30868745]
74. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
75. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
76. Jurkeviciute M, Eriksson H. Exploring the use of evidence from the development and evaluation of an electronic health (eHealth) trial: case study. J Med Internet Res 2020 Aug 28;22(8):e17718 [FREE Full text] [doi: 10.2196/17718] [Medline: 32857057]
77. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
78. Evidence standards framework (ESF) for digital health technologies. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2022-12-01]
79. Fulton E, Kwah K, Wild S, Brown K. Lost in translation: transforming behaviour change techniques into engaging digital content and design for the StopApp. Healthcare (Basel) 2018 Jul 06;6(3):75 [FREE Full text] [doi: 10.3390/healthcare6030075] [Medline: 29986396]
80. De Longh A, Fagan P, Fenner J, Kidd L. A Practical Guide to Self-Management Support Key Components for Successful Implementation. London, England, United Kingdom: Health Foundation; 2015.

Abbreviations

BCT: behavior change technique

HeLP-Diabetes: Healthy Living for People With Type 2 Diabetes
NHS: National Health Service
NHS-DDPP: National Health Service Digital Diabetes Prevention Programme
RCT: randomized controlled trial
T2DM: type 2 diabetes mellitus
TIDieR: Template for Intervention Description and Replication

Edited by G Eysenbach; submitted 11.05.22; peer-reviewed by C Reidy, C Varsi; comments to author 22.08.22; revised version received 20.09.22; accepted 07.10.22; published 07.12.22.

Please cite as:
Benton JS, Cotterill S, Hawkes RE, Miles LM, French DP
Changes in a Digital Type 2 Diabetes Self-management Intervention During National Rollout: Mixed Methods Study of Fidelity
J Med Internet Res 2022;24(12):e39483
URL: https://www.jmir.org/2022/12/e39483
doi:10.2196/39483
PMID:36476723

Original Paper

Evaluation of OPTIMISE (Online Programme to Tackle Individual's Meat Intake Through Self-regulation): Cohort Study

Cristina Stewart[1], BSc, MSc; Carmen Piernas[1], BSc, MSc, PhD; Kerstin Frie[1], BSc, MSc, DPhil; Brian Cook[1], BA, MA, PhD; Susan A Jebb[1], BSc, PhD

Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Susan A Jebb, BSc, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 01865 617826
Email: susan.jebb@phc.ox.ac.uk

Abstract

Background: There is an urgent need to reduce society's meat consumption to help mitigate climate change and reduce noncommunicable diseases.

Objective: This study aimed to investigate changes in meat intake after participation in an online, multicomponent, self-regulation intervention.

Methods: We conducted a pre-post observational study among adult meat eaters in the United Kingdom who signed up to a website offering support based on self-regulation theory to reduce meat consumption. The program lasted 9 weeks (including a 1-week baseline phase, a 4-week active intervention phase, and a 4-week maintenance phase), comprising self-monitoring, goal setting, action planning, and health and environmental feedback. Meat intake was estimated during weeks 1, 5, and 9 using a 7-day meat frequency questionnaire. We analyzed the change in mean daily meat intake from baseline to week 5 and week 9 among those reporting data using a hierarchical linear mixed model. We assessed changes in attitudes toward meat consumption by questionnaire and considered the acceptability and feasibility of the intervention.

Results: The baseline cohort consisted of 289 participants, of whom 77 were analyzed at week 5 (26.6% of the baseline sample) and 55 at week 9 (71.4% of the week 5 sample). We observed large reductions in meat intake at 5 and 9 weeks: –57 (95% CI –70 to –43) g/day (P<.001) and –49 (95% CI –64 to –34) g/day (P<.001), respectively. Participants' meat-free self-efficacy increased, meat-eating identities moved toward reduced-meat and non–meat-eating identities, and perceptions of meat consumption as the social norm reduced. Participants who completed the study reported high engagement and satisfaction with the intervention.

Conclusions: Among people motivated to engage, this online self-regulation program may lead to large reductions in meat intake for more than 2 months, with promising signs of a change in meat-eating identity toward more plant-based diets. This digital behavior change intervention could be offered to complement population-level interventions to support reduction of meat consumption.

(J Med Internet Res 2022;24(12):e37389) doi:10.2196/37389

KEYWORDS

self-regulation; self-monitoring; goal setting; meat intake; meat reduction; multi-component intervention; health; nutrition; diet

Introduction

Population-level changes in meat consumption are needed to help mitigate climate change and reduce noncommunicable diseases. The livestock sector is a leading contributor to environmental degradation [1], while a high intake of meat, particularly red and processed meat, has been linked to type 2 diabetes, cardiovascular disease, and some forms of cancer [2]. There is a growing interest in reduced-meat diets, primarily for health reasons, but also because of concerns regarding animal

welfare and the environment [3]. According to UK public attitude surveys, 65% of people surveyed in 2020 were willing to consider eating less meat [3], up from 35% in 2014 [4]. Meat substitutes are also rising in popularity; a trend analysis of the UK National Diet and Nutrition Survey (NDNS) found that their consumption has almost doubled in the last decade [5], and market research data suggests the number of British people eating these products has increased from 50% in 2017 to 65% in 2019 [6]. However, meat consumption in the United Kingdom is decreasing only slowly (–17 g/capita/day; –17% in the last decade) [7], suggesting people need more support to enact their intentions to reduce meat in their diet and close the intention-behavior gap.

Individual-level interventions (targeting our conscious and reflective decision-making processes) can complement interventions at a population level (targeting automatic, nonconscious processes) [8], but need to be delivered at scale [9-11]. Using digital technology (eg, mobile apps, interactive websites, and text messaging) is a promising approach to providing scalable, cost-effective interventions [12], and evidence suggests this approach can help promote a range of healthy behaviors [12-15]. Previous research has noted that these interventions need to be thoroughly grounded in behavior change theory [16].

We recently developed an online multicomponent intervention, OPTIMISE (Online Programme to Tackle Individual's Meat Intake Through Self-regulation), based on self-regulation theory to support individuals in reducing their meat consumption. The intervention guides individuals through a self-regulation process of self-monitoring, goal setting, learning about the health and environmental impact of their meat intake, action planning, and regular reflection. We tested its effectiveness in a randomized controlled trial (RCT) [17] among adults who ate meat very regularly (≥5 times per week), and found it led to significant reductions in meat intake: a 40 g/day greater reduction, relative to the control group, at 5 weeks. Identifying effective and potentially scalable interventions that can support people's efforts to enact their intentions to eat less meat is imperative to improve both planetary and human health.

This population-based cohort study builds upon our previous RCT and aims to investigate whether this online self-regulation intervention is effective in helping the general population in the United Kingdom who eat at least some meat to reduce their meat intake. A secondary aim was to investigate the adherence to and acceptability of the intervention.

Methods

Study Design and Setting

We conducted a cohort study among UK adults using OPTIMISE, an online program to support meat reduction based on self-regulation theory. All aspects of the study were delivered remotely through a website developed specifically for the intervention, through which all data collection took place between May 28, 2021, and December 13, 2021.

Ethics Approval

This study was granted ethical approval by the Central University Research Ethics Committee of the University of Oxford (R71430/RE003).

Participant Recruitment

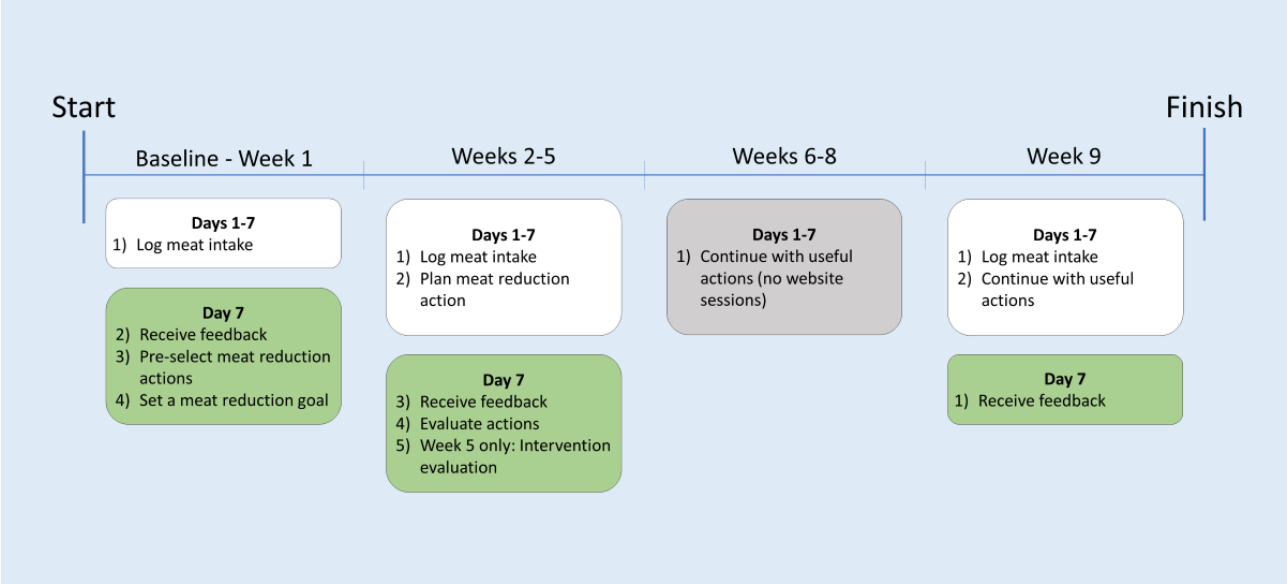
We made the website publicly available and signposted it to people over 6 months through public engagement events across the United Kingdom via our research team's website and social media presence, as well as online newsletters and volunteer databases (eg, Research for the Future) [18]. Everyone who accessed the OPTIMISE website was offered the opportunity to sign up as a study participant or use the program independently on their own. Recruitment closed 1 month after our last public engagement event.

People interested in taking part in the research completed a screening questionnaire to assess eligibility (participants were aged 18 years or older, were resident in the United Kingdom, were meat eaters, and wanted to reduce their meat intake), and they provided consent for participation in the study before registering with the OPTIMISE website using their email address. Participants who completed the program were entered into a raffle to win a £100 (US $122.53) digital gift card (1 gift card was available for every 50 participants).

Study Procedures

The study lasted 9 weeks (including a baseline week of self-monitoring meat consumption, a 4-week active intervention phase, and a 4-week maintenance phase; Figure 1), with 3 data collection weeks: baseline (week 1), first follow-up (week 5) and second follow-up (week 9). After registering with the OPTIMISE website, participants were presented with information regarding the health and environmental benefits of eating less meat (Multimedia Appendix 1). Participants then completed a baseline questionnaire that asked about their demographic characteristics, dietary restrictions, and attitudes toward meat consumption. The attitudes assessed were meat-eating identity, meat-free self-efficacy (adapted from Lacroix and Gifford's self-efficacy scale [19]), motivation to reduce meat consumption, perception of meat consumption as the social norm (consisting of the 4 *N*'s—the belief that eating meat is "natural, normal, necessary, and nice" [20]), and social support for meat reduction. Full details are provided in Multimedia Appendix 2. Participants repeated the attitude questionnaire at both follow-ups (at weeks 5 and 9). Meat consumption was measured daily during the 3 data collection weeks using a specific meat frequency questionnaire [21]. This questionnaire asked participants to report the number of servings of individual meat and seafood items they consumed in the previous 24 hours. Serving sizes were based on underlying portion size data from the UK Food Standards Agency combined with estimates of meat content from composite dishes from the UK NDNS [22,23].

Figure 1. Procedure of the OPTIMISE (Online Programme to Tackle Individual's Meat Intake Through Self-regulation) study.



The Intervention

The full intervention has been described in detail previously [17]. In short, on the last day of the baseline week, participants received feedback on the health and environmental impacts of their total meat consumption and red meat consumption. They were then presented with a list of 26 meat consumption reduction actions across 6 categories that they could preselect for the upcoming weeks: (1) "preparing to change"; (2) "try swapping out meat for veg"; (3) "try something new"; (4) "cut out specific animal products"; (5) "limit intake of animal products"; and (6) "get family and friends involved" (Multimedia Appendix 3). The participants were also prompted to set themselves a goal to reduce their meat consumption. Every day throughout the active intervention phase (ie, weeks 2-5), participants self-monitored their meat consumption and planned a meat reduction action. Each subsequent morning they were asked if they had managed to perform their action on the previous day; if they had not, they were asked to reflect on what they could do differently next time. At the end of each week of the active intervention phase, the participants received feedback on how their meat consumption compared to baseline (Multimedia Appendix 4), and they were asked to reflect on how useful they found the actions they had chosen that week. At the first follow-up (week 5), participants completed an intervention evaluation questionnaire. During the 4-week maintenance phase (ie, weeks 6-9), the participants were asked to continue performing the actions they found useful during the active intervention phase offline, with no web sessions to complete.

Outcome Measures

The main outcome measure was the change in mean daily meat consumption from baseline to week 5, measured by the daily meat frequency questionnaires [21]. We also assessed the change in (1) total mean daily meat consumption from baseline to week 9, (2) total mean daily meat consumption from week 5 to week 9, and (3) mean daily consumption of meat subtypes comprising red meat and processed meat from baseline to weeks 5 and 9. We also explored the predictors of change in meat intake and change in attitudes toward meat consumption from baseline to weeks 5 and 9. We assessed adherence to the intervention as the proportion of the 42 sessions participants completed and the acceptability of the intervention based on responses to the intervention evaluation questionnaire.

Statistical Analysis

All statistical analyses were conducted in Stata/IC (version 14.1). $P<.05$ was set to denote statistical significance. We published a statistical analysis plan on the Open Science Framework preceding the analyses on October 18, 2021 [24].

For each participant and time point (baseline, week 5, and week 9), we calculated mean total daily intakes of all meat and meat subtypes (ie, red meat and processed meat). The main analysis used a hierarchical linear mixed model with fixed effects for "time point" and random effects for "participant" to investigate whether meat consumption at weeks 5 and 9 differed significantly from baseline. As prespecified in our statistical analysis plan, days in which reported meat intake exceeded 1.5 kg were excluded, as we deemed this implausible. We identified no confounding variables through univariable regressions and so the model was unadjusted.

To analyze the predictors of change in mean daily meat consumption from baseline to week 5, we used a multivariable linear regression model with change in meat consumption as the dependent variable and possible predictors included in one single model. The predictors were age, gender, ethnic group, highest educational qualification, household size, annual household income, the response to "currently trying to lose weight" (yes/no), dietary restrictions, baseline meat consumption, baseline attitudes toward meat (ie, meat-eating identity, including non–meat eater, reduced-meat eater, and meat eater; mean meat-free self-efficacy; meat reduction motivation; mean meat consumption social norms; and meat reduction social support), tertiles of engagement (based on the percentage of sessions participants completed throughout the

active intervention phase), and the number of action categories tried at least once.

We used hierarchical linear mixed models to investigate changes in attitudes toward meat (ie, meat-free self-efficacy, meat reduction motivation, meat consumption social norms, and meat reduction social support), between baseline and weeks 5 and 9. Due to multicollinearity between tertiles of engagement and meat-eating identity changes, we used the chi-square goodness-of-fit test to explore the proportions of each meat-eating identity at both follow-ups compared to baseline. Written feedback collected from participants as part of the intervention evaluation questionnaire was analyzed qualitatively using inductive thematic analysis in NVivo 12 (QSR International) [25].

Sensitivity analyses were performed using 2-tailed independent *t* tests (for normal continuous data), Mann-Whitney *U* tests (for skewed continuous data), and chi-square tests (for categorical data) to explore baseline differences in participants who did not provide any outcome data (ie, who did not complete any sessions in week 5) and those who did.

Exploratory Analyses

To explore barriers to adherence to participants' chosen meat reduction actions, we analyzed the free-text responses to the daily action completion question when participants indicated they had not managed to perform their action using inductive thematic analysis [25].

Results

Participants

A total of 566 individuals signed up to the study website, 59 of whom requested their account (and subsequently all their data) be deleted before the end of the study. We were unable to establish which of these 59 individuals were study participants and which were independent users. Of the remaining 507 individuals for whom we had data, 120 registered as independent users and 387 registered as study participants.

Of the study participants, 82 did not complete any baseline sessions, 3 did not complete the baseline demographics questionnaire, and 7 reported no meat consumption during the baseline week. Six participants were excluded as they self-reported eating >1.5 kg of meat per day in every meat frequency questionnaire they completed. The total baseline cohort, therefore, consisted of 289 of the 387 registered participants (74.7%). Participants were aged 18 to 84 years (mean 46.8, SD 13.8 years), 72.3% (209/289) were female, and 57.1% (165/289) were White British (Table 1). Reported total meat consumption at baseline was 146 (SD 162) g/day (Table 2).

Eleven participants did not complete their goal setting, preselect their actions, or both, and a further 201 participants did not complete any sessions in week 5, leaving 77 participants in our first follow-up sample (week 5; 26.6% of the baseline sample of 289 participants). Of these participants, 22 did not complete any sessions in week 9, leaving 55 participants in our second follow-up sample (week 9; 71.4% of the first follow-up sample of 77 participants). Figure 2 depicts a flow diagram of participants throughout the study.

In the baseline cohort, the most important motivating factor to reduce meat intake on a scale from 1 (not at all important) to 10 (extremely important) was to help the environment (mean score 8.6, SD 1.5), followed by health benefits (mean score 7.9, SD 1.8) and animal welfare concerns (mean score 7.6, SD 2.3). The mean meat-consumption reduction goal shows participants on average challenged themselves to reduce their meat consumption by nearly a quarter (–23%, SD 13%; range 5%-90%).

Participants who dropped out before week 5 were more likely to be trying to lose weight (P=.03) and were less motivated to reduce their meat consumption at baseline (P=.02) compared to those who completed week 5 sessions. No other baseline measurements differed significantly between these groups (Multimedia Appendix 5).

Table 1. Baseline characteristics (N=289).

Characteristics	Values
Age[a] (years), mean (SD)	46.8 (13.8)
Gender, n (%)	
Female	209 (72.3)
Male	78 (27)
Other/prefer not to say	2 (0.7)
Ethnicity, n (%)	
White British	165 (57.1)
White other	84 (29.1)
Asian or Asian British	17 (5.9)
Black or Black British	4 (1.4)
Mixed/other	18 (6.2)
Prefer not to say	1 (0.4)
Highest educational qualification, n (%)	
University degree, NVQ[b] level 4-5 or equivalent, and above	242 (83.7)
Other post–high school qualifications	15 (5.2)
A-levels[c], NVQ level 2-3 or equivalent	21 (7.3)
Apprenticeship	1 (0.4)
GCSE[d], NVQ level 1, or equivalent	2 (0.7)
Other vocational, work-related qualifications	3 (1)
No formal qualifications	1 (0.4)
Prefer not to say	4 (1.4)
Household size, n (%)	
1 person	55 (19)
2 people	115 (39.8)
3 people	57 (19.7)
4 people	48 (16.6)
5 people	10 (3.5)
≥6 people	4 (1.4)
Annual household income, n (%)	
<£15,000 (US $18,418)	10 (3.5)
£15,000-£24,999 (US $18,418-$30,695)	24 (8.3)
£25,000-£39,999 (US $30,695-$49,113)	45 (15.6)
£40,000-£75,000 (US $49,113-$92,090)	99 (34.3)
>£75,000 (>US $92,090)	90 (31.1)
Prefer not to say	21 (7.3)
Currently trying to lose weight, n (%)	
Yes	198 (68.5)
No	91 (31.5)
Dietary restrictions[e], n (%)	
Dairy-free	14 (4.9)
Gluten-free	19 (6.6)

Characteristics	Values
Fish and shellfish allergy	3 (1)
None	259 (89.6)
How participants heard of the program, n (%)	
Public engagement events	4 (1.4)
Research team's website/social media	6 (2.1)
Friends or family members	17 (5.9)
Social media	48 (16.6)
Radio or newspaper	181 (62.6)
Volunteer databases	9 (3.1)
Other	24 (8.3)

[a]Age ranged from 18 to 84 years.

[b]NVQ: National Vocational Qualification.

[c]Advanced level (A-level) qualifications are subject-based qualifications for students aged 16 or older.

[d]GCSE: General Certificate of Secondary Education.

[e]Participants could select multiple answers to this question.

Table 2. Meat consumption and attitudes at baseline and both follow-ups. Estimates are from mixed effects models with fixed effects for "time point" and random effects for "participant." The models were unadjusted, as we identified no potential confounders in univariate analyses. Data on meat-eating identity are shown in Multimedia Appendix 6. Baseline N=289; meat consumption n=77 and n=55 participants at first and second follow-ups, respectively; attitudinal measures, n= 55 and n=41 participants at first and second follow-ups, respectively.

	Baseline, mean (SD)	First follow-up (week 5)			Second follow-up (week 9)		
		Mean (SD)	Mean difference (95% CI)	*P* value	Mean (SD)	Mean difference (95% CI)	*P* value
Meat consumption (g/day)							
Total meat	146 (162)	61 (50)	–57 (–70 to –43)	<.001	68 (51)	–49 (–64 to –34)	<.001
Red meat	53 (65)	27 (33)	–22 (–32 to –12)	<.001	28 (31)	–21 (–32 to –10)	<.001
Processed meat	40 (80)	17 (23)	–13 (–19 to –8)	<.001	18 (22)	–12 (–18 to –5)	<.001
Red and processed meat	92 (121)	44 (52)	–35 (–49 to –22)	<.001	46 (48)	–33 (–48 to –18)	<.001
Attitudinal measures							
Meat-free self-efficacy score[a]	3.2 (1.2)	2.8 (1.4)	–0.3 (–0.6 to 0.0)	.09	2.4 (1.1)	–0.8 (–1.2 to –0.5)	<.001
Meat reduction motivation score[b]	7.5 (1.6)	8.1 (1.6)	0.2 (–0.3 to 0.8)	.45	7.7 (2.1)	–0.1 (–0.7 to 0.5)	.72
Meat consumption social norms score[c]	4.4 (1.0)	4.2 (1.2)	–0.2 (–0.4 to 0.0)	.13	4.0 (1.1)	–0.4 (–0.6 to –0.1)	.001
Meat reduction social support score[d]	6.6 (2.5)	7.0 (2.6)	–0.1 (–0.8 to 0.5)	.66	6.3 (2.5)	–0.8 (–1.5 to –0.1)	.02

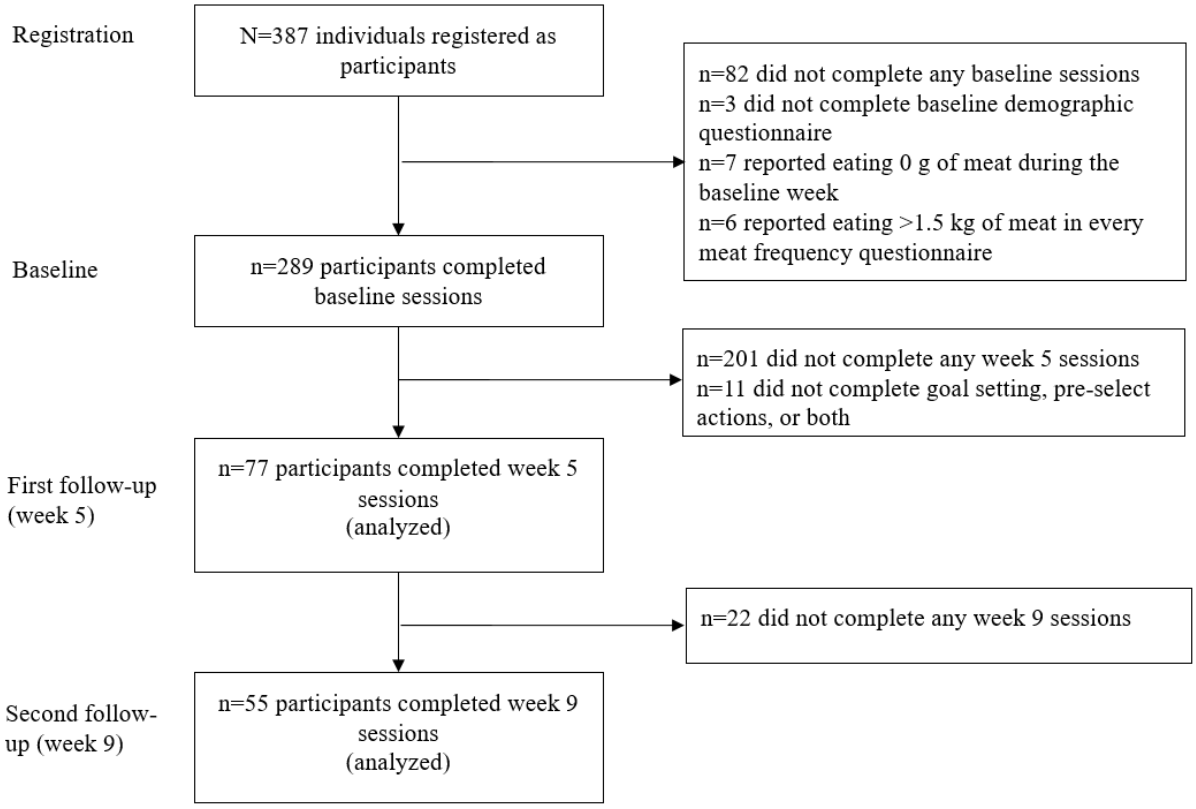
[a]Mean score of 3 self-efficacy questions ("I lack the cooking skills to prepare meat-free meals," "I don't know what to eat instead of meat," and "I don't have enough willpower to not eat meat"), measured on a scale from 1 (strongly disagree) to 7 (strongly agree).

[b]Participants were asked to respond to the question "How motivated are you to reduce your meat intake beyond the context of this programme?" on a scale from 1 (not at all motivated) to 10 (extremely motivated).

[c]Mean score of responses to 4 social norm questions using the 4 *N*'s scale (the belief that eating meat is "natural, normal, necessary, and nice") on a scale from 1 (strongly disagree) to 7 (strongly agree).

[d]Participants were asked, "How willing are the people you share your meals with to reduce their meat consumption?" on a scale from 1 (not open at all) to 10 (very open to it).

Figure 2. Flow chart of participants.



Changes in Meat Consumption

Total mean consumption of meat decreased from baseline to week 5 by –57 (95% CI –70 to –43) g/day (P<.001) and from baseline to week 9 by –49 (95% CI –64 to –34) g/day (P<.001). This included reductions in consumption of red meat and processed meat of –35 (95% CI –49 to –22) g/day (P<.001) and –33 (95% CI –48 to –18) g/day (P<.001) at weeks 5 and 9, respectively (Table 2). The reduction in total meat consumption from week 5 to week 9 was –8 (95% CI –7 to –23 g/day), but this was not a significant difference (P=.31).

Predictors of Change

Higher baseline meat consumption was associated with a greater reduction in meat intake at week 5, with every 1 g of greater baseline intake predicting a 0.9 g/day greater reduction (95% CI –1.1 to –0.7; P<.001). Choosing meat reduction actions from only one category was associated with an increase in meat consumption from baseline to week 5 of 104 (95% CI 10 to 198) g/day (P=.03). For participants choosing actions from more than one category, there was no association between the number of action categories chosen and meat intake reduction. No demographic characteristics or baseline attitudes toward meat significantly predicted change in meat intake, nor did tertiles of intervention engagement (Multimedia Appendix 7).

Changes in Attitudes Toward Meat Consumption

There was a significant change in reported meat-eating identities toward reduced-meat and non–meat-eating identities from baseline to both follow-ups (P=.005 at week 5 and P=.002 at week 9). Forty-four percent (23/52) and 43% (17/40) of participants described themselves as meat eaters at weeks 5 and 9, down 25 and 30 absolute percentage points from baseline, respectively (Multimedia Appendix 6). There was no change in any other attitudinal measures from baseline to week 5. At week 9, there was an increase in mean meat-free self-efficacy score (–0.8, 95% CI –1.2 to –0.5; P<.001), a decrease in the score for perception of meat consumption as the social norm (–0.4, 95% CI –0.6 to –0.1; P=.001), and a decrease in the score for perceived social support for meat reduction (–0.8, 95% CI –1.5 to –0.1; P=.02). There was no change in participants' motivation to reduce meat intake at either week 5 or 9 (Table 2).

Self-reported Barriers

The most commonly reported barriers for not performing meat reduction actions were (1) other people (most frequently friends and family), (2) being too busy and not having enough time, (3) eating out and the lack of meat-free options available or the temptation to opt for a meat dish, and (4) eating meat leftovers and wanting to avoid food waste. Representative quotes are as follows:

> *I was not cooking yesterday, and when you're a guest I think it's polite to eat what's been served.*
>
> *I was very exhausted today and didn't have the energy to make two dishes.*
>
> *I ate leftover food, my partner had cooked more meat than the children wanted or needed.*

Acceptability and Feasibility

More than 7 out of 10 participants dropped out before week 5 (73.4%, 212/289), but thereafter, 71% (55/77) completed the study. Of the participants who completed week 5 and week 9,

78% (60/77) and 98% (54/55) completed at least 80% (34/42) of the sessions, respectively. Fifty-five participants (71%, 55/77) completed the intervention evaluation questionnaire at week 5, rating the usefulness of the intervention components and additional resources on a scale from 1 (not useful) to 10 (very useful). Mean scores ranged from 7.2 (SD 2.7) to 9 (SD 1.7) (Table 3). Participants rated the daily meat consumption tracking to be the most useful component of the intervention (mean score 8.7, SD 1.6) and the action diary as the most useful additional resource (mean score 9.0, SD 1.7). Forty-one participants provided additional feedback as responses to the free text question; they were largely positive about their experience of the program. Responses included the following:

> *In general I've found the study interesting and important. It has been effective to chart and reflect on my meat consumption, plan for change and see my evidence of change progressively.*
>
> *It has helped me to confirm what my personal stumbling blocks are.*
>
> *To me, tracking the meat consumption and planning activities was the best way to help me out eating less meat, because I'm a naturally planned person.*

While some participants said the daily action planning was helpful, others said they would have preferred weekly actions to make planning meals in advance for the week easier. Some participants said they would have liked both social and competitive elements, allowing them to share their progress with others and compare their intake with other users or the UK average, or both.

Table 3. Intervention evaluation questionnaire results. Participants were asked how useful they found the items on a scale of 1 (not useful) to 10 (very useful). The additional resources were optional and only evaluated by those who reported using them throughout the study.

Questionnaire items	Mean score (SD)	Respondents, n
Intervention components		
Tracking your meat consumption on a daily basis	8.7 (1.6)	55
Feedback on the environmental and health impact of your meat consumption	7.6 (2.4)	55
Planning an action on a daily basis to reduce your meat consumption	7.2 (2.7)	55
Additional resources		
Weekly action evaluation	7.6 (2.4)	55
Downloadable action diary	9.0 (1.7)	3
Downloadable action overview	8.0 (2.0)	3
Links to other resources	8.3 (1.5)	10
Ability to review your journey	8.2 (1.9)	22

Discussion

Principal Results

We observed significant reductions in meat consumption when UK adult meat eaters engaged with a bespoke meat-consumption reduction website and were guided through a process of self-regulation. Participants reported marked changes in meat-eating identity toward reduced-meat and non–meat-eating identities, their meat-free self-efficacy increased, and their perception of meat consumption as the social norm decreased. There was a high dropout rate from registration to first follow-up, but the quarter of participants who provided outcome data had high engagement with the intervention and rated it highly, particularly the self-monitoring aspect.

Strengths and Limitations

Strengths of this study were that baseline meat intake was similar to that of the general UK population [7] and that we collected detailed estimates of the quantity and type of meat consumed using a specific meat frequency questionnaire [21]. We recruited participants from the general population through public engagement events, social media, and broadcast media, and our results likely reflect the characteristics of people who were attracted to this type of digital self-help support for dietary change. The OPTIMISE program was free to use, easy to sign up to, and easy to try out. To try to mimic "real-world" usage and minimize researcher bias, the participants had no direct contact with the researcher, and all aspects of the intervention were delivered remotely through our study website. Many people who initially signed up as participants did not complete any follow-up assessments, suggesting that those included in the analysis represent a particularly motivated group of people. The high dropout rate was not surprising, as previous research has noted that a high level of attrition poses a significant challenge for digital interventions [26,27], including web-based trials [28,29]. A recent systematic review of app-based interventions for chronic disease found dropout rates were high—up to 87%—with higher rates seen in observational studies than RCTs [26]. Moreover, an observational study testing a healthy-eating app found less than 3% of users were classed as "active" [27], with the majority of participants downloading the app and using it only once. As with other real-world evaluations, another limitation is that we had no randomly assigned control group and cannot infer a causal link between the website and the reduction in meat intake. The reduction in meat intake was maintained 4 weeks beyond the active intervention, and there were associated changes in meat-eating identity, factors that have been shown to be predictive of behavioral intentions [30]. Nevertheless, a longer follow-up period is needed to assess changes in habitual dietary behaviors.

Comparison With Prior Work

The absolute reduction in meat intake reported here was both large and remarkably similar to the reduction observed in the intervention group in our previous RCT (–58% vs –57% at week 5 and –53% vs –52% at week 9 in the current study and the RCT, respectively) [17]. For context, average meat intake in the United Kingdom has decreased by only 1.7% per year, on average, over the 10 years after 2008-2009 [7]. However, in both studies, we specifically recruited people seeking to reduce their meat intake, and our findings should be interpreted accordingly. We cannot infer causality or precisely identify the active components of the online program, but participants rated self-monitoring as the most useful component. In our RCT, we also observed significant reductions in meat intake in the control group, who were not offered goal setting, action planning, or feedback components but did self-monitor as part of the outcome assessments. It is plausible that the observed reductions in meat consumption are largely a result of self-monitoring. Indeed, previous research has found self-monitoring to be effective in helping individuals to reduce their meat consumption [31,32] and in promoting other positive lifestyle and dietary behavior changes [33].

Importantly, participants reported an increased meat-free self-efficacy, a marked shift toward reduced meat-eating and non–meat-eating identities, and a reduction in perception of meat consumption as the social norm. This reflects findings from a United Kingdom–based RCT that tested the effectiveness of a multicomponent behavioral intervention to reduce meat consumption [34]. That study found substantive reductions in meat intake (–61% at 4 weeks and –38% at 8 weeks in the intervention group) alongside increased intentions, positive attitudes, perceived control, and subjective norms of eating a low-meat diet. Previous research has suggested that meat-eating identity can explain intentions to reduce intake of red and processed meat [35], while higher levels of meat-free self-efficacy are an important predictor of successful meat-intake reduction [36]. We found no change in participants' motivation to reduce meat intake, though this is likely because our participants had a high level of motivation at the start of the study, with those who reached week 5 having higher motivation than those who dropped out. Participants reported a decline in perceived social support during the program consistent with their reports of friends and family being one of the greatest barriers to performing their meat-intake reduction actions.

The results of our two OPTIMISE studies, taken together, suggest this online self-regulation program may be effective for helping motivated individuals to reduce their meat intake and closing the intention-behavior gap. In comparison to in-person interventions, there is preliminary evidence to support the scalability [37] and cost-effectiveness [38] of web-based interventions. Further, the OPTIMISE website uses a self-directed format (not requiring any researcher involvement), can be hosted at a minimal cost, and is currently publicly available and open to individuals who self-select to sign up. This approach is consistent with other web-based online programs designed to be made widely accessible at scale [39]. However, since its use is likely to be restricted to individuals with sufficient intrinsic motivation to seek out support for meat-intake reduction, this can only be one part of a wider strategy to support meat-intake reduction [40]. Population-level strategies that focus on restructuring the physical microenvironments (ie, choice architecture interventions or "nudges") or economic environments to support a more plant-based diet are likely to be important complementary actions to support individual meat-intake reduction efforts [11,40].

Conclusions

An online program to encourage self-monitoring of meat consumption, together with goal setting, educative feedback, action planning, and reflection may help individuals seeking to reduce their meat intake to change their diet and foster a reduced-meat or non–meat-eating identity. This type of support could be offered at scale with minimal cost and could complement other environmental interventions to help people eat less meat.

Acknowledgments

The authors gratefully acknowledge One Ltd, who created the Online Programme to Tackle Individual's Meat Intake Through Self-regulation (OPTIMISE) website, and we thank Dr Rachel Pechey (University of Oxford) for her statistical advice regarding the analysis of meat-eating identity changes. We would also like to thank all those who helped us test the website throughout its development and provided us with feedback. This research was funded by the Wellcome Trust Our Planet Our Health program (Livestock, Environment and People [LEAP]; 205212/Z/16/Z). SAJ is funded by the National Institute for Health and Care Research (NIHR) Oxford Biomedical Research Centre. SAJ and CP are funded by the Oxford and Thames Valley NIHR Applied Research Centre. The funders of the study were not involved in the study design, conduct of the study, data collection, data analysis, data interpretation, or any aspect pertinent to the study. The funders had no role in the writing of the manuscript or in the decision to submit it for publication.

Data Availability

Data described in the manuscript will be made available upon request pending application and approval.

Authors' Contributions

CS, CP, KF, BC, and SAJ designed the research; CS conducted the research; CS analyzed the data; CS drafted the manuscript; and CS, CP, KF, and SAJ critically reviewed the manuscript. CS had primary responsibility for the final content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Health and environmental benefits presented to participants.
[DOC File , 438 KB - jmir_v24i12e37389_app1.doc]

Multimedia Appendix 2
Meat consumption attitude questions.
[DOCX File , 15 KB - jmir_v24i12e37389_app2.docx]

Multimedia Appendix 3
Meat consumption reduction actions.
[DOC File , 52 KB - jmir_v24i12e37389_app3.doc]

Multimedia Appendix 4
Example of weekly health and environmental feedback.
[DOC File , 313 KB - jmir_v24i12e37389_app4.doc]

Multimedia Appendix 5
Participant characteristics of week 5 drop-outs vs completers.
[DOCX File , 19 KB - jmir_v24i12e37389_app5.docx]

Multimedia Appendix 6
Meat-eating identity proportions.
[DOCX File , 15 KB - jmir_v24i12e37389_app6.docx]

Multimedia Appendix 7
Predictors of change in total meat intake.
[DOCX File , 18 KB - jmir_v24i12e37389_app7.docx]

References

1. Clark MA, Springmann M, Hill J, Tilman D. Multiple health and environmental impacts of foods. Proc Natl Acad Sci USA 2019 Nov 12;116(46):23357-23362 [FREE Full text] [doi: 10.1073/pnas.1906908116] [Medline: 31659030]
2. Ekmekcioglu C, Wallner P, Kundi M, Weisz U, Haas W, Hutter H. Red meat, diseases, and healthy alternatives: A critical review. Crit Rev Food Sci Nutr 2018 Jan 22;58(2):247-261. [doi: 10.1080/10408398.2016.1158148] [Medline: 27128451]
3. Public attitude survey 2020 – findings. Eating Better. URL: https://www.eating-better.org/uploads/Documents/2020/Eating%20Better%202020%20public%20survey%20analysis.docx.pdf [accessed 2022-01-16]
4. Let's talk about meat: changing dietary behaviour for the 21st century. Eating Better. 2020. URL: https://www.eating-better.org/uploads/Documents/LetsTalkAboutMeat.pdf [accessed 2022-01-16]
5. Alae-Carew C, Green R, Stewart C, Cook B, Dangour AD, Scheelbeek PFD. The role of plant-based alternative foods in sustainable and healthy food systems: Consumption trends in the UK. Sci Total Environ 2022 Feb 10;807(Pt 3):151041 [FREE Full text] [doi: 10.1016/j.scitotenv.2021.151041] [Medline: 34673070]
6. Plant-based push: UK sales of meat-free foods shot up 40% between 2014-2019. Mintel. 2020. URL: https://www.mintel.com/press-centre/food-and-drink/plant-based-push-uk-sales-of-meat-free-foods-shoot-up-40-between-2014-19 [accessed 2022-01-16]
7. Stewart C, Piernas C, Cook B, Jebb SA. Trends in UK meat consumption: analysis of data from years 1-11 (2008-09 to 2018-19) of the National Diet and Nutrition Survey rolling programme. Lancet Planet Health 2021 Oct;5(10):e699-e708 [FREE Full text] [doi: 10.1016/S2542-5196(21)00228-X] [Medline: 34627474]
8. Godfray HCJ, Aveyard P, Garnett T, Hall JW, Key TJ, Lorimer J, et al. Meat consumption, health, and the environment. Science 2018 Jul 20;361(6399):eaam5324. [doi: 10.1126/science.aam5324] [Medline: 30026199]

9. Rees JH, Bamberg S, Jäger A, Victor L, Bergmeyer M, Friese M. Breaking the habit: on the highly habitualized nature of meat consumption and implementation intentions as one effective way of reducing it. Basic Appl Soc Psych 2018 Apr 05;40(3):136-147. [doi: 10.1080/01973533.2018.1449111]
10. Bianchi F, Dorsel C, Garnett E, Aveyard P, Jebb SA. Interventions targeting conscious determinants of human behaviour to reduce the demand for meat: a systematic review with qualitative comparative analysis. Int J Behav Nutr Phys Act 2018 Oct 19;15(1):102 [FREE Full text] [doi: 10.1186/s12966-018-0729-6] [Medline: 30340498]
11. Bianchi F, Garnett E, Dorsel C, Aveyard P, Jebb SA. Restructuring physical micro-environments to reduce the demand for meat: a systematic review and qualitative comparative analysis. Lancet Planet Health 2018 Sep;2(9):e384-e397 [FREE Full text] [doi: 10.1016/S2542-5196(18)30188-8] [Medline: 30177007]
12. Chen Y, Perez-Cueto FJA, Giboreau A, Mavridis I, Hartwell H. The promotion of eating behaviour change through digital interventions. Int J Environ Res Public Health 2020 Oct 15;17(20):7488 [FREE Full text] [doi: 10.3390/ijerph17207488] [Medline: 33076239]
13. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 02;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
14. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction 2011 Feb;106(2):267-282. [doi: 10.1111/j.1360-0443.2010.03214.x] [Medline: 21083832]
15. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
16. Hedin B, Katzeff C, Eriksson E, Pargman D. A systematic review of digital behaviour change interventions for more sustainable food consumption. Sustainability 2019 May 08;11(9):2638 [FREE Full text] [doi: 10.3390/su11092638]
17. Frie K, Stewart C, Piernas C, Cook B, Jebb SA. Effectiveness of an Online Programme to Tackle Individual's Meat Intake through SElf-regulation (OPTIMISE): A randomised controlled trial. Eur J Nutr 2022 Aug 04;61(5):2615-2626 [FREE Full text] [doi: 10.1007/s00394-022-02828-9] [Medline: 35244757]
18. Research for the Future. National Institute for Health and Care Research. URL: https://www.researchforthefuture.org/ [accessed 2021-10-05]
19. Lacroix K, Gifford R. Reducing meat consumption: Identifying group-specific inhibitors using latent profile analysis. Appetite 2019 Jul 01;138:233-241. [doi: 10.1016/j.appet.2019.04.002] [Medline: 30965045]
20. Piazza J, Ruby MB, Loughnan S, Luong M, Kulik J, Watkins HM, et al. Rationalizing meat consumption. The 4Ns. Appetite 2015 Aug;91:114-128. [doi: 10.1016/j.appet.2015.04.011] [Medline: 25865663]
21. Stewart C, Frie K, Piernas C, Jebb SA. Development and reliability of the Oxford Meat Frequency Questionnaire. Nutrients 2021 Mar 12;13(3):922 [FREE Full text] [doi: 10.3390/nu13030922] [Medline: 33809192]
22. Food Standards Agency. Food Portion Sizes, 3rd Edition. Norwich, UK: TSO; 2002.
23. Fitt E, Mak TN, Stephen AM, Prynne C, Roberts C, Swan G, et al. Disaggregating composite food codes in the UK National Diet and Nutrition Survey food composition databank. Eur J Clin Nutr 2010 Nov;64 Suppl 3:S32-S36. [doi: 10.1038/ejcn.2010.207] [Medline: 21045846]
24. Stewart C, Piernas C. Evaluation of an online self-regulation intervention (OPTIMISE) to reduce meat consumption: a pre-post observational study. Statistical Analysis Plan. Open Science Framework. URL: https://osf.io/tfka6/ [accessed 2021-10-18]
25. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Vol 2 Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. Washington, DC: American Psychological Association; 2012:57-71.
26. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
27. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. J Med Internet Res 2014 Apr 15;16(4):e109 [FREE Full text] [doi: 10.2196/jmir.3084] [Medline: 24735567]
28. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. J Med Internet Res 2013 Aug 30;15(8):e162 [FREE Full text] [doi: 10.2196/jmir.2336] [Medline: 23996958]
29. Powell J, Williams V, Atherton H, Bennett K, Yang Y, Davoudianfar M, et al. Effectiveness and cost-effectiveness of a self-guided internet intervention for social anxiety symptoms in a general population sample: randomized controlled trial. J Med Internet Res 2020 Jan 10;22(1):e16804 [FREE Full text] [doi: 10.2196/16804] [Medline: 31821151]
30. Rise J, Sheeran P, Hukkelberg S. The role of self-identity in the theory of planned behavior: a meta-analysis. J Appl Soc Psychol 2010;40(5):1085-1105. [doi: 10.1111/j.1559-1816.2010.00611.x]

31. Carfora V, Caso D, Conner M. Correlational study and randomised controlled trial for understanding and changing red meat consumption: The role of eating identities. Soc Sci Med 2017 Feb;175:244-252 [FREE Full text] [doi: 10.1016/j.socscimed.2017.01.005] [Medline: 28129949]
32. Carfora V, Caso D, Conner M. Randomised controlled trial of a text messaging intervention for reducing processed meat consumption: The mediating roles of anticipated regret and intention. Appetite 2017 Oct 01;117:152-160 [FREE Full text] [doi: 10.1016/j.appet.2017.06.025] [Medline: 28651971]
33. Kassavou A, Wang M, Mirzaei V, Shpendi S, Hasan R. The association between smartphone app-based self-monitoring of hypertension-related behaviors and reductions in high blood pressure: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Jul 12;10(7):e34767 [FREE Full text] [doi: 10.2196/34767] [Medline: 35819830]
34. Bianchi F, Aveyard P, Astbury NM, Cook B, Cartwright E, Jebb SA. Replacing meat with alternative plant-based products (RE-MAPs): protocol for a randomised controlled trial of a behavioural intervention to reduce meat consumption. BMJ Open 2019 Jun 01;9(5):e027016 [FREE Full text] [doi: 10.1136/bmjopen-2018-027016] [Medline: 31154309]
35. Wolstenholme E, Carfora V, Catellani P, Poortinga W, Whitmarsh L. Explaining intention to reduce red and processed meat in the UK and Italy using the theory of planned behaviour, meat-eater identity, and the Transtheoretical model. Appetite 2021 Nov 01;166:105467 [FREE Full text] [doi: 10.1016/j.appet.2021.105467] [Medline: 34133978]
36. Verain MC, Bouwman EP, Galama J, Reinders MJ. Healthy eating strategies: Individually different or context-dependent? Appetite 2022 Jan 01;168:105759 [FREE Full text] [doi: 10.1016/j.appet.2021.105759] [Medline: 34662601]
37. Christensen H, Griffiths K. Reaching standards for dissemination: a case study. Stud Health Technol Inform 2007;129(Pt 1):459-463 [FREE Full text] [Medline: 17911759]
38. McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:55-62. [doi: 10.1192/bjp.185.1.55] [Medline: 15231556]
39. Bernecker SL, Williams JJ, Caporale-Berkowitz NA, Wasil AR, Constantino MJ. Nonprofessional peer support to improve mental health: randomized trial of a scalable web-based peer counseling course. J Med Internet Res 2020 Sep 21;22(9):e17164 [FREE Full text] [doi: 10.2196/17164] [Medline: 32955451]
40. Marteau TM. Towards environmentally sustainable human behaviour: targeting non-conscious and conscious processes for effective and acceptable policies. Philos Trans A Math Phys Eng Sci 2017 Jun 13;375(2095):20160371 [FREE Full text] [doi: 10.1098/rsta.2016.0371] [Medline: 28461435]

Abbreviations

NDNS: National Diet and Nutrition Survey
OPTIMISE: Online Programme to Tackle Individual's Meat Intake Through Self-regulation
RCT: Randomized Controlled Trial

Edited by G Eysenbach; submitted 18.02.22; peer-reviewed by N Pahlavani; comments to author 17.10.22; revised version received 06.11.22; accepted 30.11.22; published 12.12.22.

Please cite as:
Stewart C, Piernas C, Frie K, Cook B, Jebb SA
Evaluation of OPTIMISE (Online Programme to Tackle Individual's Meat Intake Through Self-regulation): Cohort Study
J Med Internet Res 2022;24(12):e37389
URL: https://www.jmir.org/2022/12/e37389
doi:10.2196/37389
PMID:36508245

Original Paper

Patient Trust in Physicians Matters—Understanding the Role of a Mobile Patient Education System and Patient-Physician Communication in Improving Patient Adherence Behavior: Field Study

Dezhi Wu[1], PhD; Paul Benjamin Lowry[2], PhD; Dongsong Zhang[3], PhD; Youyou Tao[4], PhD

[1]Department of Integrated Information Technology, University of South Carolina, Columbia, SC, United States

[2]Department of Business Information Technology, Virginia Tech, Blacksburg, VA, United States

[3]Department of Business Information Systems & Operations Management, The University of North Carolina at Charlotte, Charlotte, NC, United States

[4]Department of Information Systems and Business Analytics, Loyola Marymount University, Los Angeles, CA, United States

Corresponding Author:
Dezhi Wu, PhD
Department of Integrated Information Technology
University of South Carolina
550 Assembly Street
Columbia, SC, 29298
United States
Phone: 1 8037774691
Fax: 1 8037779564
Email: dezhiwu@cec.sc.edu

Abstract

Background: The ultimate goal of any prescribed medical therapy is to achieve desired outcomes of patient care. However, patient nonadherence has long been a major problem detrimental to patient health, and thus is a concern for all health care providers. Moreover, nonadherence is extremely costly for global medical systems because of unnecessary complications and expenses. Traditional patient education programs often serve as an intervention tool to increase patients' self-care awareness, disease knowledge, and motivation to change patient behaviors for better adherence. Patient trust in physicians, patient-physician relationships, and quality of communication have also been identified as critical factors influencing patient adherence. However, little is known about how mobile patient education technologies help foster patient adherence.

Objective: This study aimed to empirically investigate whether and how a mobile patient education system (MPES) juxtaposed with patient trust can increase patient adherence to prescribed medical therapies.

Methods: This study was conducted based on a field survey of 125 patients in multiple states in the United States who have used an innovative mobile health care system for their health care education and information seeking. Partial least squares techniques were used to analyze the collected data.

Results: The results revealed that patient-physician communication and the use of an MPES significantly increase patients' trust in their physicians. Furthermore, patient trust has a prominent effect on patient attitude toward treatment adherence, which in turn influences patients' behavioral intention and actual adherence behavior. Based on the theory of planned behavior, the results also indicated that behavioral intention, response efficacy, and self-efficacy positively influenced patients' actual treatment adherence behavior, whereas descriptive norms and subjective norms do not play a role in this process.

Conclusions: Our study is one of the first that examines the relationship between patients who actively use an MPES and their trust in their physicians. This study contributes to this context by enriching the trust literature, addressing the call to identify key patient-centered technology determinants of trust, advancing the understanding of patient adherence mechanisms, adding a new explanation of the influence of education mechanisms delivered via mobile devices on patient adherence, and confirming that the theory of planned behavior holds in this patient adherence context.

(J Med Internet Res 2022;24(12):e42941) doi:10.2196/42941

KEYWORDS

mobile health; mHealth; trust; patient adherence; mobile patient education system; MPES; patient-physician communication; theory of planned behavior; TPB; patient-centered care; mobile phone

Introduction

Background

The primary goal of any prescribed medical therapy is to achieve the desired medical outcomes of patient care, which requires a certain level of patient adherence, a critical measure of quality care. Previous studies have shown that nonadherence results in an economic burden of approximately US $100 to US $300 billion per year in the United States [1] owing to the costs of disease progression, readmissions, wasted resources, labor burden, and insurance costs [2], representing 3% to 10% of total health care costs in the United States [3,4]. Another meta-analysis of 79 studies across several countries showed that all-cause nonadherence costs ranged from US $5271 to US $52,341 per person [5]. In a medical context, adherence is defined as patients' behaviors that coincide with health care providers' health and medical advice, such as taking prescribed medication and following suggested diets [6-8]. For optimal therapeutic efficacy, adherence rates of >80% are needed, and for patients with certain more serious conditions, adherence rates of >95% are required [7,9]. Many health intervention programs have been implemented in different health care settings, but low success rates have persisted, imposing a major financial burden on the US health care system [5,10]. Studies have shown that the average adherence rate for long-term medication therapies is 40% to 50%, but the adherence rate for short-term therapies is 70% to 80% [11-13]. Even with clinical performance incentives and tremendous efforts regarding the development and examination of interventions, the current rate of long-term adherence to evidence-based medications in cardiometabolic diseases remains low [10]. Patients who comply with a treatment, even when the treatment is a placebo, have better health outcomes than those with poor adherence [14]. In contrast, patients who do not comply with the recommended treatments cause unsuccessful medical interventions and therapies, which exacerbate undesired health outcomes such as suboptimal therapeutic outcomes, delayed recovery, and more additive or aggressive treatments with the potential for more adverse events [4,15]. Thus, therapeutic adherence has been a long-standing topic of clinical concern for decades owing to the widespread nature of nonadherence.

Poor patient adherence can obscure a clinician's assessment of therapeutic effectiveness and result in avoidable hospitalization, increased mortality risk, and increased health care costs [5,8,16]. Furthermore, because of undetected or unreported therapeutic nonadherence, physicians may change the regimen, which may increase the cost or complexity of a treatment, thus further increasing the burden on patients. According to the Centers for Medicare and Medicaid Services, national prescription drug spending is expected to continue growing by approximately 6% per year from 2023 to 2028 [17]. Such nonadherence is extremely costly to global medical systems as it causes unnecessary complications and expenses [5]. Accordingly, a key goal of behavioral medicine is to find ways to increase patient adherence to prescribed treatments. Therefore, from the perspective of achieving desirable clinical and economic outcomes, the influential factors that contribute to patient adherence need to be examined and better understood for developing effective strategies to promote patient adherence. An understanding of the predictive value of these factors for patient adherence would also contribute positively to the overall planning of any disease management program.

The World Health Organization Multidimensional Adherence Model identifies five interrelated dimensions of patient medication adherence: (1) social and economic factors (eg, limited access to health care facilities), (2) health care system factors (eg, provider-patient relationship and providers' communication skills), (3) medical condition–related factors (eg, severity of symptoms), (4) therapy-related factors (eg, duration of a therapy), and (5) patient-related factors (eg, patient age, gender, and knowledge of a disease) [18,19]. Among the factors in the World Health Organization Multidimensional Adherence Model framework, considerable attention would need to be given to patients' relevant information and knowledge [19-21]. Health care providers can meet patients' information needs by reinforcing patient education on their treatment [19]. Educating patients about their disease status and general knowledge of their medications, as suggested in patient-centered health care, can also increase patient confidence, active participation, and patient adherence behavior [20,22]. Relatedly, a recent study revealed that personalized and repeated patient education interventions have modest efficacy in increasing patient adherence to medications [23].

However, patient education is not always *the more, the better* [20,24]. An *inverted U* relationship between knowledge and adherence has been found in adolescents [25]. Adolescent patients who know little about their therapies and illness are poor at adherence, whereas patients who are adequately educated about their disease and drug regimens are good at adherence; however, patients who know the lifelong adverse consequences might show poor medication adherence [26]. A recent clinical trial study showed that patient education significantly improved medication adherence but found no differences between single- and multicomponent education interventions [20]. Given the importance of patient education for patient adherence and the complexity of this relationship, further studies are needed to understand the underlying complex relationship between patient education and adherence to improve the quality of care. This knowledge gap is more meaningful, when mobile patient education programs are increasingly delivered to patients and are accessible anytime, anywhere through their mobile devices, since with traditional PC-based patient education programs, patients are primarily passive information receivers, have limited access to their health educational materials, and face difficulties in communicating with their physicians and care provider teams.

With the rapid advances and prevalence of the latest ubiquitous computing and mobile communication technologies, mobile

health (mHealth) systems have been partially or fully implemented and embedded in current health care systems to foster patient-centered care. As such, increasing research has started to explore whether mHealth technologies can help improve patients' adherence behavior. In this study, we define mHealth technologies as medical and public health practices and services, such as health care–related reminders, advice, and information delivered through mobile devices such as mobile phones [27,28]. Several studies have reported a positive relationship between the use of mHealth technologies and the increase in patients' adherence to medication [28-30], exercise advice [31,32], and a few other contexts such as dietary behavior [33,34]. Research has also shown that a key factor that influences patient adherence is patients' trust in their physicians and the quality of the relationships and communication between them [35-37]. Current literature also indicates that mHealth technologies can be used to increase patients' knowledge of medical therapies and, thus, improve patient-physician communication [28,38]. However, few studies have been conducted to systematically examine how mobile technologies for patient education can be leveraged to increase patients' trust in physicians and further improve patients' adherence behavior.

Objectives

Therefore, in this study, we aimed to fill such a knowledge gap to identify and empirically examine in-depth mechanisms through a lens of theory of planned behavior (TPB) on how and why such patient-physician trust is formed and leveraged by a mobile patient education system (MPES), leading to increased actual adherence to prescribed medical therapies. Thus, we conducted a field survey of 125 patients in multiple states in the United States who have used an innovative mHealth system for their health care education, information seeking, and communication with physicians. Our main finding was that patient-physician communication and the use of an MPES significantly increased patients' trust in their physicians, which further influenced patient attitude, intention, and actual behavior toward treatment adherence.

Methods

Research Model and Hypotheses

In this subsection, we propose our theoretical model and develop hypotheses to explain the role of an MPES in a health care setting, where an MPES affects patients' trust in physicians, which further influences their adherence. Our model is presented in Figure 1. First, we draw on the interpersonal trust literature and hypothesize the determinants of interpersonal trust. Specifically, we explain that one's general satisfaction and communication quality with one's physician helps form a patient's trust in physicians, whereas communication barriers with physicians decrease trust. Moreover, we specifically explain how the use of an MPES designed to increase patients' understanding of treatments may also increase trust. Finally, we draw on the well-known TPB [39] to explore how a patient's trust in physicians enhanced by an MPES may influence their treatment adherence, along with other factors that are derived from the TPB. On the basis of the existing literature, we also included some related covariates in our model to examine other factors that may affect patient adherence.

Trust has been widely studied and recognized as a cornerstone of effective patient-physician relationships [36,40-43]. In this study, we focus on interpersonal trust, which is defined as *the extent to which a person is confident in and willing to act on the basis of, the words, actions, and decisions of another* [44]. Interpersonal trust has been widely studied in the information systems literature [45,46]. In this context, we define trust as the acceptance of a vulnerable situation in which the patient believes that the physician will act in the patient's best interests and provide assistance and support for medical care and treatment [43,47]. The vulnerable situation and the need for trust are associated with being unhealthy, the information asymmetries of medical knowledge, and the uncertainty of risks regarding the intentions and competence of the physician [40].

Next, we focus on discussing 3 key factors and their underlying mechanisms associated with patients' trust: patients' general satisfaction, communication quality, and communication barriers between physicians and patients. Patients' general satisfaction reflects their perceptions and attitudes toward physicians and medical care in general [48]. Patients' general satisfaction with care signals good relationships between them and physicians [49]. It also reflects patients' perception of the physician's effective treatment in previous medical care, indicating that the physician will have the ability to provide high-quality medical care and treatment in the future. Hence, patients' general satisfaction affects their confidence and trust in the physician for medical care and treatment. Previous studies have also found that a higher level of patient satisfaction is associated with a higher level of patient-provider trust [49]. Thus, we hypothesize that an increase in patients' general satisfaction with their physicians is associated with an increase in their trust in their physicians (hypothesis 1).

Communication also plays an essential role in the patient-physician relationship, especially in the effectiveness of this relationship [4,36]. Effective, sufficient, and 2-way conversations between patients and physicians enable patients to decrease the information asymmetries that come from the nature of medical knowledge and facilitate the collaborative decision-making process toward patients' treatment plans [24,37,50], thus promoting patients' confidence and trust in their physicians. Indeed, empirical support has been found for the effect of communication (eg, discussing options and being open during communication) on trust [51]. Conversely, barriers to effective and sufficient communication between patients and physicians, such as 1-way conversations and small talk during communication, would lead to poor understanding of the benefits and risks associated with their condition, therapy, and treatment and a poor understanding of the proper use of the medication, which would decrease patients' confidence and trust in the physician. Previous studies have also confirmed this relationship. For example, some communication barriers may lead to a patient's low levels of trust in physicians [22], such as when physicians answer few questions or when patients find it difficult to understand a physician's language or writing. Too little time spent by physicians with patients is also found to threaten patients' motivation to maintain their therapeutic treatment plan,

which further leads to a low level of trust in physicians [52]. Thus, we hypothesize that an increase in patients' communication quality with their physicians is associated with an increase in their trust in their physicians (hypothesis 2) and that an increase in patients' communication barriers with their physicians is associated with a decrease in their trust in their physicians (hypothesis 3).

Our next hypothesis predicts that mobile patient education can help increase patients' trust in their physicians. Our contextual assumption for the design of this study was that patients would use an MPES to learn about the treatment that they were seeking in a just-in-time manner in their physicians' waiting room right before seeing a physician about the treatment. During this time, patients can learn the key terms, procedures, issues, risks, and benefits involved in a treatment and, thus, are able to communicate with a physician regarding their treatment plan with more knowledge and confidence. Furthermore, such an educational artifact allows basic questions to be answered ahead of time and, thus, enables patients to use the limited time with their physicians more effectively. Better and more effective communication enhances patients' confidence and trust in their physicians.

To further explain and justify this prediction, we used several lines of reasoning and evidence. First, one of the biggest problems in trust formation between patients and physicians is poor communication and misunderstandings between them because of information asymmetry [22,53,54]. On the basis of the assumption that information is imperfect and obtaining information can be costly, information asymmetry concerns are *critical when one party lacks information about the quality of another party or when one party is concerned about another party's behavioral tendencies* [55,56]. Information asymmetry is common in the health care sector, where physicians are equipped with their training and specialized knowledge but patients do not have similar training to physicians and usually have limited knowledge regarding their diagnoses and medical treatments [57]. Thus, we argue that an MPES can help patients educate themselves and reduce information asymmetry through the availability of relevant and timely medical information regarding their conditions and treatment plans. Enhanced knowledge and a shared understanding of the terms and issues involved in a treatment further lead to more effective communication and higher engagement during a visit. For example, by using an MPES in the waiting room, patients may have a better understanding of relevant medical information. With a more shared understanding of their condition and treatment, patients will have a 2-way conversation with their physicians to have their questions directly answered. Previous studies have also found that shared values would be tied to the production of trust, and patients who felt well informed had a generally high level of trust in their physicians [54,58].

Second, patients judge the competencies of physicians in a multifaceted way [58]. On the one hand, patients do not expect a physician to know everything. In contrast, patients seek information to judge the quality of the information they received from their physicians. Namely, they welcome evidence-based information as this helps them make self-determined decisions regarding their preferences. The possibility of making informed decisions based on rational criteria is perceived as a prerequisite for developing trust [58]. Patients can use the patient education material that they receive on their mobile devices to cross-validate the information they find from other sources and, thus, take an active role in the clinical decision-making process, helping them develop trust in their physicians [59].

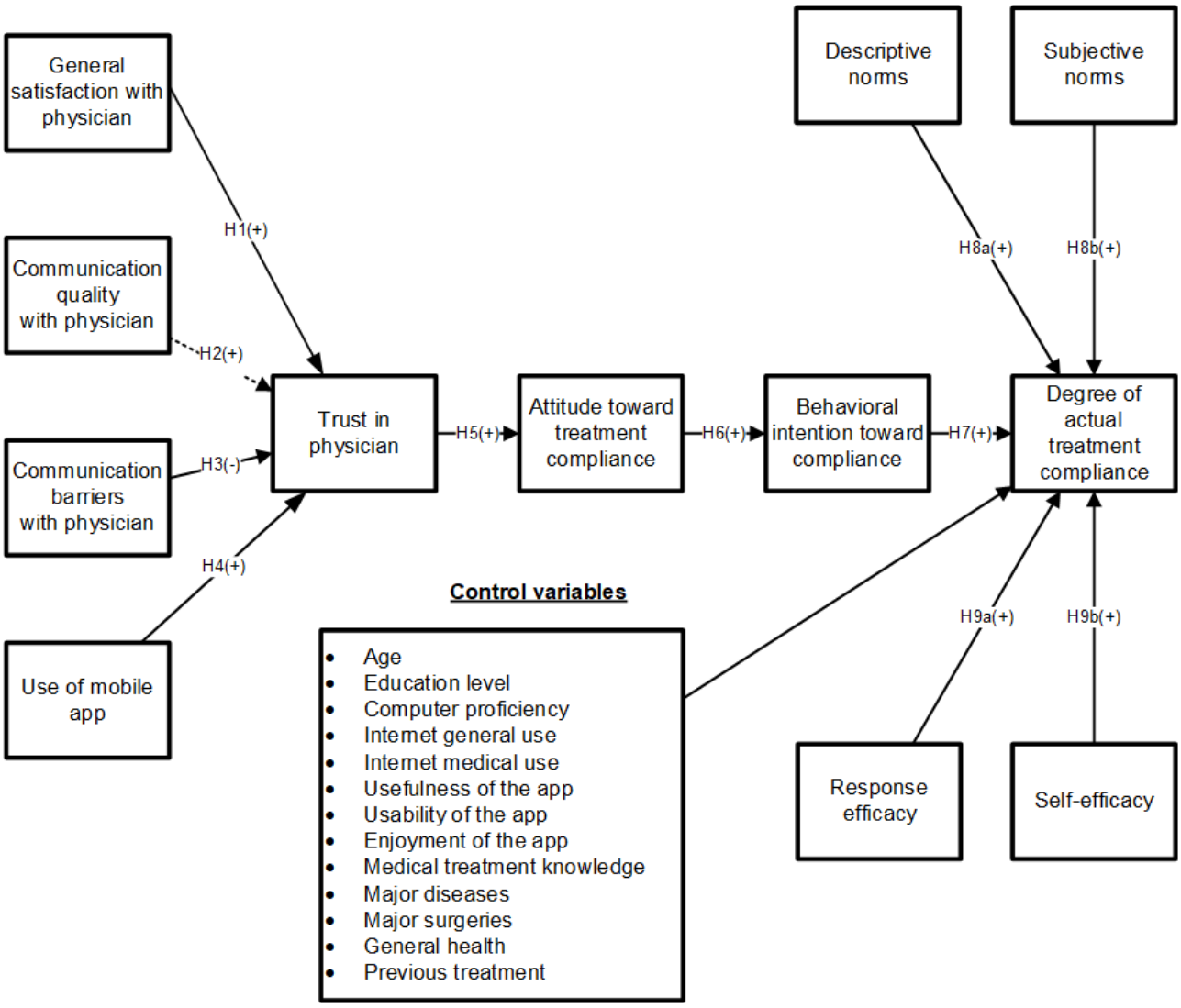
Third, a key issue in the relationship between patients and physicians is too little time spent together, which undermines communication and trust [52]. Given the limited and costly time available during patients' visits, time spent on ineffective communication would further leave less time for physicians to discuss treatment plans and address patients' needs. Using an MPES before a time-constrained visit would empower patients with necessary and relevant medical information and allow such a visit to be more effective for both the patient and physician.

Fourth, we also assert that physicians who provide such an MPES with personalized patient materials and tools in their waiting rooms not only better prepare their patients to communicate with them about their treatments but also provide a positive signal of service quality and empathy that can facilitate patient-physician communication and trust formation [60]. Consequently, a trusting patient-provider relationship enabled by an MPES may further improve patient care quality.

In summary, using an MPES is predicted to reduce information asymmetries between patients and physicians, empower patients taking an active role in clinical decision-making processes, facilitate effective communication in a time-constrained visit, and send a message of care and empathy from the physicians, thus leading to a better patient-physician relationship and enhancing patients' confidence and trust in their physicians. Therefore, we propose that the level of trust is likely to be higher when the level of patient use of a mobile patient education app is higher. As such, we posit that an increase in patients' use of a mobile patient education app designed to increase their knowledge of a specific medical treatment will increase their trust in their physicians (hypothesis 4).

A trustful relationship between patients and health care service providers is key to patient adherence. In particular, a healthy relationship is established based on patients' trust in physicians and empathy from physicians [37,60,61]. Changing patients' attitudes and beliefs to improve their adherence behavior is more likely when there is an elevated level of patient trust. Trust in a physician correlates positively with patients' perceived effectiveness of care, acceptance of new medications, and intention to follow physician instructions [37,50]. Thus, we hypothesize that an increase in patients' trust in their physicians is associated with an increase in their positive attitudes toward treatment adherence (hypothesis 5).

Figure 1. The research model. H1: hypothesis 1; H2: hypothesis 2; H3: hypothesis 3; H4: hypothesis 4; H5: hypothesis 5; H6: hypothesis 6; H7: hypothesis 7; H8a: hypothesis 8a; H8b: hypothesis 8b; H9a: hypothesis 9a; H9b: hypothesis 9b.



We used the TPB to account for the formation of attitudes from beliefs, norms, and self-efficacy, which can then be used to predict subsequent behaviors [39]. Fundamental to the TPB is the idea that attitudes are the drivers of behavioral intentions, and behavioral intentions are the drivers of actual behaviors [39,62-64]. We leveraged and applied these concepts and predictions to the health care adherence context. In our model, an attitude toward treatment adherence is the primary driver of behavioral intention toward treatment adherence, which then drives actual treatment adherence behaviors. We assume that the TPB holds in this health care context; thus, we posit that an increase in positive attitudes toward treatment adherence is associated with an increase in behavioral intention toward treatment adherence (hypothesis 6) and that an increase in behavioral intention toward treatment adherence is associated with an increase in the degree of actual treatment adherence (hypothesis 7).

Next, we followed the TPB assumption that normative beliefs will influence actual adherence behaviors. Normative beliefs represent a person's perceived social pressure to comply with a recommendation as informed by their valued social referents for the context [62,65]. Normative beliefs are also known as social influence, which comprises subjective and descriptive norms [66]. Following adherence literature [66-68], subjective norms in our context represent the degree to which patients believe that other key people (eg, family, friends, and coworkers) in their lives want them to comply with a treatment recommendation. Descriptive norms represent a patient's beliefs about what is commonly done by most patients or the public in terms of adherence to a specific medical recommendation.

According to the TPB, norms affect individuals' intentions and behaviors [39,62,69]. In a security context, Siponen [70] found that norms work because of the desire to conform to a group to which one belongs; this was also confirmed by Mishra et al [71]. Psychology researchers have long proposed that conformity to groups is attributable to norms and the pressures that norms place on individuals within a group [72-74]. Assuming that these social norms also play a role in patients' decisions regarding medical recommendation adherence, we posit that an increase in descriptive norms toward a treatment is associated with an increase in the degree of actual treatment adherence (hypothesis 8a) and that an increase in subjective norms toward a treatment is associated with an increase in the degree of actual treatment adherence (hypothesis 8b).

Next, we followed the TPB assumption that one's actual behaviors will be influenced by one's efficacy. As a long-established component of the TPB, self-efficacy is highly important in the medical treatment adherence context as it covers patients' basic self-assessment regarding their ability to effectively follow medical advice and whether they believe that a recommended treatment is efficacious [4,30]. In the TPB, efficacy is conceptualized as response efficacy and self-efficacy. On the basis of the literature on adherence [18,65-67], we define self-efficacy in our context as a patient's judgment of their personal ability, competency, and knowledge in complying with a recommended medical treatment. Similarly, response efficacy is a patient's judgment of the likely effectiveness and positive outcomes associated with a recommended medical treatment. We assume that these efficacy judgments play a role in patients' adherence to medical recommendations and hypothesize that an increase in response efficacy toward a treatment is associated with an increase in the degree of actual treatment adherence (hypothesis 9a) and an increase in self-efficacy toward a treatment is associated with an increase in the degree of actual treatment adherence (hypothesis 9b).

Ethical Considerations

The MPES on which we focused in this study was codeveloped by the first author's research group and ABC Company (anonymized), which is a software company whose main products are health care systems aiming to address the communication and trust issues between patients and physicians to improve patient adherence behavior. The MPES has been successfully sold and deployed in many clinics and hospitals in North America, South America, and Asia. Patient users can access patient education materials in physicians' clinics or anywhere else through different mobile devices. Its web portal interface is easy to navigate, and the educational contents are customized according to each patient's health situation.

This study was approved by the institutional review board of the Southern Utah University (approval number: 15-052013). We worked with the ABC Company to obtain their support for conducting this research with their patients. With the assistance of the company's attorney, we carefully followed the US Health Insurance Portability and Accountability Act's (HIPAA) Privacy, Security, and Breach Notification Rules to protect patients' rights. Finally, after obtaining all approvals, we were able to post a web-based flyer on the ABC Company's patient web portal to invite interested patients who had used the MPES to participate in our field study. We provided a US $10 honorarium for each survey respondent who provided valid and complete responses.

Data and Sample

We conducted a field study with real patients from multiple states in the United States to test our research model. These patients were able to use an MPES designed to improve patient education and experience at their physicians' clinics or hospitals. Patient participation was completely voluntary. They were given web-based instructions to fill out a web-based questionnaire distributed solely inside the MPES, where only active patient users could see our project flyer and answer our questions on their perceptions and assessments of using the MPES and its influence on their adherence behavior. During a period of 2 and a half months, after both plastic surgery and obstetrics versions of the MPES were launched, we received a total of 126 patient responses. We excluded 0.8% (1/126) of responses from a male patient from further data analysis because most questions were unanswered, resulting in 125 valid responses, all of which were not surprisingly from female patients, adequately representing the patient population that we reached out to. After the initial data collection, several duplicate responses were identified and removed before data analysis.

After the initial development of the questionnaire, we made it accessible on the web. We then circulated it among 26 senior students at a US university to obtain feedback on the relevance and clarity of the survey questions and on whether the web-based questionnaire could be accessed properly through different mobile devices, such as iPads, iPhones, and other types of smartphones and tablets. As we planned to deploy the questionnaire on the web, the first author also conducted a 20-minute face-to-face meeting with each student to verify the clarity of the web-based questionnaire instructions in that no face-to-face contact was expected to take place between the researchers and the actual patient respondents. According to the feedback that we obtained from these pilot sessions, we were able to further refine a few ambiguous questions and adjust the web-based questionnaire interface to better fit heterogeneous mobile platforms.

Of the 125 valid survey respondents, 110 (88%) patients were from the plastic surgery field, 9 (7.2%) were from obstetrics, and 6 (4.8%) were from other medical fields. All the respondents (125/125, 100%) were female. The average age was 39.6 (SD 12.9) years.

Measures

For the study constructs and measures, we adapted existing validated psychometric scales and measurement items from established research. We then tailored the questions to fit the context of this study. Multimedia Appendix 1 details the key constructs and associated detailed questions found in the questionnaire. Multimedia Appendix 2 details the factor loadings as well as the means and SDs of each factor.

Items for a patient's general satisfaction with physicians were taken from the *overall satisfaction* dimension of the patient satisfaction questionnaire scale for a specific physician designed by Ware Jr et al [48]. Items for communication barriers and constructs regarding quality of communication with physicians were adapted and modified from the scale developed by Steine et al [75]. Items related to patient trust in physicians were adopted from the study by Hall et al [76]. Use of the app was customized to the medical context based on a self-reported internet use measure [77]. Subjective and descriptive norm constructs were modified based on validated instruments originally developed by Herath and Rao [66]. The response efficacy and self-efficacy constructs were adapted from the scale developed by Workman et al [78]. Attitude toward treatment adherence, intention toward treatment adherence, and degree of actual adherence constructs were measured based on modifications of similar measures from Bulgurcu et al [65] and Hu et al [79].

Statistical Analysis

As all our constructs were reflective and our research model contained both first-order and second-order constructs, partial least squares path modeling was used to examine our research model [80]. Partial least squares has been suggested for testing novel propositions with limited previous theoretical development [81-83], which is the nature of this study. Namely, we used SmartPLS software (version 3.0; SmartPLS GmbH) [84] to examine our research model.

In this study, we used the marker variable technique suggested by Lindell and Whitney [85] and Malhotra et al [86] to examine possible common method bias as self-reported survey data may inflate variable correlations. We selected a marker variable, organizational procedural justice [87], which is theoretically irrelevant to the context of this study. After computing the average correlation between the marker variable *organizational procedural justice* and the 12 principal constructs, which was 0.17 (average *P* value=.20), we confirmed that common method bias was not a major concern in this study.

Results

Descriptive Analysis

Table 1 summarizes the descriptive statistics related to the demographic information of the participants.

Table 1. Demographic information of the participants (N=125).

Variable and category	Participants, n (%)
Sex	
Female	125 (100)
Male	0 (0)
Age (years)	
<20	2 (1.6)
20 to 29	30 (24)
30 to 39	33 (26.4)
40 to 49	35 (28)
50 to 59	14 (11.2)
≥60	11 (8.8)
Education level	
Lower than high school or secondary school	0 (0)
High school or secondary school	21 (16.8)
Some university but had not completed a degree	33 (26.4)
Associate degree	16 (12.8)
Bachelor's degree	38 (30.4)
Master's degree	13 (10.4)
Doctorate or PhD	4 (3.2)
Computer proficiency	
Poor	0 (0)
Fair	7 (5.6)
Good	28 (22.4)
Very good	53 (42.4)
Excellent	37 (29.6)
Internet general use: how often do you use the internet?	
Never	0 (0)
Seldom	0 (0)
Occasionally	7 (5.6)
Frequently	14 (11.2)
Always or every day	104 (83.2)
Internet medical use: how often do you use internet sites such as WebMD to learn about medical information?	
Never	2 (1.6)
Seldom	18 (14.4)
Occasionally	54 (43.2)
Frequently	46 (36.8)
Always or every day	5 (4)

Measurement Model

We first tested internal consistency and then examined convergent and discriminant validity of the measurement items. Composite reliability and average variance extracted (AVE) [88] were computed. Both composite reliability and AVE reached a satisfactory level (Table S1 in Multimedia Appendix 2), as suggested by Fornell and Larcker [88] that composite reliability should be >0.70 and AVE should be >0.50. As such, the internal consistency, reliabilities, and convergent validity were confirmed.

Furthermore, to verify discriminant validity, we also computed the square root of the AVE for all latent variables and compared

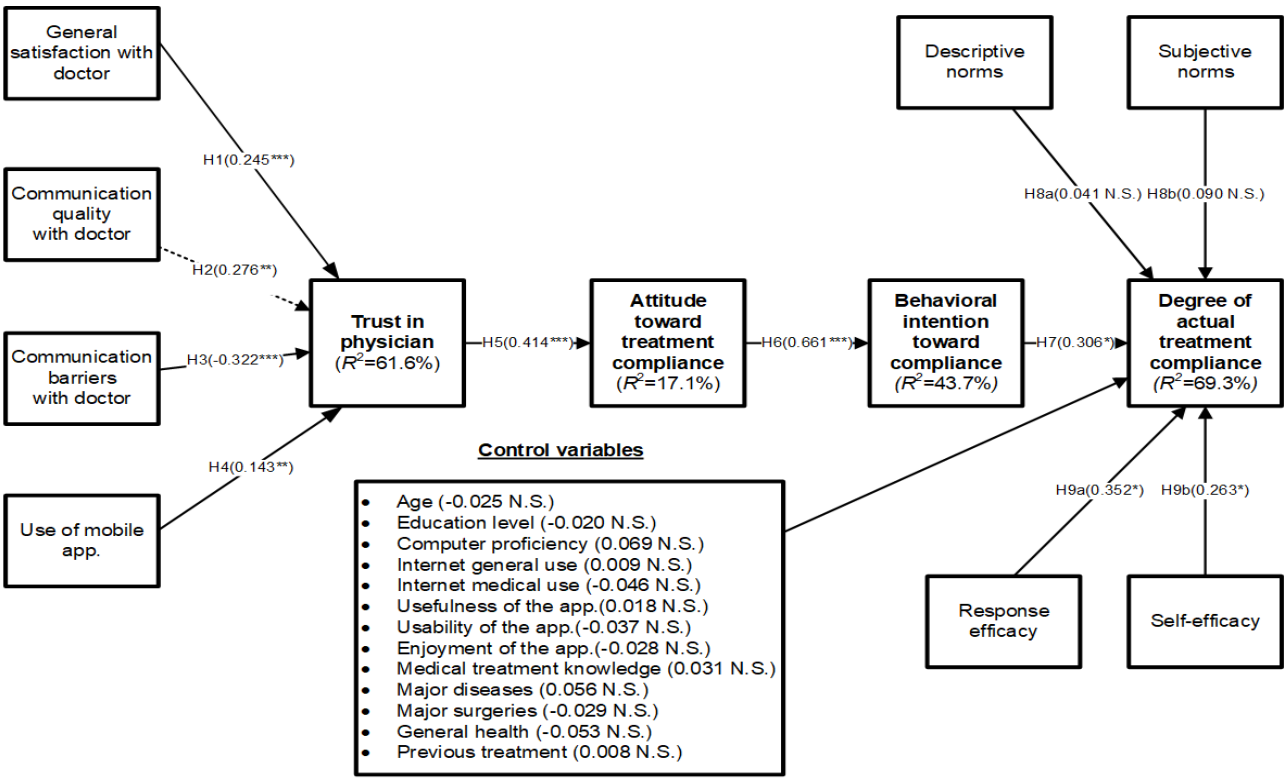
them against their correlations with other constructs [88]. Table S1 in Multimedia Appendix 2 demonstrates that all the square roots of the AVE values are greater than their correlations with any other constructs; thus, discriminant validity was also confirmed [89].

As all factor measures loaded highly (P>.50) on their associated latent constructs [90], we were able to confirm both the convergent and discriminant validity of this study. As Table S2 in Multimedia Appendix 2 shows, all items are >0.70 on their targeted constructs, which are much higher than the suggested threshold (P>.50) and other cross-loadings. Therefore, the results support convergent and discriminant validity [91].

Structural Model

In this study, we selected and examined several control variables related to our dependent variable, *degree of actual treatment adherence*, which were age, education level, computer proficiency, internet general use, internet medical use, usefulness of the app, usability of the app, enjoyment of the app, medical treatment knowledge, major diseases, major surgeries, general health, and previous treatment. However, none of them were statistically significant, as illustrated in Figure 2. Next, we assessed the structural model to examine the path coefficients (β). We used the bootstrapping method with 500 resamples to compute the statistical significance levels of the parameter estimates. Figure 2 depicts the results. Hypotheses 1 to 4 theorized the factors that influence patient trust in physicians. General satisfaction with a physician positively affected patient trust in the physician (β=.245; P<.001), supporting hypothesis 1. Communication quality with a physician significantly influenced patient trust in the physician (β=.276; P<.01), supporting hypothesis 2. Communication barriers with a physician were negatively related to patient trust in the physician (β=−0.322; P<.001), supporting hypothesis 3. The use of mobile education apps was found to affect patient trust in a physician significantly and positively (β=.143; P<.01), supporting hypothesis 4. Furthermore, patient trust in physicians had a highly positive relationship with attitude toward treatment adherence (β=.414; P<.001), supporting hypothesis 5. Finally, we found that attitude toward treatment adherence was positively related to behavioral intention toward adherence (β=.661; P<.001), which also significantly influenced the degree of actual treatment adherence (β=.306; P<.05), supporting hypotheses 6 and 7, respectively. The data results also indicated that response efficacy (β=.352; P<.01) and self-efficacy (β=.263; P<.05) significantly influenced the degree of actual treatment adherence, supporting hypotheses 9a and 9b, respectively. However, subjective and descriptive norms had no influence on patients' actual treatment adherence behavior. Thus, hypotheses 8a and 8b were not supported.

Figure 2. Results of research model testing. *P<.05, **P<.01, ***P<.001; H1: hypothesis 1; H2: hypothesis 2; H3: hypothesis 3; H4: hypothesis 4; H5: hypothesis 5; H6: hypothesis 6; H7: hypothesis 7; H8a: hypothesis 8a; H8b: hypothesis 8b; H9a: hypothesis 9a; H9b: hypothesis 9b; N.S.: nonsignificant.



Discussion

Principal Findings

In this paper, we aimed to systemically examine how and why patient-physician trust is formed and how patient education delivered via an MPES influences patients' trust and patient adherence. We leveraged the TPB along with fundamental concepts of trust to propose a model that explains how an MPES, along with other factors, can help foster trust of patients in physicians and eventually increase treatment adherence. We first confirmed that general satisfaction with physicians, communication quality, and use of an MPES jointly facilitate and foster patients' trust in physicians, whereas communication barriers decrease trust. We also found that patients' trust in their physicians was indeed a significant determinant of positive attitude formation. We found compelling evidence for our expanded model based on the TPB. Attitudes toward treatment adherence were positively related to intentions toward treatment adherence, and these intentions were positively related to actual treatment adherence behaviors. In addition to behavioral intention toward adherence, we found that response efficacy and self-efficacy enhance actual treatment adherence behaviors. A key exception in our model was the insignificant influence of social norms (ie, subjective and descriptive norms). This was particularly surprising in the context of plastic surgery visits, in which we expected social norms to play a stronger role. As most of our study participants were patients of plastic surgery who primarily sought treatment to improve their physical appearance instead of for medical reasons, the insignificant results associated with social norms were likely caused by their privacy concerns, which requires further research. Our research model explained 61.9% of the variance (R^2) in patient trust in physicians, 17.1% of the attitude toward treatment adherence, 43.7% of the behavioral intention toward adherence, and 69.3% of the variance of patients' actual adherence behavior.

Comparison With Prior Work

First, as one of the first studies that systematically examine the relationship between the use of mobile technologies and patients' trust in physicians in the context of patients being active users of technologies, we found that patient use of an MPES positively influences patient trust in physicians. Previous studies have examined patients' trust in technologies used solely by providers, such as electronic health records, electronic monitors, and web-based health communities supported by the internet [92-94]. In most previous studies, patients had limited control over the technologies and were passive users of them [92]. In our study, patients were active users of technologies and had control over how and when to use and access the content on mobile devices. In current patient-centered health care, the role of patients has evolved to becoming those active partners; thus, it is crucial to consider the influence of patients' use of health care technology in this relationship when mobile technologies become more prevalent ways of conducting patient outreach and intervention programs. Our study enriches the trust literature by addressing the call to identify key determinants of the use of patient-centered mobile technologies on trust.

Second, our study proposed an integrated model to explain how communication, patient satisfaction, MPES use, and trust foster the degree of actual treatment adherence. We examined these 4 factors in the patient adherence context. We found that communication, patient satisfaction, and MPES use jointly influence trust, which further fosters actual patient adherence. Overall, the integration of factors from multiple streams of the literature to explain patient adherence behavior is one of our core theoretical contributions, especially from a TPB theoretical lens. This finding advances our understanding of underlying patient adherence mechanisms.

Third, we examined an MPES in a clinical setting. Although the focus of our study, patient education mechanisms, is not new to the patient adherence literature, previous relevant patient adherence research has not examined patient education mechanisms delivered through mobile systems. Given the popularity and availability of mobile devices, patient education is increasingly delivered and communicated through mobile devices. Previous patient education studies have primarily focused on engaging patients in mHealth intervention programs to improve their adherence behaviors regarding medication [28-30], exercise advice [31,32], and dietary behavior [33,34] and have mostly studied 1-way communications between patients as passive users and their physicians. However, none of these studies explored the role of personalized patient education intervention programs in an mHealth environment to systematically understand how to establish and improve patients' trust in physicians, which is critical to improving patient-centered care. Our study provides an in-depth understanding of the influence of patient education mechanisms delivered through an MPES on patient adherence.

Finally, we extended and empirically tested the TPB in the context of patient adherence. From a theoretical standpoint, few empirical studies have examined the TPB in the context of patient adherence [30]; thus, our study extended the TPB to an mHealth context and provided empirical evidence on how an MPES can leverage patient trust in physicians to improve patient adherence, confirming that the TPB also holds in this context.

Strengths

It is critical for physicians to understand how to enhance patient adherence to their treatment recommendations. Our study indicates that an approach that may help is for physicians to proactively leverage mobile technologies such as an MPES to enhance the provider-patient relationship and foster patient trust in physicians. We recommend that physicians consider the trust-building process both on the web and in the office. Web-based trust can be developed by providing quality apps for patients to adopt and use while considering patients as active partners in the care process. Offline trust can be built through 2-way effective conversations with patients by addressing their personalized treatment needs. This study also highlights the importance of choosing the appropriate mobile technologies and apps for patients to use given that a message of care and empathy to patients disseminated through an MPES can further increase patients' trust in physicians and enhance patient adherence.

Our results also imply that additional changes to clinical workflows in hospitals and clinics may enhance patient-centered care. For example, the Mayo Clinic, as the leading hospital system in the United States, has implemented a secure patient message portal system to improve communication quality between physicians and patients, patient engagement, and patient-centered care [95]. Moreover, in practice, teams of caregivers—including hospitals, clinicians, nursing practitioners, and physician assistants—may also need to find effective ways to balance their main workload to take care of patients and handle increasingly growing communication loads with their patients to achieve the desired clinical operation efficiency and higher care quality driven by improved patient adherence. Thus, our study provides in-depth mechanisms to further this area of health care practice to not only foster patient engagement and communication but also establish physician-patient trust, which is critical to patient adherence and care quality.

Limitations

Our study has several limitations owing to the nature of the field study but also provides new opportunities for future research. First, this field study was restricted to physicians who had adopted the MPES in their practices and were willing to offer us access to their patients. Although all valid survey respondents were female, they were representative of the population in the related medical practices—plastic surgery and obstetrics.

Second, after going through many complicated legal and research coordination processes, we were only allowed to conduct the study with patients who had tried the MPES; thus, we were not able to reach patients who were still using traditional patient education systems as a control group. Patient adherence behavior has not been well studied across populations, diseases, and settings, thus making it difficult for health professionals and patients to know which strategies work and which do not [96]. We suggest that researchers further study the use of MPES in this context but involve other types of patients—including a balance of gender—such as those with chronic diabetes, mental illness, or cardiovascular disease who have serious adherence challenges in an mHealth setting.

Third, according to the behavior model of persuasive design by Fogg [97] and more personally controlled help-seeking features suggested by Lau et al [98], it would be valuable to conduct a longitudinal field study to observe patients' adherence behaviors during their different treatment stages in comparison with a control group. Moreover, we may also explore how different types of mobile interface designs and new technologies such as radio-frequency identification and the Internet of Things affect both patients' adherence behavior and physicians' decision-making processes [99]. Conversely, as mHealth systems represent innovative technology, many hospitals and physicians' offices have not adopted such systems to benefit their patients. However, how physicians can make more informed decisions to address patients' nonadherence issues through an MPES more effectively warrants further investigation.

Conclusions

In summary, extensive research has been conducted that examines factors associated with patient adherence, some of which has examined the relationship between patient education and adherence. However, as highlighted by extant literature, the underlying relationship between patient education and adherence is complex, and no studies to date have been conducted that explore and explain the underlying mechanisms of patient education delivered through mobile devices on patient adherence [24,100]. Achieving a more in-depth understanding of the effects of these mechanisms on adherence can have theoretical and practical implications on how to leverage an MPES to improve patient adherence, which in turn may improve health care outcomes. Thus, our study aimed to bridge this compelling knowledge gap.

In this study, our MPES provided additional patient care at the physician's office or clinic through a real-time mobile personalized patient education intervention program, which enabled 2-way physician-patient communication beyond the patients' in-person office visits. Our study participants were active patient users of the MPES, on which they could access their individual patient education materials and directly interact with their physicians and caregiver teams on the web. The results of our study imply that the MPES can be effectively leveraged by physicians' offices or clinics for more seamless high-quality care. It can also be a trade-off for physicians' offices or clinics to handle additional workload to provide more personalized services to their patients through an MPES. Nevertheless, our study findings indicate that the extended service lines provided on the MPES beyond regular in-person office visits may significantly improve patient-physician communication quality and increase patients' trust in their physicians, thus leading to more optimal health outcomes such as enhanced patient adherence to their therapy or treatment plans.

In conclusion, our study is one of the first that examines the relationship between patients who actively use an MPES and their trust in their physicians. This study contributes to this context by (1) enriching the trust literature addressing the call to identify key patient-centered technology determinants of trust, (2) advancing the understanding of patient adherence mechanisms, (3) adding a new explanation for the influence of education mechanisms delivered through mobile devices on patient adherence, and (4) confirming that the TPB holds in this patient adherence context.

Acknowledgments

The authors would like to acknowledge the funding support provided by the University of South Carolina (USC), Columbia, South Carolina, United States (grant 80002838), and partial support from the USC Big Data Health Science Center, a USC excellence initiative program (grant BDHSC-2021-14 and BDHSC-2022); the USC Advancing Chronic Care Outcome through

Research and Innovation Center (ACORN-2022); and a Utah Science Technology and Research Initiative grant (USTAR-TCG). The content is solely the responsibility of the authors and does not necessarily represent the views of the funding agencies.

Data Availability

Access to the data set analyzed in this study may not be available to researchers because of Health Insurance Portability and Accountability Act and institutional review board constraints.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient survey instrument and questions and measures for the theoretical model.
[DOCX File , 34 KB - jmir_v24i12e42941_app1.docx]

Multimedia Appendix 2
Measurement model statistics and validity details.
[DOCX File , 40 KB - jmir_v24i12e42941_app2.docx]

References

1. Neiman AB, Ruppar T, Ho M, Garber L, Weidle PJ, Hong Y, et al. CDC grand rounds: improving medication adherence for chronic disease management - innovations and opportunities. MMWR Morb Mortal Wkly Rep 2017 Nov 17;66(45):1248-1251 [FREE Full text] [doi: 10.15585/mmwr.mm6645a2] [Medline: 29145353]
2. The Healthcare Costs of Poor Patient Compliance. Oxitone. 2022 Jul 27. URL: https://www.blog.oxitone.com/2021/04/23/poor-patient-compliance/ [accessed 2022-08-30]
3. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [doi: 10.1097/01.mlr.0000114908.90348.f9] [Medline: 15076819]
4. Iuga AO, McGuire MJ. Adherence and health care costs. Risk Manag Healthc Policy 2014 Feb 20;7:35-44 [FREE Full text] [doi: 10.2147/RMHP.S19801] [Medline: 24591853]
5. Cutler RL, Fernandez-Llimos F, Frommer M, Benrimoj C, Garcia-Cardenas V. Economic impact of medication non-adherence by disease groups: a systematic review. BMJ Open 2018 Jan 21;8(1):e016982 [FREE Full text] [doi: 10.1136/bmjopen-2017-016982] [Medline: 29358417]
6. DiMatteo MR, Giordani PJ, Lepper HS, Croghan TW. Patient adherence and medical treatment outcomes: a meta-analysis. Med Care 2002 Sep;40(9):794-811. [doi: 10.1097/00005650-200209000-00009] [Medline: 12218770]
7. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
8. Khunti K, Seidu S, Kunutsor S, Davies M. Association between adherence to pharmacotherapy and outcomes in type 2 diabetes: a meta-analysis. Diabetes Care 2017 Nov;40(11):1588-1596. [doi: 10.2337/dc16-1925] [Medline: 28801474]
9. Haynes RB, Taylor DW, Sackett DL, Gibson ES, Bernholz CD, Mukherjee J. Can simple clinical measurements detect patient noncompliance? Hypertension 1980;2(6):757-764. [doi: 10.1161/01.hyp.2.6.757] [Medline: 7007235]
10. Lauffenburger JC, Haff N, McDonnell ME, Solomon DH, Antman EM, Glynn RJ, et al. Exploring patient experiences coping with using multiple medications: a qualitative interview study. BMJ Open 2021 Nov 22;11(11):e046860 [FREE Full text] [doi: 10.1136/bmjopen-2020-046860] [Medline: 34810179]
11. DiMatteo MR. Patient adherence to pharmacotherapy: the importance of effective communication. Formulary 1995 Oct;30(10):596-605. [Medline: 10151723]
12. Sabaté E. Adherence to long-term therapies : evidence for action. World Health Organization. Geneva, Switzerland: World Health Organization; 2003. URL: https://apps.who.int/iris/handle/10665/42682 [accessed 2022-09-10]
13. Morris LS, Schulz RM. Patient compliance--an overview. J Clin Pharm Ther 1992 Oct;17(5):283-295. [doi: 10.1111/j.1365-2710.1992.tb01306.x] [Medline: 1464632]
14. Horwitz RI, Horwitz SM. Adherence to treatment and health outcomes. Arch Intern Med 1993 Aug 23;153(16):1863-1868. [Medline: 8250647]
15. Harnett J, Wiederkehr D, Gerber R, Gruben D, Bourret J, Koenig A. Primary nonadherence, associated clinical outcomes, and health care resource use among patients with rheumatoid arthritis prescribed treatment with injectable biologic disease-Modifying antirheumatic drugs. J Manag Care Spec Pharm 2016 Mar;22(3):209-218. [doi: 10.18553/jmcp.2016.22.3.209] [Medline: 27003550]
16. Wroth TH, Pathman DE. Primary medication adherence in a rural population: the role of the patient-physician relationship and satisfaction with care. J Am Board Fam Med 2006;19(5):478-486 [FREE Full text] [doi: 10.3122/jabfm.19.5.478] [Medline: 16951297]

17. NHE Fact Sheet - Projected NHE 2019-2028. Centers for Medicare & Medicaid Services. 2022. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet [accessed 2022-09-20]
18. Mondesir FL, Levitan EB, Malla G, Mukerji R, Carson AP, Safford MM, et al. Patient perspectives on factors influencing medication adherence among people with coronary heart disease (CHD) and CHD risk factors. Patient Prefer Adherence 2019;13:2017-2027 [FREE Full text] [doi: 10.2147/PPA.S222176] [Medline: 31819383]
19. 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]
20. Taibanguay N, Chaiamnuay S, Asavatanabodee P, Narongroeknawin P. Effect of patient education on medication adherence of patients with rheumatoid arthritis: a randomized controlled trial. Patient Prefer Adherence 2019 Jan 11;13:119-129 [FREE Full text] [doi: 10.2147/PPA.S192008] [Medline: 30666095]
21. Moinul P, Barbosa J, Qian J, Chen ML, Mohaghegh M, Kaur H, et al. Does patient education improve compliance to routine diabetic retinopathy screening? J Telemed Telecare 2020 Apr;26(3):161-173. [doi: 10.1177/1357633X18804749] [Medline: 30373446]
22. Rubin RR. Adherence to pharmacologic therapy in patients with type 2 diabetes mellitus. Am J Med 2005 May;118 Suppl 5A:27S-34S. [doi: 10.1016/j.amjmed.2005.04.012] [Medline: 15850551]
23. Kini V, Ho PM. Interventions to improve medication adherence: a review. JAMA 2018 Dec 18;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
24. Jin J, Sklar GE, Min Sen Oh V, Chuen Li S. Factors affecting therapeutic compliance: a review from the patient's perspective. Ther Clin Risk Manag 2008 Mar;4(1):269-286 [FREE Full text] [doi: 10.2147/tcrm.s1458] [Medline: 18728716]
25. Wiese CW, Tay L, Duckworth AL, D'Mello S, Kuykendall L, Hofmann W, et al. Too much of a good thing? Exploring the inverted-U relationship between self-control and happiness. J Pers 2018 Jun;86(3):380-396 [FREE Full text] [doi: 10.1111/jopy.12322] [Medline: 28480971]
26. Hamburg BA, Inoff GE. Relationships between behavioral factors and diabetic control in children and adolescents: a camp study. Psychosom Med 1982 Sep;44(4):321-339. [doi: 10.1097/00006842-198209000-00002] [Medline: 7146242]
27. Galligioni E, Piras EM, Galvagni M, Eccher C, Caramatti S, Zanolli D, et al. Integrating mHealth in oncology: experience in the province of Trento. J Med Internet Res 2015 May 13;17(5):e114 [FREE Full text] [doi: 10.2196/jmir.3743] [Medline: 25972226]
28. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020 Jan 13;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
29. Qudah I, Leijdekkers P, Gay V. Using mobile phones to improve medication compliance and awareness for cardiac patients. In: Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments. 2010 Jun Presented at: PETRA '10; June 23-25, 2010; Samos, Greece p. 1-7. [doi: 10.1145/1839294.1839337]
30. Wu D, Lowry PB, Zhang D, Parks RF. Patients' compliance behavior in a personalized mobile patient education system (PMPES) setting: rational, social, or personal choices? Int J Med Inform 2021 Jan;145:104295. [doi: 10.1016/j.ijmedinf.2020.104295] [Medline: 33129124]
31. Newell D, MChiro RB. Increasing compliance toward home exercise in chiropractic patients using SMS texting: a pilot study. Clin Chiropr 2012 Dec;15(3-4):107-111. [doi: 10.1016/j.clch.2012.10.042]
32. Chen HC, Chuang TY, Lin PC, Lin YK, Chuang YH. Effects of messages delivered by mobile phone on increasing compliance with shoulder exercises among patients with a frozen shoulder. J Nurs Scholarsh 2017 Jul;49(4):429-437. [doi: 10.1111/jnu.12308] [Medline: 28632975]
33. Ifejika NL, Noser EA, Grotta JC, Savitz SI. Swipe out Stroke: feasibility and efficacy of using a smart-phone based mobile application to improve compliance with weight loss in obese minority stroke patients and their carers. Int J Stroke 2016 Jul;11(5):593-603. [doi: 10.1177/1747493016631557] [Medline: 26956032]
34. Orlemann T, Reljic D, Zenker B, Meyer J, Eskofier B, Thiemt J, et al. A novel mobile phone app (OncoFood) to record and optimize the dietary behavior of oncologic patients: pilot study. JMIR Cancer 2018 Nov 20;4(2):e10703 [FREE Full text] [doi: 10.2196/10703] [Medline: 30459139]
35. Piette JD, Heisler M, Krein S, Kerr EA. The role of patient-physician trust in moderating medication nonadherence due to cost pressures. Arch Intern Med 2005 Aug;165(15):1749-1755. [doi: 10.1001/archinte.165.15.1749] [Medline: 16087823]
36. Lu X, Zhang R, Wu W, Shang X, Liu M. Relationship between internet health information and patient compliance based on trust: empirical study. J Med Internet Res 2018 Aug 17;20(8):e253 [FREE Full text] [doi: 10.2196/jmir.9364] [Medline: 30120087]
37. Chandra S, Mohammadnezhad M, Ward P. Trust and communication in a doctor- patient relationship: a literature review. J Healthc Commun 2018 Jan;3(3):36. [doi: 10.4172/2472-1654.100146]
38. Kashgary A, Alsolaimani R, Mosli M, Faraj S. The role of mobile devices in doctor-patient communication: a systematic review and meta-analysis. J Telemed Telecare 2017 Sep;23(8):693-700. [doi: 10.1177/1357633X16661604] [Medline: 27632990]

39. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
40. Rowe R, Calnan M. Trust relations in health care--the new agenda. Eur J Public Health 2006 Mar;16(1):4-6. [doi: 10.1093/eurpub/ckl004] [Medline: 16446297]
41. Ozawa S, Sripad P. How do you measure trust in the health system? A systematic review of the literature. Soc Sci Med 2013 Aug;91:10-14. [doi: 10.1016/j.socscimed.2013.05.005] [Medline: 23849233]
42. Caterinicchio RP. Testing plausible path models of interpersonal trust in patient-physician treatment relationships. Soc Sci Med Med Psychol Med Sociol 1979 Jan;13A(1):81-99. [doi: 10.1016/0160-7979(79)90011-0] [Medline: 551529]
43. Thom DH, Hall MA, Pawlson LG. Measuring patients' trust in physicians when assessing quality of care. Health Aff (Millwood) 2004;23(4):124-132. [doi: 10.1377/hlthaff.23.4.124] [Medline: 15318572]
44. McAllister DJ. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Acad Manag J 1995 Feb 01;38(1):24-59. [doi: 10.2307/256727]
45. Benbasat I, Wang W. Trust in and adoption of online recommendation agents. J Assoc Inf Syst 2005 Mar;6(3):72-101. [doi: 10.17705/1jais.00065]
46. Paul DL, McDaniel Jr RR. A field study of the effect of interpersonal trust on virtual collaborative relationship performance. MIS Q 2004 Jun;28(2):183-227. [doi: 10.2307/25148633]
47. Anderson LA, Dedrick RF. Development of the Trust in Physician scale: a measure to assess interpersonal trust in patient-physician relationships. Psychol Rep 1990 Dec;67(3 Pt 2):1091-1100. [doi: 10.2466/pr0.1990.67.3f.1091] [Medline: 2084735]
48. Ware Jr JE, Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. Eval Program Plann 1983;6(3-4):247-263. [doi: 10.1016/0149-7189(83)90005-8] [Medline: 10267253]
49. Shan L, Li Y, Ding D, Wu Q, Liu C, Jiao M, et al. Patient satisfaction with hospital inpatient care: effects of trust, medical insurance and perceived quality of care. PLoS One 2016 Oct 18;11(10):e0164366 [FREE Full text] [doi: 10.1371/journal.pone.0164366] [Medline: 27755558]
50. Mitchell AJ, Selmes T. Why don't patients take their medicine? Reasons and solutions in psychiatry. Adv Psychiatr Treat 2018 Jan 02;13(5):336-346. [doi: 10.1192/apt.bp.106.003194]
51. LoCurto J, Berg GM. Trust in healthcare settings: scale development, methods, and preliminary determinants. SAGE Open Med 2016 Sep 1;4:2050312116664224 [FREE Full text] [doi: 10.1177/2050312116664224] [Medline: 27635245]
52. Lawson VL, Lyne PA, Harvey JN, Bundy CE. Understanding why people with type 1 diabetes do not attend for specialist advice: a qualitative analysis of the views of people with insulin-dependent diabetes who do not attend diabetes clinic. J Health Psychol 2005 May;10(3):409-423. [doi: 10.1177/1359105305051426] [Medline: 15857871]
53. McQuaid EL. Barriers to medication adherence in asthma: the importance of culture and context. Ann Allergy Asthma Immunol 2018 Jul;121(1):37-42. [doi: 10.1016/j.anai.2018.03.024] [Medline: 29580846]
54. Xu Y, He W. More information = less aggression? impact of information asymmetry on Chinese patients' aggression. Front Public Health 2019 May 22;7:118 [FREE Full text] [doi: 10.3389/fpubh.2019.00118] [Medline: 31192181]
55. Stiglitz JE. The contributions of the economics of information to twentieth century economics. Q J Econ 2000 Nov 01;115(4):1441-1478. [doi: 10.1162/003355300555015]
56. Wang N, Liang H, Xue Y, Ge S. Mitigating information asymmetry to achieve crowdfunding success: signaling and online communication. J Assoc Inf Syst 2021;22(3):773-796. [doi: 10.17705/1jais.00679]
57. Gao G, Greenwood BN, Agarwal R, McCullough JS. Vocal minority and silent majority: how do online ratings reflect population perceptions of quality? MIS Q 2015;39(3):565-589.
58. Kuhlmann E. Traces of doubt and sources of trust: health professions in an uncertain society. Curr Sociol 2006 Jul;54(4):607-620. [doi: 10.1177/0011392106065092]
59. Gulbrandsen P, Clayman ML, Beach MC, Han PK, Boss EF, Ofstad EH, et al. Shared decision-making as an existential journey: aiming for restored autonomous capacity. Patient Educ Couns 2016 Sep;99(9):1505-1510. [doi: 10.1016/j.pec.2016.07.014] [Medline: 27460801]
60. Wu Q, Jin Z, Wang P. The relationship between the physician-patient relationship, physician empathy, and patient trust. J Gen Intern Med 2022 May;37(6):1388-1393. [doi: 10.1007/s11606-021-07008-9] [Medline: 34405348]
61. Gonzalez J, Williams Jr JW, Noël PH, Lee S. Adherence to mental health treatment in a primary care clinic. J Am Board Fam Pract 2005;18(2):87-96 [FREE Full text] [doi: 10.3122/jabfm.18.2.87] [Medline: 15798137]
62. Bosnjak M, Ajzen I, Schmidt P. The theory of planned behavior: selected recent advances and applications. Eur J Psychol 2020 Aug;16(3):352-356 [FREE Full text] [doi: 10.5964/ejop.v16i3.3107] [Medline: 33680187]
63. Jenkins J, Durcikova A, Nunamaker Jr J. Mitigating the security intention-behavior gap: the moderating role of required effort on the intention-behavior relationship. J Assoc Inf Syst 2021 Jan 8;22(1):246-272. [doi: 10.17705/1jais.00660]
64. Dinev T, Hu Q. The centrality of awareness in the formation of user behavioral intention toward protective information technologies. J Assoc Inf Syst 2008 Oct 1;8(7):386-408. [doi: 10.17705/1jais.00133]
65. Bulgurcu B, Cavusoglu H, Benbasat I. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS Q 2010 Sep;34(3):523-548. [doi: 10.2307/25750690]

66. Herath T, Rao HR. Protection motivation and deterrence: a framework for security policy compliance in organisations. Eur J Inf Syst 2009;18(2):106-125. [doi: 10.1057/ejis.2009.6]
67. Johnston AC, Warkentin M. Fear appeals and information security behaviors: an empirical study. MIS Q 2010;34(3):549-566. [doi: 10.2307/25750691]
68. Lu X, Zhang R. Impact of physician-patient communication in online health communities on patient compliance: cross-sectional questionnaire study. J Med Internet Res 2019 May 13;21(5):e12891 [FREE Full text] [doi: 10.2196/12891] [Medline: 31094342]
69. Leone L, Perugini M, Ercolani AP. A comparison of three models of attitude-behavior relationships in the studying behavior domain. Eur J Soc Psychol 1999 Mar;29(2-3):161-189. [doi: 10.1002/(sici)1099-0992(199903/05)29:2/3<161::aid-ejsp919>3.0.co;2-g]
70. Siponen MT. A conceptual foundation for organizational information security awareness. Inf Manag Comput Secur 2000 Mar;8(1):31-41. [doi: 10.1108/09685220010371394]
71. Mishra S, Dhillon G. Information systems security governance research: a behavioral perspective. In: 1st Annual Symposium on Information Assurance, Academic Track of 9th Annual NYS Cyber Security Conference. 2006 Presented at: ASIA '06; June 14-15, 2006; New York, NY, USA.
72. Reicher S, Levine M. On the consequences of deindividuation manipulations for the strategic communication of self: identifiability and the presentation of social identity. Eur J Soc Psychol 1994 Jul;24(4):511-524. [doi: 10.1002/ejsp.2420240408]
73. Postmes T, Spears R. Deindividuation and antinormative behavior: a meta-analysis. Psychol Bull 1998 May;123(3):238-259. [doi: 10.1037/0033-2909.123.3.238]
74. Wiatrowski MD, Griswold DB, Roberts MK. Social control theory and delinquency. Am Sociol Rev 1981 Oct;46(5):525-541. [doi: 10.2307/2094936]
75. Steine S, Finset A, Laerum E. A new, brief questionnaire (PEQ) developed in primary health care for measuring patients' experience of interaction, emotion and consultation outcome. Fam Pract 2001 Aug;18(4):410-418. [doi: 10.1093/fampra/18.4.410] [Medline: 11477049]
76. Hall MA, Zheng B, Dugan E, Camacho F, Kidd KE, Mishra A, et al. Measuring patients' trust in their primary care providers. Med Care Res Rev 2002 Sep;59(3):293-318. [doi: 10.1177/1077558702059003004] [Medline: 12205830]
77. Horton RP, Buck T, Waterson PE, Clegg CW. Explaining intranet use with the technology acceptance model. J Inf Technol 2001 Dec 01;16(4):237-249. [doi: 10.1080/02683960110102407]
78. Workman M, Bommer WH, Straub D. Security lapses and the omission of information security measures: a threat control model and empirical test. Comput Human Behav 2008 Sep;24(6):2799-2816. [doi: 10.1016/j.chb.2008.04.005]
79. Hu Q, Xu Z, Dinev T, Ling H. Does deterrence work in reducing information security policy abuse by employees? Commun ACM 2011 Jun;54(6):54-60. [doi: 10.1145/1953122.1953142]
80. 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]
81. Gefen D, Rigdon EE, Straub D. Editor's comments: an update and extension to SEM guidelines for administrative and social science research. MIS Q 2011 Jun;35(2):iii-xiv. [doi: 10.2307/23044042]
82. Rai A, Patnayakuni R, Seth N. Firm performance impacts of digitally enabled supply chain integration capabilities. MIS Q 2006 Jun;30(2):225-246. [doi: 10.2307/25148729]
83. Lowry PB, Gaskin J. Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: when to choose it and how to use it. IEEE Trans Profess Commun 2014 Jun;57(2):123-146. [doi: 10.1109/tpc.2014.2312452]
84. Ringle C, da Silva D, Bido D. Structural equation modeling with the SmartPLS. Brazil J Mark 2015 Oct 22;13(2):56-73. [doi: 10.5585/remark.v13i2.2717]
85. 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]
86. Malhotra NK, Kim SS, Patil A. Common method variance in IS research: a comparison of alternative approaches and a reanalysis of past research. Manag Sci 2006 Dec;52(12):1865-1883. [doi: 10.1287/mnsc.1060.0597]
87. Colquitt JA, Rodell JB. Justice, trust, and trustworthiness: a longitudinal analysis integrating three theoretical perspectives. Acad Manag J 2011 Dec;54(6):1183-1206. [doi: 10.5465/amj.2007.0572]
88. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Market Res 1981 Feb;18(1):39-50. [doi: 10.1177/002224378101800104]
89. Barclay DW, Higgins CA, Thompson R. The partial least squares approach to causal modeling: personal computer adoption and use as an illustration. Technol Stud 1995;2(2):285-324.
90. Wixom BH, Watson HJ. An empirical investigation of the factors affecting data warehousing success. MIS Q 2001 Mar;25(1):17-41. [doi: 10.2307/3250957]
91. Gefen D, Straub D. A practical guide to factorial validity using PLS-Graph: tutorial and annotated example. Commun Assoc Inf Syst 2005;16(5):91-109. [doi: 10.17705/1cais.01605]

92. Montague E, Asan O. Trust in technology-mediated collaborative health encounters: constructing trust in passive user interactions with technologies. Ergonomics 2012;55(7):752-761 [FREE Full text] [doi: 10.1080/00140139.2012.663002] [Medline: 22506847]
93. Qiao Y, Asan O, Montague E. Factors associated with patient trust in electronic health records used in primary care settings. Health Policy Technol 2015 Dec;4(4):357-363. [doi: 10.1016/j.hlpt.2015.08.001]
94. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
95. Huang M, Fan J, Prigge J, Shah ND, Costello BA, Yao L. Characterizing patient-clinician communication in secure medical messages: retrospective study. J Med Internet Res 2022 Jan 11;24(1):e17273 [FREE Full text] [doi: 10.2196/17273] [Medline: 35014964]
96. Ryan R, Santesso N, Hill S, Lowe D, Kaufman C, Grimshaw J. Consumer-oriented interventions for evidence-based prescribing and medicines use: an overview of systematic reviews. Cochrane Database Syst Rev 2011 May 11(5):CD007768. [doi: 10.1002/14651858.CD007768.pub2] [Medline: 21563160]
97. Fogg BJ. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Apr Presented at: Persuasive '09; April 26-29, 2009; Claremont, CA, USA p. 1-7. [doi: 10.1145/1541948.1541999]
98. Lau AY, Proudfoot J, Andrews A, Liaw ST, Crimmins J, Arguel A, et al. Which bundles of features in a Web-based personally controlled health management system are associated with consumer help-seeking behaviors for physical and emotional well-being? J Med Internet Res 2013 May 06;15(5):e79 [FREE Full text] [doi: 10.2196/jmir.2414] [Medline: 23649790]
99. Zhu X, Tao Y, Zhu R, Wu D, Ming WK. Impact of hospital characteristics and governance structure on the adoption of tracking technologies for clinical and supply chain use: longitudinal study of US hospitals. J Med Internet Res 2022 May 26;24(5):e33742 [FREE Full text] [doi: 10.2196/33742] [Medline: 35617002]
100. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021 May 25;23(5):e24190. [doi: 10.2196/24190] [Medline: 34032583]

Abbreviations

AVE: average variance extracted
mHealth: mobile health
MPES: mobile patient education system
TPB: theory of planned behavior

Edited by T Leung, G Eysenbach; submitted 24.09.22; peer-reviewed by X Lu, Y Jeem; comments to author 25.10.22; revised version received 13.11.22; accepted 28.11.22; published 20.12.22.

Please cite as:
Wu D, Lowry PB, Zhang D, Tao Y
Patient Trust in Physicians Matters—Understanding the Role of a Mobile Patient Education System and Patient-Physician Communication in Improving Patient Adherence Behavior: Field Study
J Med Internet Res 2022;24(12):e42941
URL: https://www.jmir.org/2022/12/e42941
doi:10.2196/42941
PMID:36538351

Original Paper

The Role of Access Type and Age Group in the Breadth of Use of Patient Portals: Observational Study

Theophile Ndabu[1]; Lavlin Agrawal[2], MS; Raj Sharman[2], PhD

[1]Belk College of Business, University of North Carolina at Charlotte, Charlotte, NC, United States

[2]State University of New York, University at Buffalo, Buffalo, NY, United States

Corresponding Author:
Theophile Ndabu
Belk College of Business
University of North Carolina at Charlotte
9209 Mary Alexander Rd
Charlotte, NC, 28262
United States
Phone: 1 5852855316
Email: tnndabu@buffalo.edu

Abstract

Background: Health care delivery and patient satisfaction are improved when patients engage with their medical information through patient portals. Despite their wide availability and multiple functionalities, patient portals and their functionalities are still underused.

Objective: We seek to understand factors that lead to patient engagement through multiple portal functionalities. We provide recommendations that could lead to higher patients' usage of their portals.

Methods: Using data from the Health Information National Trends Survey 5, Cycle 3 (N=2093), we performed descriptive statistics and used a chi-square test to analyze the association between the demographic variables and the use of mobile health apps for accessing medical records. We further fitted a generalized linear model to examine the association between access type and the use of portal functionalities. We further examined the moderation effects of age groups on the impact of access type on portal usage.

Results: Our results show that accessing personal health records using a mobile health app is positively associated with greater patient usage of access capabilities (β=.52; *P*<.001), patient-provider interaction capabilities (β=.24, *P*=.006), and patient–personal health information interaction capabilities (β=.23, *P*=.009). Patients are more likely to interact with their records and their providers when accessing their electronic medical records using a mobile health app. The impacts of mobile health app usage fade with age for tasks consisting of viewing, downloading, and transmitting medical results to a third party (β=–.43, *P*=.005), but not for those involving patient-provider interaction (β=.05, *P*=.76) or patient–personal health information interaction (β=–.15, *P*=.19).

Conclusions: These findings provide insights on how to increase engagement with diverse portal functionalities for different age groups and thus improve health care delivery and patient satisfaction.

(J Med Internet Res 2022;24(12):e41972) doi:10.2196/41972

KEYWORDS

patient portal; mobile health apps; electronic medical records; personal health record functionalities; patient satisfaction; health information; Health Information National Trends Survey; healthcare delivery; app use; electronic medical record

Introduction

Background

Increased collaboration between patients and providers in the delivery of health care results in patient engagement and ultimately patient satisfaction [1-4]. Thus, hospitals and medical practices proactively involve patients in decision-making and all other aspects of care by offering them access to their personal health information (PHI) stored in electronic health records—a health care provider–facing digital copy of patients' medical records. Patients can access, share, and interact with both their medical records and their providers through PHI management tools, often called patient portals (hereinafter referred to as "portals") [5-10]. In the United States, portals are used to meet

calls from the Office of the National Coordinator (ONC) to provide patients with the capabilities to view, download, and transmit (VDT) their records to a third party of their choice [11,12]. Patients' usage of VDT and other functionalities (exchanging messages with providers, scheduling medical appointments, etc) helps health care practitioners receive financial incentives. However, health care practitioners must meet qualifying usage thresholds set forth by the Promoting Interoperability Programs of the US Centers for Medicare and Medicaid Services [3,7,11]. Engagement with functionalities of portals improves the quality of care delivery and patient satisfaction [13-25]. Thus, understanding factors that lead to patient engagement through the use of multiple portal functionalities is warranted.

Portal usage remains low despite its benefits [7,9]. Several studies have also shown that portal usage varies on the basis of patients' chronic disease and demographic statuses, such that patients from ethnic minorities are less likely to access and interact with their medical records electronically [6,26-31]. Previous research also shows that health information technology characteristics, such as usability and convenience, influence the use of portals [32-36]. A recent survey on consumers' use of mobile apps and mobile browsers showed that people spend nearly 3 hours daily on mobile apps but less than an hour using mobile browsers [37]. This trend aligns with prior research that patients, including older patients, are interested in using mobile apps [31-33]. The extent to which the convenience of mobile health apps translates into the diversity of use of portal functionalities has not been studied. Building on these studies and using a nationally representative sample, we examine the role of portal access type and mobile health app use on the likelihood that patients of different age groups will use multiple portal engagement functionalities. These functionalities include VDT and tasks allowing users to interact with their PHI and providers. We further examine which functionalities benefit from each portal access type.

To the best of our knowledge, prior studies have not described or examined factors that lead to the use of diverse capabilities of portals in a nonintegrated context among patients using a mobile health app. Understanding these factors could help designers and health care practitioners facilitate both the frequency and the breadth of the use of portal functionalities by patients and also develop proper access-based intervention for underused capabilities.

Engagement Functionalities and Patient Engagement

The ONC classifies portal engagement functionalities into 2 types based on their utility. The first type consists of capabilities that allow patients to access their medical records to perform VDT tasks. The second type of engagement functionality includes capabilities that facilitate web-based interactions between both patients and their providers and patients and their PHI. For example, interaction functionalities include capabilities such as secure messaging between the patient and provider, refilling prescriptions, and amending personal records [11]. Since the use of functionalities depends in part on their utility, the study adopts the ONC classification as well as the term engagement functionalities to describe portal capabilities that allow patients to (1) access their personal health information or (2) interact with their providers and (3) interact with their data. Similarly, we refer to patients' use of engagement functionalities as patient engagement with their care in line with extant research [3,7,11,36].

Our study examines the role of portal access type in promoting patient engagement through the use of portal engagement functionalities. We specifically hypothesize that patients accessing their electronic medical records using a mobile health app are more likely to participate in all 3 aspects of engagement: VDT, patient-provider interaction (PPI), and patient-PHI interaction (PPHI). We also propose that the effects of access type on portal usage vary by functionality and age group. Specifically, the intensity and the breadth of the use of engagement functionalities differ on the basis of whether portal users are younger than 65 years. This study controls for demographic- and health-related behavioral variables.

Methods

Data Source and Study Population

We used data from the Health Information National Trends Survey (HINTS) 5, Cycle 3, collected by the National Cancer Institute between January and May 2019. HINTS 5 Cycle 3 surveys noninstitutionalized civilian US adults using a 2-stage sampling design. Data were collected from 5438 respondents out of 23,430 targeted addresses (overall response rate 30.3%) [38]. The survey methods and detailed reports have been published elsewhere [38-40]. HINTS data sets have been used in the literature to study health-related behaviors, including information seeking and sharing, patient-provider trust, and HIT adoption and use in health care [34-36,41-43]. This study focuses on patients who indicated on the survey questionnaire whether they use a smartphone health app to access their electronic medical records. We filtered survey responses to exclude those who did not know whether they had used an app and missing values for this parameter. The resulting sample consisted of 2093 observations.

Variables

Predictor and Control Variables

The main predictor in this study was whether users accessed their web-based medical records using a smartphone health app. Participants who responded as having used smartphones to access their information were recoded as 1 and those who had not as 0. The study also controlled for demographic and health-related variables, including self-reported gender, income, age (younger than 65 years having been recoded as 0 and 65 years or older as 1), education, race and ethnicity, general health status, and insurance status. In addition, in line with previous HIT studies, we controlled for the chronic disease status [35,36] and propensity to search for health information on the internet [6,36,41]. The *propensity for searching health information on the internet* parameter was derived from questionnaire responses on whether respondents had used a computer, smartphone, or other electronic means to search for medical information on the internet. Survey questions and variable measurements are reported in Multimedia Appendix 1.

Outcome Variables

The role of app use on portal usage was examined using 3 engagement functionalities: VDT, PPI, and PPHI.

VDT scores were obtained by summing up answers to 3 dichotomous survey questions regarding the use of electronic medical records. First, respondents were asked whether they used their portals to view laboratory results and download medical records. "Yes" answers were recoded as 1 and "No" answers as 0. Second, respondents were also asked whether they had used their portals to share their electronic records with another health care provider, a family member, or a health app to manage or store the data. Those who had shared their records with any of the options were recoded as 1 and those who had not as 0. Finally, the recoded responses to the 3 questions were summed to obtain a VDT score ranging from 0 to 3.

We conceptually grouped responses to questions regarding the use of interaction functionalities into 2 categories. The first category consisted of functionalities related to care and communication convenience. For example, respondents were asked whether they requested medication refills or sent a secure message to their providers. We called this parameter PPI. The variable ranged from 0 to 2 after summing the dichotomous values of the questionnaire responses. The second parameter, PPHI, consisted of answers to questions on whether users (1) requested the correction of erroneous information in their records, (2) added information to their records to share with their health care provider, or (3) used their health records in deciding on a treatment for an illness or health condition. We used the values of the 3 items to compute the PPHI factor score, ranging from 0 to 3. The acronym VDT has been used extensively in the literature to represent the outcomes of VDT functionalities and, thus, yield little or no confusion [3,7,11]. However, this is not the case with interaction functionalities. We, therefore, used themes underlying these capabilities to represent interaction capabilities in line with extant literature [36].

Statistical Analyses

We used descriptive statistics and performed a chi-square test to analyze the association between the demographic variables and app use for accessing medical records. Using a generalized linear model, we examined the association between app use and portal engagement functionalities and the moderating effect on respondents' age group. Our analyses incorporated replicate weights to account for the survey design methodology. R (The R Foundation) statistical software was used to conduct the analyses. A P value of ≤.05 was considered significant.

Ethical Considerations

This study uses publicly available (secondary) data; hence, institutional review board review was not required.

Results

Outcome Variables

PPI and PPHI were further validated with exploratory factor analysis (EFA) in line with previous research [36]. The suitability of the data for an EFA was assessed with the nonsignificant Little missing value test ($^2_{68}$=61, P=.71), indicating that the data were missing completely at random (significant Bartlett test for sphericity: $^2_{21}$= 1617.25, P<.001; Kaiser-Meyer-Olkin measure of sample adequacy: overall index=0.74). As shown in Table 1, a survey-weighted EFA with maximum likelihood suggested 2 main themes from the questionnaire responses. A scree plot and eigenvalues (>1) supported selecting 2 themes for the analysis.

Table 1. Exploratory factor analysis to extract common themes from the data.

Variables	Factor loadings	
	Factor 1	Factor 2
Refill prescriptions	N/A[a]	0.45
Send secure message	N/A	0.64
Request correction in personal health information	0.37	N/A
Add info to medical records	0.81	N/A
Make care decisions	0.44	N/A

[a]N/A: not applicable.

Descriptive Analysis

Our sample consisted of 2093 observations. As shown in Table 2, of all the respondents, only 714 (37.2%) patients reported using an app to access their portals. Females accounted for the majority of respondents (n=1186, 57.3%). The association between app use and age group variables was significant (P<.001), indicating that the impact of this association was not uniform across age group levels. As shown in Table 2, respondents younger than 65 years were more likely to report using an app (n=553, 88.5%) than those aged 65 years or older (n=154, 11.5%). The data also show that most respondents (n=1824, 84.1%) had more than a high school degree. Most app users earned US $50,000 or more (n=467, 64.4%) and were essentially non-Hispanic White (n=399, 62.5%). Most respondents self-reported their general health status as *good* or better (n=620, 89.4%).

Table 2. Demographic and health-related categorical variables.

Variables	Overall, n (%)	Access without app, n (%)	Access with app, n (%)	*P* value
Participants	2093 (100)	1379 (62.8)	714 (37.2)	N/A[a]
Gender				.55
Female	1186 (57.3)	769 (56.3)	417 (58.9)	
Male	771 (42.7)	512 (43.7)	259 (41.1)	
Age group (years)				<.001
<65	1411 (82.8)	858 (79.4)	553 (88.5)	
≥65	651 (17.2)	497 (20.6)	154 (11.5)	
Education				.78
Less than high school	36 (1.7)	25 (1.7)	11 (1.6)	
High school graduate	195 (14.2)	132 (15.2)	63 (12.5)	
Some college	584 (43.2)	377 (42.4)	207 (44.7)	
College graduate or more	1240 (40.9)	823 (40.7)	417 (41.1)	
Income (US $)				.26
<20,000	173 (10.5)	112 (9.6)	61 (12.0)	
20,000 to <35,000	169 (8.2)	105 (6.2)	64 (11.6)	
35,000 to <50,000	233 (13.5)	174 (14.3)	59 (12.1)	
50,000 to <75,000	382 (19.3)	255 (20.5)	127 (17.3)	
≥75,000	945 (48.5)	605 (49.4)	340 (47.1)	
Health status				.86
Poor	36 (2.4)	21 (2.3)	15 (2.6)	
Fair	214 (8.9)	143 (9.4)	71 (8.1)	
Good	706 (34.3)	462 (35.2)	244 (32.9)	
Very good	825 (41.0)	551 (40.0)	274 (42.8)	
Excellent	284 (13.3)	182 (13.1)	102 (13.7)	
Race and ethnicity				.14
Non-Hispanic White	1339 (69.4)	940 (73.4)	399 (62.5)	
Non-Hispanic Black or African American	219 (8.6)	132 (8.2)	87 (9.1)	
Hispanic	233 (12.9)	113 (9.6)	120 (18.4)	
Non-Hispanic Asian	96 (5.4)	56 (5.3)	40 (5.5)	
Non-Hispanic other	73 (3.8)	46 (3.4)	27 (4.6)	
Chronic disease status (number of diseases)				.29
0	804 (40.0)	528 (41.3)	276 (37.8)	
1	669 (33.0)	442 (32.8)	227 (33.4)	
2	419 (18.6)	277 (17.6)	142 (20.2)	
3	149 (5.5)	98 (5.8)	51 (4.8)	
4	37 (1.7)	23 (0.9)	14 (3.0)	
5	10 (1.3)	7 (1.6)	3 (0.8)	
Search health information on the internet				.002
No	227 (12.3)	177 (15.1)	50 (7.4)	
Yes	1845 (87.7)	1186 (84.9)	659 (92.6)	
Insurance status				.66
Uninsured	63 (4.9)	41 (4.4)	22 (5.6)	

Variables	Overall, n (%)	Access without app, n (%)	Access with app, n (%)	*P* value
Insured	2030 (95.1)	1338 (95.6)	692 (94.4)	

[a]N/A: not applicable.

Regarding the 3 engagement activities, disparities were also apparent between respondents who used the mobile app and those who did not. The mean VDT score for those who reported using an app was 1.56 (SD 0.83) compared to 1.11 (SD 0.71) for those who did not (*P*<.001). For PPI score, the average was 1.15 (SD 0.79) for those who used an app versus 0.87 (SD 0.81) for those who did not (*P*<.001), while, for the PPHI score, the average was 0.77 (SD 0.93) for respondents who used an app versus 0.48 (SD 0.77) for those who did not (*P*<.001). The significant *P* value for all engagement factors indicates that the use of engagement functionalities differed between users who reported accessing their portals through a health app and those who did not.

Figure 1 shows that the usage of individual functionalities varied with access type. Among respondents who answered "Yes" when asked whether they used electronic medical records to make decisions about health care options, the number of participants who accessed their portal using a mobile app was higher (n=376, 51.7%) than that of non–app users. Similarly, the numbers of app users who downloaded medical results or shared records with a third party were also higher (n=196, 63.2% and n=288, 54.9%, respectively) than those of non–app users. Furthermore, web access was higher for other functionalities, including adding additional information to one's medical records (n=278, 53.7%), correcting medical records (n=94, 61.1%), sending secure messages to providers (n=645, 56.7%), requesting prescription refills (n=559, 55.8%), and viewing medical results (n=1108, 60.2%).

Figure 2 depicts the distribution of how the respondents used multiple functionalities. The usage of portal functionalities in general and the interaction functionalities, in particular, remained low. Nearly 54% (n=374) of app users and 70% (n=925) of nonusers did not engage in any PPHI task. Similarly, over 25% (n=177) of app users and 40% (n=549) of nonusers did not engage in any PHI task. The fraction of respondents who did not use any VDT tasks remained under 20% (n=263) among both app users and nonusers. Overall, app users were more likely to engage in at least one task within each outcome score.

Figure 1. Comparison of usage by access type among users who claimed to have used personal health record functionalities. PPHI: patient–personal health information interaction; PPI: patient-provider interaction; VDT: view, download, and transmit.

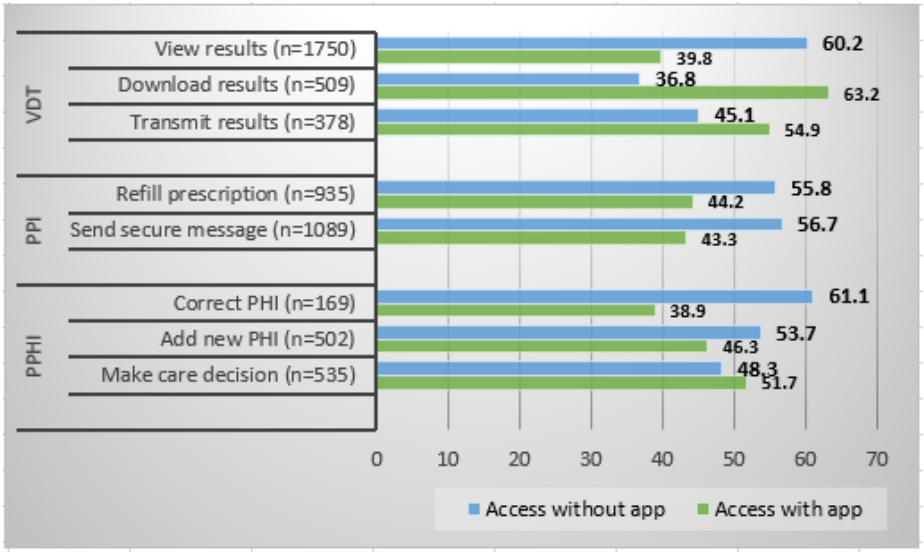
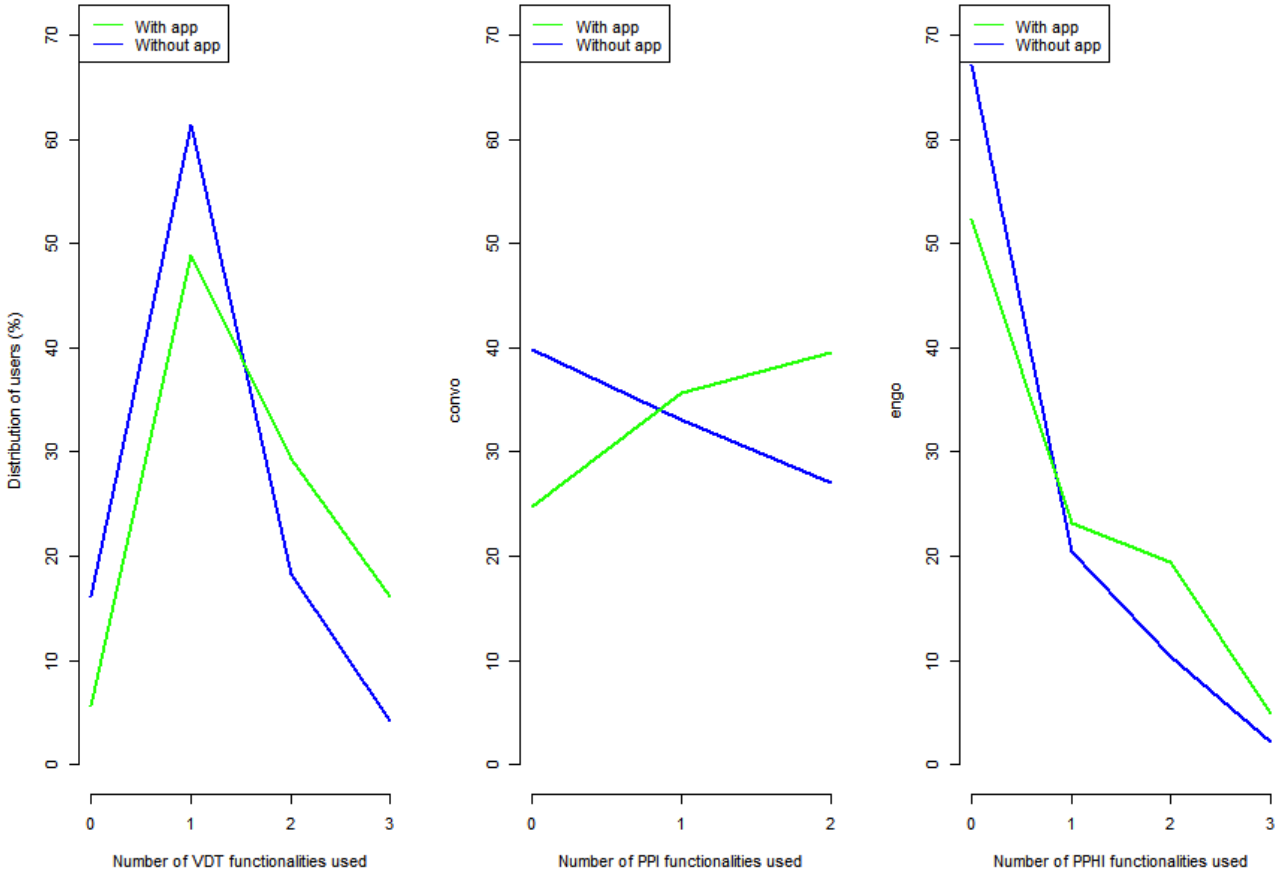


Figure 2. Comparison of the number of functionalities used by access type. PPHI: patient–personal health information interaction; PPI: patient-provider interaction; VDT: view, download, and transmit.



Generalized Linear Model

Results from the survey-weighted linear model are presented in Table 3. Models 1, 3, and 5 examine the effect of app usage without an interaction term. The results show that accessing portals using a health app is positively associated with higher usage of VDT, PPI, and PPHI functionalities. Specifically, app usage is associated with an increase of 0.52 (P<.001) in the VDT score, 0.24 (P=.006) in the PPI score, and 0.21 (P=.009) in the PPHI interaction score. These results indicate that for every 100 patients who use a health app to access their medical records, 52 more VDT tasks, 24 more PPI tasks, and 21 more PPHI tasks are performed. Models 2, 4, and 6 account for the interaction effects between age group and app use. The coefficients of the main independent variable (app use) changed slightly but not significantly (0.52 to 0.58 for the VDT score, 0.24 to 0.23 for the PPI score, and 0.21 to 0.23 for the PPHI score). The R^2 value increased for the VDT score from 0.19 to 0.20 but did not change for the PPI and PPHI scores, indicating that only the VDT model is better explained when the model includes an interaction term.

Respondents who reported general health statuses of *good* (β=–.53, P=.03) and *excellent* (β=–.58, P=.04) were less likely to engage in PPHI tasks than those who reported poor health. Chronic disease status was associated with the usage of PPI and PPHI functionalities. The propensity to search health information on the internet was also associated with higher usage of VDT and PPI tasks. Insurance status, income, education, race and ethnicity, and gender were not significant.

Table 3. Regression results.

	Engagement functionalities (participants, n=1753)					
	View, download, and transmit score (95% CI)		Patient-provider information score (95% CI)		Patient–personal health information interaction score (95% CI)	
	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[c]	Model 5[d]	Model 6[d]
Access type (reference: access without an app)	0.52[e] (0.35 to 0.69)	0.58[e] (0.39 to 0.76)	0.24[f] (0.08 to 0.40)	0.23[g] (0.04 to 0.42)	0.21[f] (0.06 to 0.37)	0.23[f] (0.07 to 0.40)
Age group of ≥65 years (reference: <65 years old)	–0.01 (–0.12 to 0.11)	0.11 (–0.02 to 0.24)	–0.01 (–0.14 to 0.13)	–0.02 (–0.16 to 0.12)	0.07 (–0.08 to 0.21)	0.11 (–0.07 to 0.28)
Male gender (reference: female gender)	0.11 (–0.02 to 0.24)	0.11 (–0.01 to 0.24)	0.07 (–0.09 to 0.24)	0.07 (–0.09 to 0.24)	–0.08 (–0.21 to 0.04)	–0.08 (–0.21 to 0.04)
Race and ethnicity (reference: White)						
Hispanic	–0.07 (–0.33 to 0.19)	–0.07 (–0.33 to 0.18)	–0.02 (–0.22 to 0.17)	–0.02 (–0.22 to 0.17)	0.02 (–0.16 to 0.20)	0.02 (–0.16 to 0.20)
Non-Hispanic Asian	0.37 (–0.07 to 0.81)	0.39 (–0.05 to 0.82)	0.12 (–0.33 to 0.56)	0.12 (–0.33 to 0.56)	0.61 (–0.09 to 1.30)	0.61 (–0.08 to 1.31)
Non-Hispanic other	0.54 (–0.09 to 1.17)	0.54 (–0.09 to 1.17)	0.27 (–0.16 to 0.70)	0.27 (–0.16 to 0.71)	0.86 (–0.58 to 2.29)	0.86 (–0.59 to 2.30)
Non-Hispanic Black or African American	–0.15 (–0.35 to 0.06)	–0.14 (–0.34 to 0.06)	0.05 (–0.19 to 0.29)	0.05 (–0.19 to 0.29)	0.20 (–0.01 to 0.42)	0.21 (–0.01 to 0.43)
Education level (reference: less than high school)						
High school graduate	0.08 (–0.34 to 0.49)	0.07 (–0.34 to 0.47)	–0.35 (–1.00 to 0.29)	–0.35 (–1.00 to 0.30)	–0.22 (–0.68 to 0.24)	–0.22 (–0.68 to 0.23)
Some college	–0.10 (–0.49 to 0.30)	–0.10 (–0.48 to 0.27)	–0.27 (–0.84 to 0.31)	–0.27 (–0.84 to 0.31)	–0.18 (–0.68 to 0.32)	–0.18 (–0.68 to 0.31)
College graduate or more	–0.06 (–0.45 to 0.33)	–0.07 (–0.44 to 0.30)	–0.30 (–0.90 to 0.29)	–0.30 (–0.90 to 0.29)	–0.19 (–0.66 to 0.28)	–0.20 (–0.66 to 0.27)
Income (US $; reference: less than US $20,000)						
20,000 to <35,000	–0.06 (–0.40 to 0.27)	–0.06 (–0.39 to 0.27)	–0.05 (–0.46 to 0.35)	–0.06 (–0.46 to 0.35)	–0.04 (–0.35 to 0.27)	–0.04 (–0.35 to 0.27)
35,000 to <50,000	0.13 (–0.14 to 0.39)	0.12 (–0.15 to 0.38)	0.02 (–0.31 to 0.36)	0.02 (–0.31 to 0.36)	0.09 (–0.25 to 0.43)	0.08 (–0.26 to 0.43)
50,000 to <75,000	0.08 (–0.19 to 0.35)	0.09 (–0.18 to 0.35)	0.19 (–0.15 to 0.53)	0.19 (–0.15 to 0.53)	0.09 (–0.23 to 0.42)	0.09 (–0.23 to 0.42)
≥75,000	0.10 (–0.15 to 0.34)	0.10 (–0.14 to 0.34)	0.12 (–0.16 to 0.41)	0.12 (–0.16 to 0.41)	0.16 (–0.14 to 0.45)	0.16 (–0.14 to 0.46)
General health status (reference: poor)						
Fair	–0.10 (–0.59 to 0.38)	–0.11 (–0.59 to 0.36)	–0.40 (–0.97 to 0.16)	–0.40 (–0.97 to 0.16)	–0.37 (–0.96 to 0.22)	–0.38 (–0.97 to 0.22)
Good	–0.32 (–0.83 to 0.18)	–0.33 (–0.82 to 0.17)	–0.35 (–0.90 to 0.20)	–0.35 (–0.90 to 0.20)	–0.53[g] (–1.00 to –0.05)	–0.53[g] (–1.00 to –0.05)
Very good	–0.18 (–0.68 to 0.33)	–0.18 (–0.68 to 0.32)	–0.27 (–0.82 to 0.29)	–0.27 (–0.82 to 0.29)	–0.44 (–0.94 to 0.05)	–0.44 (–0.94 to 0.06)
Excellent	–0.29 (–0.79 to 0.22)	–0.29 (–0.79 to 0.20)	–0.43 (–0.98 to 0.12)	–0.43 (–0.98 to 0.12)	–0.57[g] (–1.11 to –0.03)	–0.58[g] (–1.12 to –0.03)
Chronic disease status (reference: 0 diseases)						
1	–0.01 (–0.16 to 0.14)	–0.01 (–0.17 to 0.14)	0.24[f] (0.07 to 0.41)	0.24[f] (0.07 to 0.41)	–0.04 (–0.24 to 0.15)	–0.04 (–0.24 to 0.15)

	Engagement functionalities (participants, n=1753)					
	View, download, and transmit score (95% CI)		Patient-provider information score (95% CI)		Patient–personal health information interaction score (95% CI)	
	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[c]	Model 5[d]	Model 6[d]
2	0.02 (–0.14 to 0.19)	0.03 (–0.13 to 0.19)	0.29[f] (0.09 to 0.48)	0.29[f] (0.09 to 0.48)	–0.03 (–0.22 to 0.16)	–0.03 (–0.21 to 0.16)
3	–0.05 (–0.22 to 0.11)	–0.05 (–0.20 to 0.11)	0.20 (–0.08 to 0.48)	0.20 (–0.08 to 0.48)	0.08 (–0.21 to 0.37)	0.08 (–0.21 to 0.38)
4	0.08 (–0.47 to 0.63)	0.06 (–0.48 to 0.60)	0.22 (–0.24 to 0.67)	0.22 (–0.24 to 0.68)	–0.08 (–0.52 to 0.36)	–0.09 (–0.53 to 0.35)
5	–0.13 (–0.73 to 0.47)	–0.13 (–0.71 to 0.45)	0.46 (–1.20 to 2.12)	0.46 (–1.20 to 2.11)	1.17[g] (0.27 to 2.08)	1.18[g] (0.26 to 2.09)
Insurance status: insured (reference: uninsured)	0.28 (–0.13 to 0.68)	0.27 (–0.13 to 0.67)	0.13 (–0.46 to 0.73)	0.13 (–0.46 to 0.73)	–0.16 (–0.51 to 0.19)	–0.16 (–0.50 to 0.19)
Searched health information on the web (ref: did not search)	0.33[g] (0.08 to 0.58)	0.33[g] (0.08 to 0.58)	0.31[g] (0.08 to 0.58)	0.31[g] (0.07 to 0.55)	0.10 (–0.35 to 0.55)	0.10 (–0.35 to 0.55)
App use × age group	Reference	–0.43[f] (–0.71 to –0.15)	Reference	0.05 (–0.29 to 0.39)	Reference	–0.15 (–0.38 to 0.08)
Constant	0.70 (–0.00 to 1.41)	0.69 (–0.01 to 1.39)	0.80 (–0.05 to 1.66)	0.81 (–0.05 to 1.66)	1.04[g] (0.18 to 1.89)	1.03[g] (0.18 to 1.89)

[a]R^2=0.19.
[b]R^2=0.20.
[c]R^2=0.09.
[d]R^2=0.19.
[e]P<.001.
[f]P<.01.
[g]P<.05.

Interaction Effects of Age Group

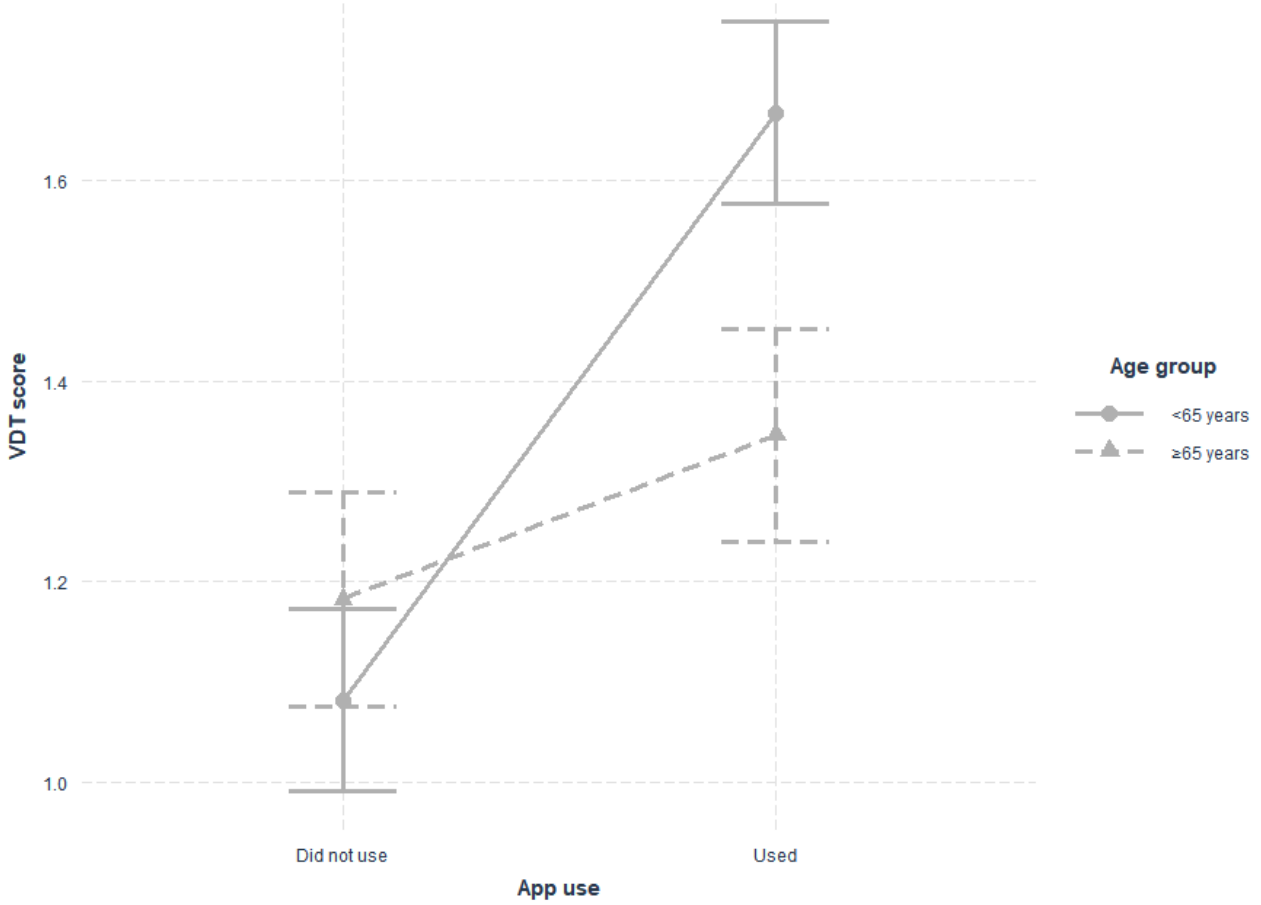
Models 2, 4, and 6 introduce an interaction term between app use and age group. The results indicate that the age group moderates the relationship (β=–.43, P=.004) between app use and VDT usage; that is, app use effects will differ between patients younger than 65 years and older patients. Table 4 and Figure 3 show the marginal means of VDT usage for all combinations of app use and age group. App users younger than 65 years exhibited higher averages than nonusers (1.67 vs 1.08, respectively). For users aged 65 years and older, Table 4 also shows a slight positive difference between app users (1.35) and nonusers (1.18). However, as shown in Figure 3, the mean difference is higher for users younger than 65 years, indicating an interaction between app use and age group variables.

The analysis of simple effects revealed a significant difference between those who used an app to perform VDT tasks and those who did not. The impact of app use was positive for both groups; however, the effect was significant only for participants younger than 65 years (β=.58, P<.001). Thus, app use was positively related to VDT use for users younger than 65 years but not for older users.

Table 4. Marginal means of VDT by app use and age group.

App usage	Age group (years)	Marginal mean (SE)
Did not use	<65	1.08 (0.0452)
Used	<65	1.67 (0.0639)
Did not use	≥65	1.18 (0.0528)
Used	≥65	1.35 (0.0952)

Figure 3. Interaction between functionalities and age. VDT: view, download, and transmit.



Discussion

Principal Findings

Health care delivery and outcomes can be improved with the increased participation of patients in health care decisions [44-47]. Portal functionalities, if sufficiently used, can be a driving force of patient-provider collaboration [4,14]. Thus, it is essential to integrate new and widely available technologies with portals to enhance patient engagement. To this end, our study has several significant findings. First, our study indicates that most portal users still do not use a health app. This finding aligns with the ONC report that states that only a small percentage of hospitals offers access to portals through health apps [11]. The positive effect of app use on patient engagement calls for more comprehensive portal mobile access functionalities. Second, regardless of how they access their medical records, many patients are still not engaging in any functionalities. Our findings show that the proportion of individuals who have been granted access privileges but do not access their portals is higher among non–app users.

Third, using a health app to access portals strongly and positively impacts the use of the multiple engagement functionalities. The regression coefficients of VDT, PPI, and PPHI engagement functionalities were positive and significant. Furthermore, Figure 2 shows that app users are more likely to use at least one engagement task and have higher usage of VDT, PPI, and PPHI tasks. This finding extends extant research that indicates that app users log into their portals sooner and more frequently than non–app users [41] by showing that beyond timeliness and frequency of use, the use of apps to access portals is associated with the breadth of the usage of portal functionalities. Fourth, the association of portal access type on individual portal tasks still favors non–app users. Portal usage of app users for individual tasks was greater only for 3 out of 8 functionalities, namely, using medical records to make care decisions, downloading, and transmitting results. Downloading and sharing medical records are associated with increased efficiency in clinical workflows [48]. Offering app access to patients who underuse these functionalities could help improve their portal usage.

Fifth, the impacts of mobile health app usage on the use of diverse functionalities fade with age only for VDT tasks. This finding is in line with that of previous research that the adoption of newer technologies and the breadth of use of technology tends to decline with age, even though users continue to frequently use technologies they are acquainted with [49]. However, this study shows that the age group of portal users does not condition the impact of access type on PPI and PPHI tasks. This finding extends previous research on the effects of age on mobile health app usage by showing that the result of mobile access is not uniform across all age groups and tasks. While mobile access increases VDT tasks for users younger than 65 years and decreases them, for others, it does not impact portal users differently on the breadth of use of interaction functionalities regardless of their age groups. App access increases the usage of interaction tasks similarly for all users.

Mobile health app access could be a better intervention for portal users younger than 65 years on VDT tasks and all portal users for interaction tasks. Understanding the nuances in using portal capabilities is essential for providers to develop equitable interventions for each age group and to increase the use of underused functionalities. Other interventions, such as providers' encouragement that have been associated with higher usage of portals [36], could be used among patients who are aged 65 years or older to increase the use of VDT tasks.

Logging into portals is only a first step toward patient engagement through their medical records. Once logged in, the number of tasks patients perform contributes to their empowerment and engagement. In this study, while a high percentage of users (46.5% and 62.8% of app users and non–app users, respectively) performed at least one VDT task, less than 20% of users in both groups engaged in all 3 VDT tasks. This low involvement with VDT tasks was more pronounced among non–app users. Similarly, less than 5% and less than 40% in each group performed all PPHI and PPI tasks, respectively. Additional research is needed to improve patient engagement with multiple portal functionalities.

Mobile apps provide convenience for their users [50]. For the most part, the effort required to perform VDT tasks is minimal. Therefore, VDT tasks can be more easily accomplished with a click or touch of a button for a health app designed with usability considerations. The same is not valid for interaction tasks that require the use of a computer or cellphone keyboard. Prior research comparing the use of smartphones with that of desktops in business settings shows that users of smartphones can perform reading tasks better than typing tasks [51]. However, the need to manipulate the keyboard and the smaller window size of a smartphone increase the time and difficulty of performing such tasks. Hence, performing interaction tasks via a health app is not as convenient as doing so via a computer and may explain the lesser impact of app usage on the interaction functionalities, as shown in Table 3. While app usage was significantly and positively associated with all engagement types, its impact on the VDT tasks was far more significant than that on the interaction tasks, as reflected in the regression coefficients. It is also possible that app interfaces were not easy to use to port data from one provider to another. Therefore, there is a need for apps to connect with platforms that make data portability seamless.

Even though app usage predicted all engagement types, the easier a task that could be performed via an app, the more it was used. Previous studies have recommended that portal designers simplify data entry into their systems [52,53]. App designers also should endeavor to decrease the amount of effort needed to perform interaction tasks, especially PPHI tasks, which are essential in this new era of interoperability across health information systems and information exchanges [19]. The ability of users to amend or request amendments to their records could be vital in reconciling fragmented information from one system to another and ensuring data integrity and completeness.

Limitations

This study used secondary data. Thus, covariates were limited to variables available in the data. Similarly, this study was based on cross-sectional data and cannot infer causality. Future research could examine the impact of health app usage on the portal engagement functionalities using longitudinal data.

These secondary data did not provide details on the functional characteristics of the mobile apps and portals that were used. It was also unclear whether the portals were tethered to an electronic health record. Since the data were representative of the national US population [54], this study assumed that the mobile health apps and portals used by the study population were diverse.

Conclusions

Patients with access to mobile-optimized portals log into their medical records sooner after being granted access and more frequently than those who only use a computer [38]. This study shows that patients who use a health app to access their records are also more likely to engage in multiple VDT, PPI, and PPHI tasks. The convenience and wide availability of health apps can also improve VDT functionalities among adults younger than 65 years and interaction functionalities among all portal users. Previous studies show that engagement with portals leads to better health outcomes and effective and efficient care delivery [21,22]. Although the likelihood of engaging in at least one task is higher when using an app, portal usage remains low. More research is needed to determine other factors and characteristics of health apps, which could lead to greater portal usage, especially among adults aged 65 years or older.

Data Availability

All data analyzed during this study are publicly available on the HINTS website [54].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions and variable measurements.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v24i12e41972_app1.xlsx]

References

1. Butler J, Gibson B, Lewis L, Reiber G, Kramer H, Rupper R, et al. Patient-centered care and the electronic health record: exploring functionality and gaps. JAMIA Open 2020 Oct;3(3):360-368 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa044] [Medline: 33215071]
2. Becker KL. Tell me your dreams and goals: Structuring communication exchanges to improve patient-centered care with chronic pain patients. Appl Nurs Res 2020 Jun;53:151248. [doi: 10.1016/j.apnr.2020.151248] [Medline: 32451006]
3. Charles D, Gabriel M, Henry JW. Electronic capabilities for patient engagement among U.S. Non-Federal acute care hospitals: 2012-2014. Office of the National Coordinator for Health Information Technology. 2015. URL: https://www.healthit.gov/sites/default/files/briefs/oncdatabrief29_patientengagement.pdf [accessed 2022-12-14]
4. Alsahafi YA, Gay V. An overview of electronic personal health records. Health Policy Technol 2018 Dec;7(4):427-432. [doi: 10.1016/j.hlpt.2018.10.004]
5. Bouayad L, Ialynytchev A, Padmanabhan B. Patient health record systems scope and functionalities: literature review and future directions. J Med Internet Res 2017 Nov 15;19(11):e388 [FREE Full text] [doi: 10.2196/jmir.8073] [Medline: 29141839]
6. Alturkistani A, Qavi A, Anyanwu PE, Greenfield G, Greaves F, Costelloe C. Patient portal functionalities and patient outcomes among patients with diabetes: systematic review. J Med Internet Res 2020 Sep 22;22(9):e18976 [FREE Full text] [doi: 10.2196/18976] [Medline: 32960182]
7. Patel V, Barker W, Siminerio E. Individuals' access and use of their online medical record nationwide. Office of the National Coordinator for Health Information Technology. 2014. URL: https://www.healthit.gov/sites/default/files/consumeraccessdatabrief_9_10_14.pdf [accessed 2022-12-14]
8. Patel V, Johnson C. Trends in individuals' access, viewing and use of online medical records and other technology for health needs: 2017-2018. Office of the National Coordinator for Health Information Technology. 2019. URL: https://www.healthit.gov/sites/default/files/page/2019-05/Trends-in-Individuals-Access-Viewing-and-Use-of-Online-Medical-Records-and-Other-Technology-for-Health-Needs-2017-2018.pdf [accessed 2022-12-14]
9. Lin SC, Lyles CR, Sarkar U, Adler-Milstein J. Are patients electronically accessing their medical records? Evidence from national hospital data. Health Aff (Millwood) 2019 Nov 01;38(11):1850-1857. [doi: 10.1377/hlthaff.2018.05437] [Medline: 31682494]
10. Heart T, Ben-Assuli O, Shabtai I. A review of PHR, EMR and EHR integration: a more personalized healthcare and public health policy. Health Policy Technol 2017 Mar;6(1):20-25. [doi: 10.1016/j.hlpt.2016.08.002]
11. Henry J, Barker W, Kachay L. Electronic capabilities for patient engagement among U.S. non-Federal acute care hospitals: 2013-2017. Office of the National Coordinator for Health Information Technology. 2019. URL: https://www.healthit.gov/sites/default/files/page/2019-04/AHApatientengagement.pdf [accessed 2022-12-14]
12. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]
13. Lyles C, Sarkar U, Schillinger D, Ralston JD, Allen JY, Nguyen R, et al. Refilling medications through an online patient portal: consistent improvements in adherence across racial/ethnic groups. J Am Med Inform Assoc 2016 Apr;23(e1):e28-e33 [FREE Full text] [doi: 10.1093/jamia/ocv126] [Medline: 26335983]
14. Shimada S, Petrakis B, Rothendler J, Zirkle M, Zhao S, Feng H, et al. An analysis of patient-provider secure messaging at two Veterans Health Administration medical centers: message content and resolution through secure messaging. J Am Med Inform Assoc 2017 Sep 01;24(5):942-949 [FREE Full text] [doi: 10.1093/jamia/ocx021] [Medline: 28371896]
15. Agrawal L, Ndabu T, Mulgund P, Sharman R. Factors affecting the extent of patients' electronic medical record use: an empirical study focusing on system and patient characteristics. J Med Internet Res 2021 Oct 28;23(10):e30637 [FREE Full text] [doi: 10.2196/30637] [Medline: 34709181]
16. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 02;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
17. Sarkar U, Bates DW. Care partners and online patient portals. JAMA 2014 Jan 22;311(4):357-358. [doi: 10.1001/jama.2013.285825] [Medline: 24394945]
18. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012 Sep 01;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]
19. Smith SG, Pandit A, Rush SR, Wolf MS, Simon C. The association between patient activation and accessing online health information: results from a national survey of US adults. Health Expect 2015 Dec 05;18(6):3262-3273 [FREE Full text] [doi: 10.1111/hex.12316] [Medline: 25475371]
20. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
21. Graetz I, Huang J, Brand RJ, Hsu J, Yamin CK, Reed ME. Bridging the digital divide: mobile access to personal health records among patients with diabetes. Am J Manag Care 2018 Jan;24(1):43-48 [FREE Full text] [Medline: 29350505]

22. Rief JJ, Hamm ME, Zickmund SL, Nikolajski C, Lesky D, Hess R, et al. Using Health Information Technology to Foster Engagement: Patients' Experiences with an Active Patient Health Record. Health Commun 2017 Mar 25;32(3):310-319. [doi: 10.1080/10410236.2016.1138378] [Medline: 27223684]
23. Feldman SS, Buchalter S, Hayes LW. Health Information Technology in Healthcare Quality and Patient Safety: Literature Review. JMIR Med Inform 2018 Jun 04;6(2):e10264 [FREE Full text] [doi: 10.2196/10264] [Medline: 29866642]
24. Fossa A, Bell S, DesRoches C. OpenNotes and shared decision making: a growing practice in clinical transparency and how it can support patient-centered care. J Am Med Inform Assoc 2018 Sep 01;25(9):1153-1159 [FREE Full text] [doi: 10.1093/jamia/ocy083] [Medline: 29982659]
25. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
26. Lyles CR, Allen JY, Poole D, Tieu L, Kanter MH, Garrido T. "I Want to Keep the Personal Relationship With My Doctor": Understanding Barriers to Portal Use among African Americans and Latinos. J Med Internet Res 2016 Oct 03;18(10):e263 [FREE Full text] [doi: 10.2196/jmir.5910] [Medline: 27697748]
27. Ancker J, Hafeez B, Kaushal R. Socioeconomic disparities in adoption of personal health records over time. Am J Manag Care 2016 Aug;22(8):539-540 [FREE Full text] [Medline: 27541700]
28. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN. Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 2014 Oct 24;16(10):e241 [FREE Full text] [doi: 10.2196/jmir.3371] [Medline: 25344621]
29. Mafi J, Mejilla R, Feldman H, Ngo L, Delbanco T, Darer J, et al. Patients learning to read their doctors' notes: the importance of reminders. J Am Med Inform Assoc 2016 Sep;23(5):951-955 [FREE Full text] [doi: 10.1093/jamia/ocv167] [Medline: 26911830]
30. Lyles CR, Tieu L, Sarkar U, Kiyoi S, Sadasivaiah S, Hoskote M, et al. A Randomized Trial to Train Vulnerable Primary Care Patients to Use a Patient Portal. J Am Board Fam Med 2019 Mar 08;32(2):248-258 [FREE Full text] [doi: 10.3122/jabfm.2019.02.180263] [Medline: 30850461]
31. Grossman L, Masterson Creber RM, Benda N, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
32. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016 Apr 19;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
33. Fernandez N, Copenhaver DJ, Vawdrey DK, Kotchoubey H, Stockwell MS. Smartphone use among postpartum women and implications for personal health record utilization. Clin Pediatr (Phila) 2017 Apr 22;56(4):376-381. [doi: 10.1177/0009922816673438] [Medline: 27798390]
34. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: an empirical analysis of a national cancer-related survey. Decis Support Syst 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
35. Kisekka V, Giboney JS. The Effectiveness of Health Care Information Technologies: Evaluation of Trust, Security Beliefs, and Privacy as Determinants of Health Care Outcomes. J Med Internet Res 2018 Apr 11;20(4):e107 [FREE Full text] [doi: 10.2196/jmir.9014] [Medline: 29643052]
36. Shimoga S, Lu Y. Role of provider encouragement on patient engagement via online portals. J Am Med Inform Assoc 2019 Oct 01;26(10):968-976 [FREE Full text] [doi: 10.1093/jamia/ocz026] [Medline: 30925585]
37. Mobile App Trends: How publishers are navigating stricter privacy regulations. Insider Intelligence. URL: https://www.emarketer.com/Article/eMarketer-Unveils-New-Estimates-Mobile-App-Usage/1015611 [accessed 2020-11-19]
38. Health Information National Trends Survey 5 (HINTS 5): Cycle 3 Methodology Report. Westat. URL: https://hints.cancer.gov/docs/Instruments/HINTS5_Cycle3_MethodologyReport.pdf [accessed 2022-11-02]
39. Nelson D, Kreps G, Hesse B, Croyle R, Willis G, Arora N, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004 Sep;9(5):443-460. [doi: 10.1080/10810730490504233] [Medline: 15513791]
40. Turner K, Clary A, Hong Y, Alishahi Tabriz A, Shea CM. Patient Portal Barriers and Group Differences: Cross-Sectional National Survey Study. J Med Internet Res 2020 Sep 17;22(9):e18870 [FREE Full text] [doi: 10.2196/18870] [Medline: 32940620]
41. Graetz I, Huang J, Brand R, Hsu J, Reed ME. Mobile-accessible personal health records increase the frequency and timeliness of PHR use for patients with diabetes. J Am Med Inform Assoc 2019 Jan 01;26(1):50-54 [FREE Full text] [doi: 10.1093/jamia/ocy129] [Medline: 30358866]
42. Strekalova YA. Electronic health record use among cancer patients: Insights from the Health Information National Trends Survey. Health Informatics J 2019 Mar 23;25(1):83-90 [FREE Full text] [doi: 10.1177/1460458217704246] [Medline: 28434278]
43. Chae J. How we use the Internet matters for health: The relationship between various online health-related activities and preventive dietary behaviors. Health Informatics J 2019 Sep 19;25(3):973-983 [FREE Full text] [doi: 10.1177/1460458217735675] [Medline: 29047292]

44. Seo D, Park YR, Lee Y, Kim JY, Park J, Lee J. The use of mobile personal health records for hemoglobin A1c regulation in patients with diabetes: retrospective observational study. J Med Internet Res 2020 Jun 02;22(6):e15372 [FREE Full text] [doi: 10.2196/15372] [Medline: 32484447]
45. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
46. Gerard M, Chimowitz H, Fossa A, Bourgeois F, Fernandez L, Bell SK. The importance of visit notes on patient portals for engaging less educated or nonwhite patients: survey study. J Med Internet Res 2018 May 24;20(5):e191 [FREE Full text] [doi: 10.2196/jmir.9196] [Medline: 29793900]
47. Sarkar U, Lyles C, Parker M, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
48. O'Donnell A, Kaner E, Shaw C, Haighton C. Primary care physicians' attitudes to the adoption of electronic medical records: a systematic review and evidence synthesis using the clinical adoption framework. BMC Med Inform Decis Mak 2018 Nov 13;18(1):101 [FREE Full text] [doi: 10.1186/s12911-018-0703-x] [Medline: 30424758]
49. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of technology: frequency of use for younger and older adults. Ageing Int 2011 Mar 28;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
50. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
51. Adepu S, Adler R. A comparison of performance and preference on mobile devices vs. desktop computers. 2016 Presented at: 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON); October 20-22, 2016; New York, NY. [doi: 10.1109/uemcon.2016.7777808]
52. Clarke MA, Schuetzler RM, Windle JR, Pachunka E, Fruhling A. Usability and cognitive load in the design of a personal health record. Health Policy Technol 2020 Jun;9(2):218-224. [doi: 10.1016/j.hlpt.2019.10.002]
53. Fernando M, Fidge C, Sahama T. Design guidelines for effective occupation-based personal health records. 2020 Presented at: ACSW '20: Australasian Computer Science Week 2020; February 4-6, 2020; Melbourne. [doi: 10.1145/3373017.3373042]
54. What is HINTS? National Cancer Institute. Health Information National Trends Survey. URL: https://hints.cancer.gov/ [accessed 2020-01-01]

Abbreviations

EFA: exploratory factor analysis
HINTS: Health Information National Trends Survey
ONC: Office of the National Coordinator
PHI: personal health information
PPHI: patient–personal health information interaction
PPI: patient-provider interaction
VDT: view, download, and transmit

Edited by T Leung; submitted 16.08.22; peer-reviewed by S Van den Bulck, C Smith; comments to author 08.09.22; revised version received 06.11.22; accepted 25.11.22; published 27.12.22.

Please cite as:
Ndabu T, Agrawal L, Sharman R
The Role of Access Type and Age Group in the Breadth of Use of Patient Portals: Observational Study
J Med Internet Res 2022;24(12):e41972
URL: https://www.jmir.org/2022/12/e41972
doi:10.2196/41972
PMID:36574284

Original Paper

Accuracy and Systematic Biases of Heart Rate Measurements by Consumer-Grade Fitness Trackers in Postoperative Patients: Prospective Clinical Trial

Philipp Helmer[1], MD; Sebastian Hottenrott[1], MD; Philipp Rodemers[1]; Robert Leppich[2], MSc; Maja Helwich[1]; Rüdiger Pryss[3], Prof Dr; Peter Kranke[1], Prof Dr; Patrick Meybohm[1], Prof Dr; Bernd E Winkler[1*], MD; Michael Sammeth[1,4*], Prof Dr

[1]Department of Anaesthesiology, Intensive Care, Emergency and Pain Medicine, University Hospital Würzburg, Würzburg, Germany

[2]Department of Software Engineering, Faculty of Computer Science, University of Würzburg, Würzburg, Germany

[3]Institute for Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

[4]Department of Applied Sciences, Coburg University, Coburg, Germany

[*]these authors contributed equally

Corresponding Author:
Philipp Helmer, MD
Department of Anaesthesiology, Intensive Care, Emergency and Pain Medicine
University Hospital Würzburg
Oberdürrbacher Str. 6
Würzburg, 97070
Germany
Phone: 49 93120130574
Email: helmer_p@ukw.de

Abstract

Background: Over the recent years, technological advances of wrist-worn fitness trackers heralded a new era in the continuous monitoring of vital signs. So far, these devices have primarily been used for sports.

Objective: However, for using these technologies in health care, further validations of the measurement accuracy in hospitalized patients are essential but lacking to date.

Methods: We conducted a prospective validation study with 201 patients after moderate to major surgery in a controlled setting to benchmark the accuracy of heart rate measurements in 4 consumer-grade fitness trackers (Apple Watch 7, Garmin Fenix 6 Pro, Withings ScanWatch, and Fitbit Sense) against the clinical gold standard (electrocardiography).

Results: All devices exhibited high correlation ($r \geq 0.95$; $P < .001$) and concordance ($r_c \geq 0.94$) coefficients, with a relative error as low as mean absolute percentage error <5% based on 1630 valid measurements. We identified confounders significantly biasing the measurement accuracy, although not at clinically relevant levels (mean absolute error<5 beats per minute).

Conclusions: Consumer-grade fitness trackers appear promising in hospitalized patients for monitoring heart rate.

Trial Registration: ClinicalTrials.gov NCT05418881; https://www.clinicaltrials.gov/ct2/show/NCT05418881

(J Med Internet Res 2022;24(12):e42359) doi:10.2196/42359

KEYWORDS

health tracker; smartwatch; internet of things; personalized medicine; photoplethysmography; wearable; Garmin Fenix 6 Pro; Apple Watch 7; Fitbit Sense; Withings ScanWatch

Introduction

Fitness trackers are usually wrist-worn devices equipped with photoplethysmography (PPG) sensors and motion sensors, among complementary sensor units. These devices paved the way for continuous monitoring of diverse fitness parameters including various vital signs [1]. In contrast to conventional PPG measurement methods based on transmissive pulse oximetry (TPO), fitness trackers use reflective pulse oximetry. Therefore, wearing a finger clip is obsolete because both the light-emitting diode and the photodiode (light sensor) can be

combined side by side in one measuring unit that can be worn, for example, on the wrist, offering more mobility to users or patients.

The option of continuous heart rate monitoring without impairing the mobility of patients opens up a range of new opportunities, especially for hospitalized patients. For example, the Early Warning Score can be calculated from heart rate and other parameters and is used for the early detection of deterioration in patients [2]. Yet, vital signs are only monitored continuously in hospitalized patients requiring intensive care, as the technical, personal, and financial requirements do not enable the current methods to be expanded to a peripheral ward. Traditional monitoring also makes patients more difficult to mobilize, which runs counter to the idea of early rehabilitation according to the guidelines of the Enhanced Recovery After Surgery. Particularly, patients undergoing surgical procedures are a vulnerable patient cohort requiring close monitoring. Recently, 2 systematic reviews demonstrated that, based on continuous measuring of vital parameters in hospitalized patients, the length of stay in the hospital [3] and, in combination with automated alerting systems, even mortality [4] could be reduced. Such reports raise the evident question to which degree fitness trackers could be used in hospitalized patients for continuous monitoring of vital signs. Due to their general availability, cost efficiency, and long battery life, fitness trackers could offer a feasible solution. To date, fitness trackers have primarily been used for sports and leisure purposes [5], but their opportunities in the continuous monitoring of various vital signs during the entire hospital stay have already been highlighted [6].

Obviously, in order to establish fitness trackers in the medical sector, a rigorous validation of their measurement accuracy is of high importance. However, so far, relatively little effort has been made in this direction, and most of the currently available trials show one or more of the following shortcomings: the study was primarily conducted with healthy volunteers [7,8], it compared different devices with each other but not with an established medical gold standard [9], it examined non–consumer-grade wearables [10], and it assessed only a very limited sample size [11].

Studies on the use of fitness trackers in a perioperative setting or among patients with multiple pre-existing diseases are rare [12] and, according to systematic reviews, also hampered by a high risk of bias [13] and suffer from low quality [14]. In particular, it has been shown that motion artifacts influence the mean absolute error (MAE) of the measurements by up to 30% [15]. In order to exclude such interferences, we evaluated the accuracy of vital signs measured by fitness trackers in resting patients. We therefore set up—for the very first time—a study that aims to benchmark the heart rate measurements of 4 consumer-grade fitness trackers against the clinical gold standard under controlled conditions in postoperative patients undergoing moderate to major surgery.

Methods

Study Design

The primary objective of our study is the evaluation of the accuracy of heart rate measurements by consumer-grade fitness trackers against the clinical gold standard. The study population consisted of nonsedated postoperative patients who had undergone moderate to major surgery. This prospective validation study took place at the Department of Anaesthesiology, Intensive Care, Emergency and Pain Medicine at the University Hospital Würzburg, Germany, between November 2021 and May 2022. The study protocol was designed in accordance with the guidelines for wrist-worn consumer wearables [16]. This paper presents the results of the heart rate validation in the "Monitor Trial," registered on ClinicalTrials.org (accession No. NCT05418881).

Patients (aged ≥18 years old) scheduled for elective surgery requiring placement of an arterial line were screened prior to the procedure. Exclusion criteria for participation included critically ill patients (ie, American Society of Anesthesiologists V [ASA V]), those with a BMI of >40 kg/m^2, outpatient surgery, infectious patients (due to hygienical regulations), those who previously participated in this study, those incapable of giving written informed consent, those who did not speak and read German, and those with extensive pathological skin lesions at the forearms or with known allergies to latex, silicone, or nickel.

Ethical Considerations

The study protocol had been reviewed and approved by the local ethics committee of Würzburg (reference number 145/21_c). We conducted our study in accordance with good clinical practice guidelines and the Declaration of Helsinki. Our study was planned, carried out, analyzed, and interpreted independently of any industrial partners. All participants provided written informed consent before surgery took place.

Study Procedures

Following surgical procedures, the vital parameters of study participants were continuously monitored according to hospital standards during their stay at the postanesthesia care unit (PACU). We used medical-grade TPO at the finger as well as noninvasive and invasive blood pressure monitoring and 3-lead electrocardiography (ECG), all measured by Philips devices (IntelliVue X3, Philips Healthcare). The measured parameters were streamed to a bedside patient monitor (MX750, Philips Healthcare). Simultaneously, patients were equipped with 4 different consumer-grade fitness trackers (Table 1), attached randomly to either wrist according to the manufacturer's instructions. In doing so, we aimed to eliminate any systematic bias from our results, for example, small but potentially present differences in pulse measurements between the 2 hands. During a patient's stay at the PACU, a total of 3 on demand measurements were collected by 2 trained members of the research staff. The measured values were acquired manually from the screens of the fitness trackers and the bedside monitors (ECG and TPO) simultaneously. Patients who had no arterial line placed or those who were admitted to an intensive care unit immediately (eg, sedated, ventilated, or temporarily critically

ill patients) were excluded. The placement of an arterial line ensured that only patients with moderate to major surgery were included.

In order to set up each of the fitness trackers, an anonymized, patient-unrelated user account had been created at the corresponding manufacturer. Immediately after the initial setup, the firmware of each device was updated (Table 1). Subsequently, the connection via Bluetooth and Wi-Fi was deactivated to ensure that no further firmware updates were installed during the course of the study, preventing any possible changes to algorithms from affecting the results [17,18]. Of note, although some of the manufacturers offer customized firmware for research purposes, we decided to stick to the consumer-grade firmware to enable the comparability of our results with complementary studies.

Table 1. Wrist-worn consumer-grade fitness trackers investigated in this trial, specified by the respective manufacturer (headquarters' address), the device's model, and the firmware version used for the study.

Manufacturer	Model	Firmware version
Apple	Watch 7	watchOS8.1
Fitbit	Sense	5.3 (44.128.6.12)
Garmin	Fenix 6 Pro	19.20 (0fe794a)
Withings	ScanWatch	2291

Data Collection

Patient characteristics were recorded after performing measurements according to the guidelines for wrist-worn devices [16], including age, sex, wrist circumference, BMI, height, body weight, ASA classification, Fitzpatrick scale, and heart rhythm. As there is no generally established metrics for the density of forearm hair, we segregated the forearm hairiness of patients into 4 categories—0: no forearm hair; 1: minimal; 2: moderate; and 3: extensive hairiness. Measurements of the devices were recorded manually and transferred to an Excel (Microsoft Corp) spreadsheet later on.

Statistical Analysis

If not further specified, all statistical analyses were carried out using standard R (version 4.2.0; R Core Team) functions and using the ggplot2 package (version 3.3.6; MIT license) for visualization. For descriptive analysis of the patient cohort, we assessed the median and the IQR of each of the attributes. In addition to the fitness tracker measurements of the heart rate, TPO as the established clinical standard for heart frequency measurement was used as a control and compared to the ECG gold standard. We assessed the measurement accuracy of each device by Bland–Altman plots [19]. After visual inspection, we excluded 5 outliers from further analysis, defined as deviations of >30 beats per minute (bpm) between the gold standard and the respective benchmarked measurement. For all of the remaining paired data points (p_i,r_i), the absolute error (AE) was determined as abs($p_i - r_i$) and, inherently, the absolute percentage error as abs($p_i - r_i$) × 100/r_i, where r_i corresponds to the gold standard reference measurements by ECG. Correspondingly, MAE and mean absolute percentage error (MAPE) were computed according to standard definitions using the Metrics package (version 0.1.4).

For each of the benchmarked devices, we further computed the linear regression, determined the Pearson correlation coefficient (PCC) as r, and used the DescTools package (version 0.99.45) to determine the Lin concordance coefficient (CCC) as r_c. The PCC algorithm also provides the residual sum of squares (RSS) measure of discrepancy between the data and the prediction by the model. Comparing the distribution of benchmarked values with the distribution of gold standard reference measurements, we assessed the following hypotheses: (1) both data series are uncorrelated according to the Pearson model (standard association test, Cor-Test), (2) data are obtained from the same distribution (2-tailed Kolmogorov-Smirnov test), and (3) the 2 data vectors are shifted against each other (2-sample Mann-Whitney-Wilcoxon test). As all these tests are nonparametric, no further assumption on the nature of the compared distributions has been implied, and we generally accepted P<.05 as statistically significant.

Results

Overview of the Cohort

During the course of the study, 288 patients were screened (Figure 1A), of whom 201 gave written informed consent (initially excluded: n=87; Figure 1B). Subsequently, a further 89 patients were excluded (Figure 1B), resulting in 112 patients successfully included in the study (Figure 1C). For each of these 112 included patients (Figure 1C), 3 attempts of measurement by each measuring method (ECG, TPO, Apple, Fitbit, Garmin, and Withings) were performed. This resulted in 2016 measurements, of which the 336 gold standard measurements (ECG) served as a reference to evaluate the remaining 1680 measurements by the benchmarked devices. Some of these measurements failed (n=45) and were classified as "dropouts." After quality control, we removed another 5 measurements (2 TPO, 2 Fitbit, and 1 Withings), obtaining a final data set comprising 1630 data points (Figure 1D).

In our cohort, 62.5% (n=70) of participants were male and 37.5% (n=42) were female. The median age of patients was 68 years, height 172 cm, weight 77 kg, BMI 26.4 kg/m^2, and wrist circumference 18 cm. Patients were further stratified by ASA score, skin pigmentation (Fitzpatrick scale), and a custom scale on the degree of hairiness on their forearm (Table 2, Figure S1 in Multimedia Appendix 1).

Most of the patients (n=92; 82.1%) presented with sinus heart rhythm during the measurements; hence, merely 20 (17.9%)

patients presented with arrhythmias. Of them, 10 patients presented with atrial fibrillation, 5 with pacemaker-triggered ECG, 1 with bigeminus, 1 with clustered extrasystoles, 1 with a left ventricular assist device, and 2 patients were not further classified by the attending physician. No adverse or serious adverse events were observed during the study.

Figure 1. Study design. Flowchart of patient recruitment and data acquisition. After (A) screening and initially excluding patients, (B) 201 patients gave written informed consent. Of them, (C) 112 patients were successfully included in the study, resulting in 1680 benchmark measurements. Disregarding (C) missing data due to dropouts and (D) removing outliers during quality control resulted in the analyzed data set of 1630 data points. ICU: intensive care unit; m: number of measurements; n: number of patients; TPO: transmissive pulse oximetry.

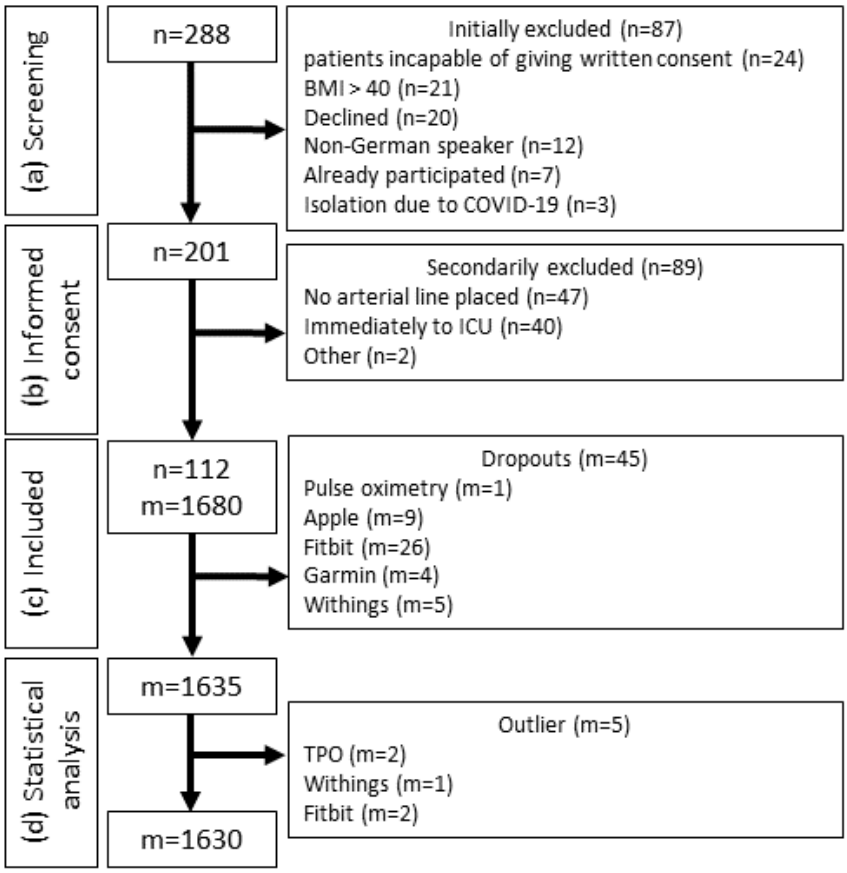


Table 2. Attributes of the patient cohort.

	Value, median (IQR)	Range (minimum-maximum)
Age (years)	68 (58-74)	24-92
Wrist circumference (cm)	18 (17-19)	15-23
BMI (kg/m^2)	26.4 (24.05-30.18)	17.7-39.1
Height (cm)	172 (165-176)	152-192
Weight (kg)	77 (68-90)	45-122
ASA[a]	2 (2-3)	1-4
Fitzpatrick scale	2 (2-3)	1-4
Degree of forearm hair density	1 (0-2)	0-3

[a]ASA: American Society of Anesthesiologists.

Overall Deviation

We used the 1630 valid measurements to determine the general deviation of the heart frequency measured by fitness trackers compared to the clinical gold standard. To this end, we first computed the cumulative dropout rate (CDR), taking failed measurements and data points removed during quality control into account. TPO showed the lowest dropouts (CDR<1%) among the benchmarked devices, whereas the measurements of fitness trackers yielded CDR>1%, ranging from 1.2% (Garmin) to 8.3% (Fitbit) (Table 3).

Table 3. Overall deviation of fitness tracker heart rate measurements and the clinical gold standard.

	Philips	Apple	Fitbit	Garmin	Withings
Valid measurement points, n	333	327	308	332	330
Failed measurements, n	3	9	28	4	6
CDR[a] (%)	0.89	2.67	8.33	1.19	1.79
MAE[b]	0.92	1.59	2.31	2.47	1.71
MAPE[c] (%)	1	2	4	3	2
Bias (95% CI)	–0.25 (–0.42 to –0.08)	0.36 (0.09 to 0.63)	0.77 (0.28 to 1.26)	–1.21 (–1.65 to –0.77)	0.05 (–0.28 to 0.40)

[a]CDR: cumulative dropout rate.

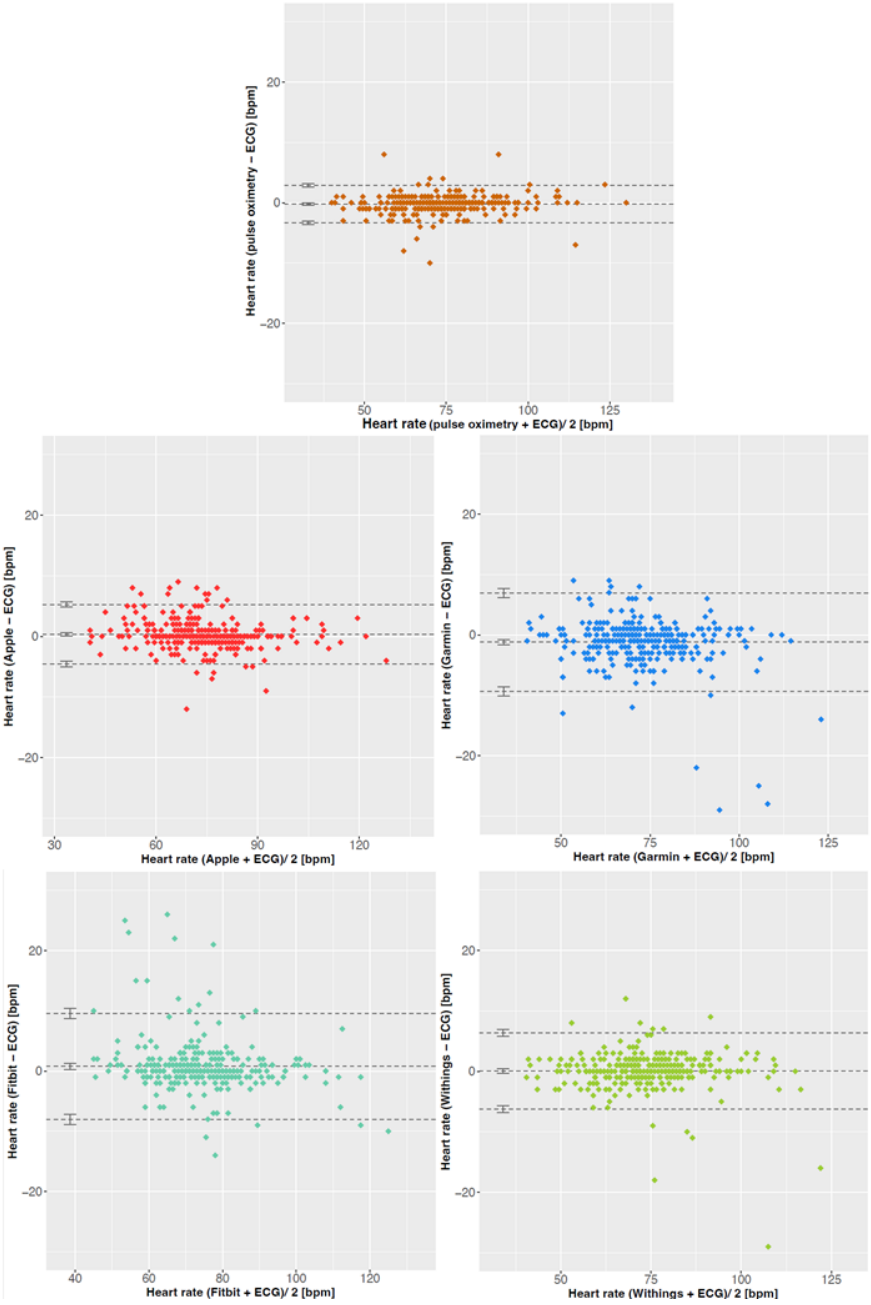
[b]MAE: median absolute error.

[c]MAPE: mean absolute percentage error.

Next, we calculated the MAE and the relativized indicator of the MAPE between all paired measurements of a benchmarked device and the reference values. As it can be assumed that the measurements by the TPO meet clinical standards, these measurements were used as a positive control of performing the measurements accurately. As anticipated, the correlation between the measurement results of TPO and ECG was very high (r=0.99; P<.001) with an MAE of <1 bpm. TPO performs better than the fitness trackers, with an absolute deviation of ~1.5 to ~2.5 bpm on average. However, the deviation by fitness tracker measurements is overall not clinically relevant. The marginal character of the deviation is further underlined by MAPE values not reaching 5% for any of the benchmarked devices. Of note, MAPE indicators are not always proportional to the CDR indicators determined for each of the devices. Although Fitbit shows the highest CDR and MAPE, Apple exhibits the second-highest CDR but has one of the lowest MAPE (Table 3).

The overall bias and the SD of the measurements by the benchmarked trackers based on the ECG reference values were determined by Bland–Altman plots (Figure 2, Table S2 in Multimedia Appendix 1). The Withings tracker readings showed even less deviation from the reference than the TPO measurements (–0.25 vs 0.05; Table 3), although exhibiting an SD twice as high. Thereby, the high SD values resulted from outliers (deviation >10 bpm or even of >20 bpm), hampering particularly the Fitbit, Garmin, and Withings measurements (Figure 2). However, no systematic biases of these outliers toward high or low measurements could be identified. Overall, tracker measurements are more frequently biased to estimate higher values compared to the gold standard (ie, for Apple, Fitbit, and Withings). However, the Garmin device exhibits the absolute highest bias in the opposite direction; that is, underestimating the true heart rate. Connected by their calculation, SDs rank expectedly similar to the MAPE indicators (Table 3).

Figure 2. Bland–Altman plots presenting systematic bias of the investigated fitness trackers compared to ECG with the upper and lower limits of agreement and their respective CIs (upper and lower dashed line), as well as bias with the CI (middle dashed line). bpm: beats per minute; ECG: electrocardiography.



Linear Agreement

In addition, the first-order correlation between benchmarked heart rate measurements and the ECG reference values was assessed. All benchmarked devices exhibited a good linear fitting of the paired data vectors, with data points scattered closely around a straight line (Figure 3). This is directly reflected by the PCCs (r) computed on each pair of vectors, where, in agreement with our previous results, TPO yielded the highest correlation coefficient (r=0.99), followed closely by Apple (r=0.98), Withings (r=0.97), Garmin (r=0.96), and Fitbit (r=0.95).

Due to the numerical proximity of the highly condensed PCC values, we also considered the RSS measures, constituting the base values for computing r. As can be seen from Table 4, RSS values are able to resolve more precisely the spread observed in each of the scatter plots (Figure 3), ranking the variability of measurements by the benchmarked devices more clearly from low to high: TPO (RSS=803), Apple (RSS=1830), Withings (RSS=3106), Garmin (RSS=4757), and Fitbit (RSS=5133).

The high values we observe for the PCCs indicate a strong linear fit, but do not provide further details about the slope and shift of the linear dependency. Comparing these indicators of a correspondingly regressed linear model reveals shifts of <10 and slopes of approximately 1 for each of the benchmarked devices (Table 4). We also computed CCC as a measure of deviation from direct proportionality (ie, $y=x$), obtaining coherent coefficients close to 1 (Table 4).

Based on these results, it is not surprising that assessing statistically the hypothesis of data being correlated (*C* test) yields a very low $P<10^{-100}$ (Table 4). We used a 2-tailed Kolmogorov-Smirnov test, which supported, with a *P* value of .01, the hypothesis that the TPO measurements are pairwisely indistinguishable from the distribution of ECG reference values. This highlights a very high concordance of the measurements obtained by consumer-grade tracker devices; for example, TPO and ECG (Table 4).

Figure 3. Scatter plots demonstrating good linear agreement and low dispersion between the heart rate measurements by the fitness trackers (y-axis) compared to electrocardiography (ECG) (x-axis). The respective devices are color coded. bpm: beats per minute.

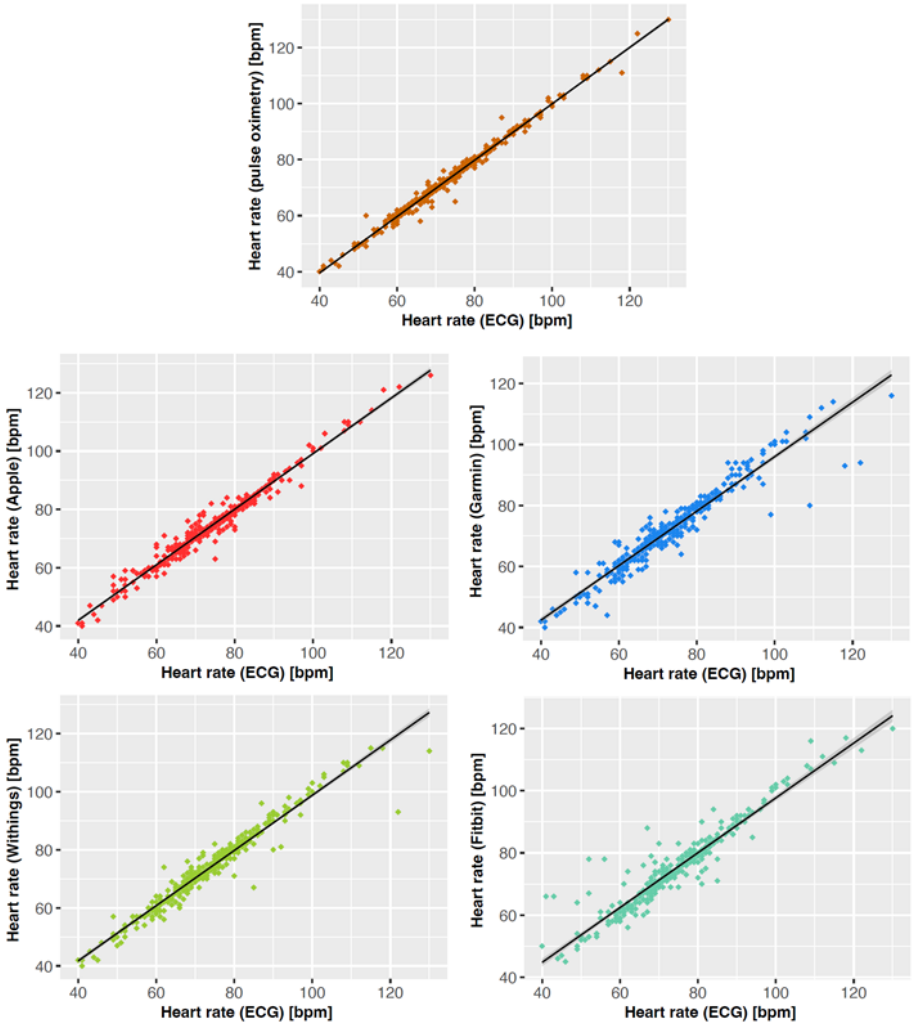


Table 4. Assessment of linear correlation.

Indicator	TPO[a]	Apple	Fitbit	Garmin	Withings
PCC[b], *r* (95% CI)	0.99 (0.99-0.99)	0.98 (0.98-0.99)	0.95 (0.93-0.96)	0.96 (0.95-0.96)	0.97 (0.97-0.98)
RSS[c]	804	1830	5133	4757	3106
P value (*C* test)	5.7×10^{-317}	2.15×10^{-247}	2.4×10^{-153}	1.26×10^{-177}	2.78×10^{-212}
Slope	1.01	0.95	0.88	0.89	0.95
Shift	–0.72	3.81	9.49	6.68	3.72
CCC[d], r_c (95% CI)	0.99 (0.99-0.99)	0.98 (0.98-0.99)	0.94 (0.93-0.95)	0.95 (0.94-0.96)	0.97 (0.97-0.98)

[a]TPO: transmissive pulse oximetry.

[b]PCC: Pearson correlation coefficient.

[c]RSS: residual sum of squares.

[d]CCC: Lin concordance coefficient.

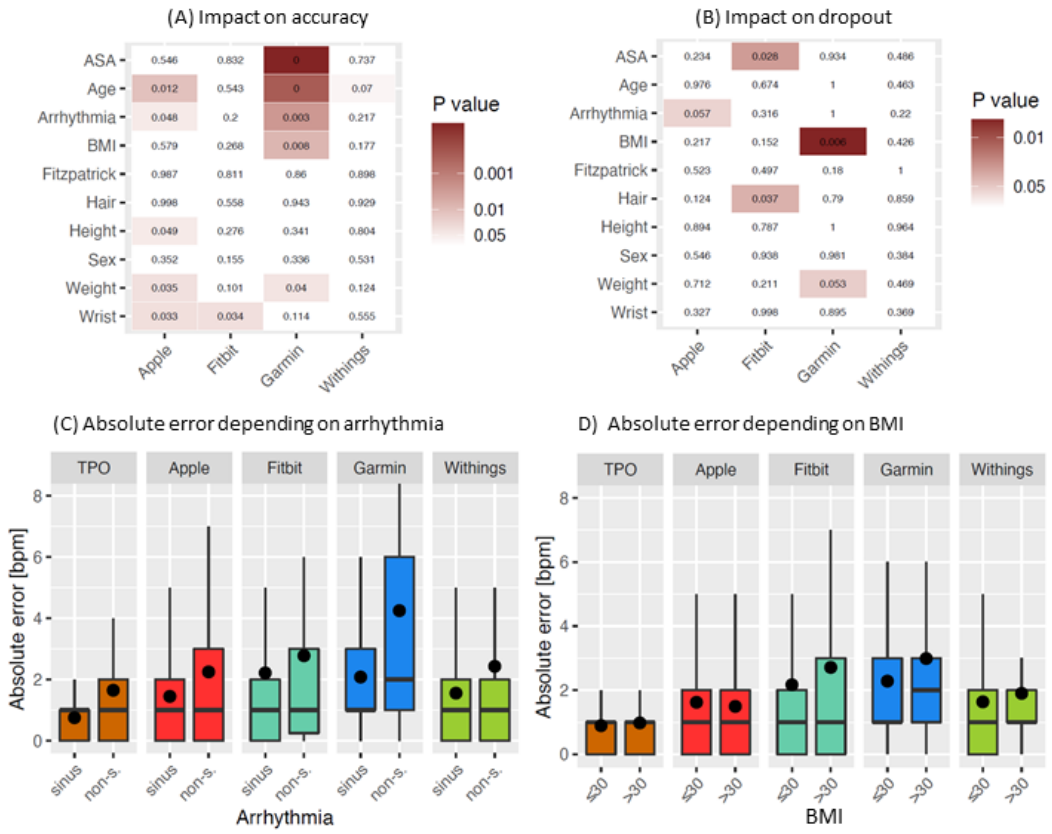
Systematic Biases

We searched for systemic factors influencing the measurement accuracy of the different fitness trackers. To this end, we divided each of the attributes recorded from the patients (Table 2) into 2 subgroups (Table S1 in Multimedia Appendix 1). In theory, an adverse factor can impact the measurements of a device in two ways: (1) either the measured value is influenced negatively, resulting in a higher observed error compared to the ECG reference (ie, impact on accuracy), or (2) the device is perturbed by the factor that no measurement is produced at all (ie, impact on dropout). In order to investigate both possibilities in a comparable manner, we used, on the one hand, Mann-Whitney-Wilcoxon tests to assess the distribution of AEs in group 1 versus 2 and, on the other hand, the Fisher exact test to assess the change in dropouts between both groups.

Figures 4A and 4B summarizes the results of our analyses. Respective box plots are presented in Figure S2 in Multimedia Appendix 1. As expected, the observed deviations in accuracy as well as changes in the number of dropouts are far from statistical significance when comparing male participants with female participants. More surprisingly, dividing patients according to the Fitzpatrick scale assigned to their skin tonality did not lead to the observation of significant differences in any indicator. Using a significance threshold of P=.05, we identified higher ASA scores, age, arrhythmias (Figure 4C), obesity (Figure 4D), and a wrist circumference of >18 cm as confounders, significantly worsening the accuracy of some tracker measurements (Figure 4A). Concordantly, higher ASA scores, obesity, and the hair density on the forearm exhibited significant differences in the number of dropouts (Figure 4B).

The identified confounders primarily affected the Garmin tracker. Particularly, negative impacts were seen in the higher age and higher ASA cohorts (Figure 4A), and in the arrhythmia and higher BMI cohorts (Figures 4C and 4D). Further, the Apple tracker exhibits negative influences by higher age and arrhythmia, albeit of less statistical significance. However, putting these statistics on scale with the total deviation, we found the largest bias caused by cardiac arrhythmia when using the Garmin tracker corresponding to an MAE of 2.17 bpm (Figure 4C). Although the presence of some confounders also increases the MAE of Fitbit and Withings measurements (Figures 4C and 4D), these differences were in general not significantly higher than errors of measurement in the background (Figures 4A and 4B).

Figure 4. Statistical assessment of measuring failures. Upper panels: heat maps visualizing the significance level of different attributes depending on the investigated wearables (1-sided Mann-Whitney U test). The darker the color, the lower the corresponding P value. (A) Attributes influencing the measurement accuracies of the investigated wearables with the respective P values. (B) Attributes influencing the dropout rates of the investigated devices with the respective P values. Lower panels: Box plots for the distribution of absolute errors in binary subgroups of patients, segregated according to their health status. (C) Arrhythmia sinus versus nonsinus rhythm. (D) BMI discriminating patients with obesity from those without obesity. bpm: beats per minute; TPO: transmissive pulse oximetry.



Discussion

Accuracy of the Heart Rate Measurements

The primary objective of our study was to evaluate the measurement accuracy of consumer-grade fitness trackers. According to Navalta et al [20], thresholds of MAPE≤5% and CCC≥0.90 can be considered as sufficiently high measurement accuracy. In our study, all of the benchmarked devices are within these threshold boundaries (Tables 3 and 4). In order to assess the clinical relevance of the deviations we observed between the benchmarked devices and the gold standard, we used the American National Standards Institute/Association for the Advancement of Medical Instrumentation standards for "cardiac monitors, heart rate meters, and alarms" (Association for the Advancement of Medical Instrumentation 2002) based on which an AE<5 bpm or relative (ie, percentage) error of <10% is required [21]. Our results (Figure 2) demonstrate that for each of the benchmarked devices, >92% of the measurements are within these limits (98.5% of TPO, 97.6% of Apple, 92.9% of Fitbit, 94% of Garmin, and 96.7% of Withings measurements). For upcoming trials, standardization of these thresholds is highly desirable in order to objectively decide on an "acceptable measurement accuracy" of a PPG-based device.

Overall, the measurement accuracy of consumer-grade fitness trackers is marginally inferior to that of TPO readings in postoperative patients while being at rest. However, the consumer-grade devices exhibit a wider dispersion in their measurements (Figure 3), as well as higher dropout rates than TPO (Table 3). Since the measurement accuracy of fitness trackers from different manufacturers depends on various technical details, we empirically tested potential confounders of heart rate measurements. Although we identified some factors that significantly decreased the accuracy of measurement (Figure 4), the observed deviations did not reach a clinically relevant level (MAE<5 bpm). To summarize, our observations support the use of fitness trackers for heart rate monitoring in postsurgical immobilized patients.

In general, the comparability of our results with previous studies is hampered by differences in methodological approaches, study designs, differences of the investigated collectives, etc. A systematic review estimated an overall MAPE between 1% and 7% for heart rate measurements of the Apple Watch [22]. In healthy test participants, Lauterbach et al [23] demonstrated an acceptable heart rate measurement accuracy with a bias <–1 bpm for the Garmin Fenix 5x plus. In patients with pre-existing cardiovascular disease, the Apple Watch Sport showed an MAE of 6.34 bpm compared to a 12-lead ECG, leading Falter et al [24] to conclude clinically acceptable accuracy. Focusing on the use of the Apple Watch 6 in patients with lung diseases or cardiovascular diseases, heart rate measurements showed a bias of –0.11 bpm and achieved a PCC of r=0.98 compared to standard finger pulse oximeters [25]. A further study comparing Apple Watch against pulse oximeters, including 100 pulmonary prediseased patients in a sitting position, demonstrated a concordance of r_c=0.995 in heart rate measurements [26]. Additionally, when comparing the Apple Watch against a telemetry monitor (CARESCAPE Monitor, GE Healthcare) in patients with atrial fibrillation and obstructive sleep apnea, authors concluded acceptable measurement accuracy [27]. On the other hand, wrist-worn devices were considered unsuitable for supraventricular tachycardia detection, if these last for less than 60 seconds [28]. Additionally, the Fitbit tracker, when compared to the clinical gold standard in patients requiring intensive care, exhibited a bias of –4.7 bpm (95% CI –4.91 to –4.44) and a relatively low correlation of r=0.74 [29]. To our best knowledge, there are currently no comparable results from other studies investigating heart rate measurement accuracy based on PPG signals by the Apple Watch 7, the Garmin Fenix 6 Pro, the Withings ScanWatch, and the Fitbit Sense. To date, there is equally poor evidence on the clinical use of further parameters measured by fitness trackers, for example, heart rate variability, blood pressure, oxygen saturation, and cardiac output.

Wearables in Digital Health Care

As part of clinical trials, an increasing number of systems that enable continuous monitoring of patients' vital signs are finding their way into clinical settings. In particular, wearables were used for early diagnostics in clinical studies during the COVID-19 pandemic [30], demonstrating that an infection can be detected by wearables even before a positive nose swab [31]. Techniques to detect certain cardiac arrhythmias with consumer-grade devices are currently being validated [32]. A randomized trial involving older adult patients in this area of application demonstrated that the detection rate of atrial fibrillation is increased by one order of magnitude compared to the standard care group [33]. Moreover, ongoing efforts on developing artificial intelligence models are using data collected from consumer-grade wearables in order to detect and to predict cardiovascular-related diseases [34]. A further meta-analysis focusing on the early detection of sepsis concluded that even mortality is reduced (risk ratio 0.56) by automated alerts when comparing artificial intelligence–based continuous vital sign monitoring systems to standard care [4]. However, wearables provide the possibility of early diagnosis and therefore of initiating timely therapies, but obviously do not alone constitute a therapeutic tool [35]. Furthermore, the compliance of patients using such wearables is of fundamental importance. In this regard, an average wear time of 23.1 hours per day has been reported in patients with dementia, who also demonstrated a high degree of satisfaction according to a survey [36]. Other challenges that need to be resolved in order to implement wearable systems at a large scale concern the financing concepts. Although the devices are significantly more cost efficient than the current standard monitoring, concrete concepts will require further development.

Limitations

There are several limitations to our study. First, even though some cardiac applications of the devices we used are approved by the US Food and Drug Administration, manufacturers generally discourage using them for diagnostic testing in a medical setting. Next, some important technical details—particularly the length of the time interval over which the heart rate is measured by the consumer-grade trackers as well as the delay between measuring and displaying the

result—are not disclosed publicly by the manufacturers. This could result in hidden biases when time matching the measurements of different fitness trackers with each other and with the gold standard reference values. With respect to this concern, we also could not fully address the question of up to which degree dropouts in the measurements of fitness trackers are related to technical problems, problems in the usage, or internal quality control mechanisms of the underlying algorithms.

We collected 3 consecutive measurements per patient during a comparatively short interval. Therefore, conclusions about long-term use are clearly beyond the scope of this trial. Furthermore, our study is underpowered to assess the measurement accuracy of the devices at extreme values of the heart rate because 78.2% of our ECG data can be considered of regular heart rate (60-90 bpm), 11.9% are bradycardic (<60 bpm), and 9.9% are tachycardic (>90 bpm). Similarly, although our results support the hypothesis of higher BMI values impairing the measurement performance, our data ultimately cannot elucidate the effects of obesity to its full impact because our study design did not include patients with a BMI of >40 kg/m^2. Additionally the median of the skin pigmentation in our cohort corresponds to Fitzpatrick scale 2, therefore, no final conclusions can be drawn about the impact of dark skin on the accuracy of the trackers. Since we focused on resting patients in the supine position, no conclusions can be drawn about the measurement accuracy of mobile patients [8]. Therefore, future studies are essential to evaluate wearables in mobile patients.

Conclusions

We summarize that consumer-grade wearables demonstrate promising accuracy for heart rate monitoring in postsurgical patients after moderate to major surgery. The confounders identified in this study did not affect heart rate measurements to a clinically relevant extent.

Acknowledgments

We would like to express our special thanks to the Vogel Foundation Dr. Eckernkamp for funding the project (Special Research Award “100 Jahre Universitätsbund”). This project was supported by the Interdisciplinary Center for Clinical Research (IZKF) at the University of Würzburg (IZKF-Clinician Scientist Program CSP-19). The manufacturers of the devices were not involved in the funding at any time. This publication was supported by the Open Access Publication Fund of the University of Wuerzburg.

Data Availability

Deidentified data can be requested for noncommercial research to qualified academic institutions. A formal request for data access must be sent to the principal investigator (helmer_p@ukw.de), including a standard data use agreement and guarantee that data will be used for research purposes only.

Authors' Contributions

PH, BEW, and MS are responsible for the conceptualization of the work. PH, SH, PR, and MH collected the data. MS and RL performed the statistical analysis. PH, BEW, MS, and RP interpreted the data. The first version of the manuscript was drafted by PH and MS and critically reviewed by SH, PR, RL, BEW, MH, RP, PK, and PM. The design of the graphics was realized by MS. PK and PM supervised the study. PH, PK, and PM managed the funding acquisition. All authors gave substantial contributions to the conception of the work and interpretation of data. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

SH, PR, RL, BEW, MH, RP, and MS declare no conflicts of interest. PH received a research award from Vogel-Foundation and is a member of the Clinician Scientist Programme, Wuerzburg. PM received honoraria for scientific lectures from CSL Behring GmbH, Haemonetics, Werfen GmbH, and ViforPharma GmbH. PK received lecturing fees from TEVA, Sintetica, CSL Behring GmbH, Vifor Pharma GmbH, Pharmacosmos, and Grünenthal and consulted for TEVA and Milestone Scientific Inc. All mentioned funders had no role in the design of the study; collection, analyses, or interpretation of data; writing of the manuscript; or in the decision to publish the results.

Multimedia Appendix 1
Supplementary figures and tables.
[DOCX File , 908 KB - jmir_v24i12e42359_app1.docx]

References

1. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth 2022 Jan 25;10(1):e34384 [FREE Full text] [doi: 10.2196/34384] [Medline: 35076409]
2. Youssef Ali Amer A, Wouters F, Vranken J, de Korte-de Boer D, Smit-Fun V, Duflot P, et al. Vital signs prediction and early warning score calculation based on continuous monitoring of hospitalised patients using wearable technology. Sensors (Basel) 2020;20(22):6593 [FREE Full text] [doi: 10.3390/s20226593] [Medline: 33218084]

3. Downey CL, Chapman S, Randell R, Brown J, Jayne DG. The impact of continuous versus intermittent vital signs monitoring in hospitals: a systematic review and narrative synthesis. Int J Nurs Stud 2018;84:19-27 [FREE Full text] [doi: 10.1016/j.ijnurstu.2018.04.013] [Medline: 29729558]
4. Zhang Z, Chen L, Xu P, Wang Q, Zhang J, Chen K, et al. Effectiveness of automated alerting system compared to usual care for the management of sepsis. NPJ Digit Med 2022;5(1):101 [FREE Full text] [doi: 10.1038/s41746-022-00650-5] [Medline: 35854120]
5. Guk K, Han G, Lim J, Jeong K, Kang T, Lim E, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials (Basel) 2019;9(6):813 [FREE Full text] [doi: 10.3390/nano9060813] [Medline: 31146479]
6. Khanna AK, Hoppe P, Saugel B. Automated continuous noninvasive ward monitoring: future directions and challenges. Crit Care 2019;23(1):194 [FREE Full text] [doi: 10.1186/s13054-019-2485-7] [Medline: 31146792]
7. Pasadyn SR, Soudan M, Gillinov M, Houghtaling P, Phelan D, Gillinov N, et al. Accuracy of commercially available heart rate monitors in athletes: a prospective study. Cardiovasc Diagn Ther 2019;9(4):379-385 [FREE Full text] [doi: 10.21037/cdt.2019.06.05] [Medline: 31555543]
8. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
9. Sañudo B, De Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot study assessing the influence of skin type on the heart rate measurements obtained by photoplethysmography with the Apple Watch. J Med Syst 2019;43(7):195. [doi: 10.1007/s10916-019-1325-2] [Medline: 31119387]
10. Mestrom E, Deneer R, Bonomi AG, Margarito J, Gelissen J, Haakma R, et al. Validation of heart rate extracted from wrist-based photoplethysmography in the perioperative setting: prospective observational study. JMIR Cardio 2021;5(2):e27765 [FREE Full text] [doi: 10.2196/27765] [Medline: 34734834]
11. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019;7(3):e10828. [doi: 10.2196/10828]
12. Patel V, Orchanian-Cheff A, Wu R. Evaluating the validity and utility of wearable technology for continuously monitoring patients in a hospital setting: systematic review. JMIR Mhealth Uhealth 2021;9(8):e17411 [FREE Full text] [doi: 10.2196/17411] [Medline: 34406121]
13. Knight SR, Ng N, Tsanas A, Mclean K, Pagliari C, Harrison EM. Mobile devices and wearable technology for measuring patient outcomes after surgery: a systematic review. NPJ Digit Med 2021;4(1):157 [FREE Full text] [doi: 10.1038/s41746-021-00525-1] [Medline: 34773071]
14. Chan PY, Ryan NP, Chen D, McNeil J, Hopper I. Novel wearable and contactless heart rate, respiratory rate, and oxygen saturation monitoring devices: a systematic review and meta-analysis. Anaesthesia 2022;77(11):1268-1280. [doi: 10.1111/anae.15834] [Medline: 35947876]
15. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
16. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020;3(1):90 [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
17. Collins T, Woolley SI, Oniani S, Pires IM, Garcia NM, Ledger SJ, et al. Version reporting and assessment approaches for new and updated activity and heart rate monitors. Sensors (Basel) 2019;19(7):1705 [FREE Full text] [doi: 10.3390/s19071705] [Medline: 30974755]
18. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
19. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]
20. Navalta JW, Montes J, Bodell NG, Salatto RW, Manning JW, DeBeliso M. Concurrent heart rate validity of wearable technology devices during trail running. PLoS One 2020;15(8):e0238569 [FREE Full text] [doi: 10.1371/journal.pone.0238569] [Medline: 32866216]
21. Association for the Advancement of Medical Instrumentation. Cardiac monitors, heart rate meters, and alarms. In: ANSI/AAMI EC13-2002. New York, NY: American National Standards Institute, Inc; 2002.
22. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022;24(1):e30791. [doi: 10.2196/30791]
23. Lauterbach CJ, Romano PA, Greisler LA, Brindle RA, Ford KR, Kuennen MR. Accuracy and reliability of commercial wrist-worn pulse oximeter during normobaric hypoxia exposure under resting conditions. Res Q Exerc Sport 2021;92(3):549-558. [doi: 10.1080/02701367.2020.1759768] [Medline: 32633688]

24. Falter M, Budts W, Goetschalckx K, Cornelissen V, Buys R. Accuracy of Apple Watch measurements for heart rate and energy expenditure in patients with cardiovascular disease: cross-sectional study. JMIR Mhealth Uhealth 2019;7(3):e11889. [doi: 10.2196/11889]
25. Spaccarotella C, Polimeni A, Mancuso C, Pelaia G, Esposito G, Indolfi C. Assessment of non-invasive measurements of oxygen saturation and heart rate with an apple smartwatch: comparison with a standard pulse oximeter. J Clin Med Res 2022;11(6):1467. [doi: 10.3390/jcm11061467]
26. Pipek LZ, Nascimento RFV, Acencio MMP, Teixeira LR. Comparison of SpO2 and heart rate values on Apple Watch and conventional commercial oximeters devices in patients with lung disease. Sci Rep 2021;11(1):18901. [doi: 10.1038/s41598-021-98453-3]
27. Huynh P, Shan R, Osuji N, Ding J, Isakadze N, Marvel FA, et al. Heart rate measurements in patients with obstructive sleep apnea and atrial fibrillation: prospective pilot study assessing Apple Watch's agreement with telemetry data. JMIR Cardio 2021;5(1):e18050 [FREE Full text] [doi: 10.2196/18050] [Medline: 33555260]
28. Sequeira N, D'Souza D, Angaran P, Aves T, Dorian P. Common wearable devices demonstrate variable accuracy in measuring heart rate during supraventricular tachycardia. Heart Rhythm 2020;17(5 Pt B):854-859. [doi: 10.1016/j.hrthm.2020.02.018]
29. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
30. Ates HC, Yetisen AK, Güder F, Dincer C. Wearable devices for the detection of COVID-19. Nat Electron 2021;4(1):13-14. [doi: 10.1038/s41928-020-00533-1]
31. Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, et al. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study. J Med Internet Res 2021;23(2):e26107 [FREE Full text] [doi: 10.2196/26107]
32. Campo D, Elie V, de Gallard T, Bartet P, Morichau-Beauchant T, Genain N, et al. Validation of an algorithm for atrial fibrillation detection with an analog smartwatch: prospective interventional clinical study. JMIR Form Res 2022;6(11):e37280. [doi: 10.2196/37280]
33. Gladstone DJ, Wachter R, Schmalstieg-Bahr K, Quinn FR, Hummers E, Ivers N, SCREEN-AF Investigators and Coordinators. Screening for atrial fibrillation in the older population: a randomized clinical trial. JAMA Cardiol 2021;6(5):558-567 [FREE Full text] [doi: 10.1001/jamacardio.2021.0038] [Medline: 33625468]
34. Lee S, Chu Y, Ryu J, Park YJ, Yang S, Koh SB. Artificial intelligence for detection of cardiovascular-related diseases from wearable devices: a systematic review and meta-analysis. Yonsei Med J 2022;63(Suppl):S93-S107. [doi: 10.3349/ymj.2022.63.s93]
35. Posthuma LM, Visscher MJ, Hollmann MW, Preckel B. Monitoring of high- and intermediate-risk surgical patients. Anesth Anal 2019;129(4):1185-1190. [doi: 10.1213/ane.0000000000004345]
36. Farina N, Sherlock G, Thomas S, Lowry RG, Banerjee S. Acceptability and feasibility of wearing activity monitors in community‐dwelling older adults with dementia. Int J Geriatr Psychiatry 2019;34(4):617-624. [doi: 10.1002/gps.5064]

Abbreviations

AE: absolute error
ASA: American Society of Anesthesiologists
bpm: beats per minute
CCC: Lin concordance coefficient
CDR: cumulative dropout rate
ECG: electrocardiography
MAE: mean absolute error
MAPE: mean absolute percentage error
PACU: postanesthesia care unit
PCC: Pearson correlation coefficient
PPG: photoplethysmography
RSS: residual sum of squares
TPO: transmissive pulse oximetry

Edited by G Eysenbach; submitted 01.09.22; peer-reviewed by M Kapsetaki, M Gholami; comments to author 07.10.22; revised version received 25.10.22; accepted 15.11.22; published 30.12.22.

Please cite as:
Helmer P, Hottenrott S, Rodemers P, Leppich R, Helwich M, Pryss R, Kranke P, Meybohm P, Winkler BE, Sammeth M
Accuracy and Systematic Biases of Heart Rate Measurements by Consumer-Grade Fitness Trackers in Postoperative Patients: Prospective Clinical Trial
J Med Internet Res 2022;24(12):e42359
URL: https://www.jmir.org/2022/12/e42359
doi:10.2196/42359
PMID:36583938

Original Paper

Characterizing Help-Seeking Searches for Substance Use Treatment From Google Trends and Assessing Their Use for Infoveillance: Longitudinal Descriptive and Validation Statistical Analysis

Thomas Patton[1], PhD; Daniela Abramovitz[2], MS; Derek Johnson[3], PhD; Eric Leas[4], PhD; Alicia Nobles[1], PhD; Theodore Caputi[5], MPH; John Ayers[1], PhD; Steffanie Strathdee[1], PhD; Annick Bórquez[1], PhD

[1]Division of Infectious Diseases and Global Public Health, University of California San Diego, La Jolla, CA, United States

[2]Department of Medicine, University of California San Diego, La Jolla, CA, United States

[3]Medecins Sans Frontieres, Genève, Switzerland

[4]Herbert Wertheim School of Public Health & Human Longevity Science, University of California San Diego, La Jolla, CA, United States

[5]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, United States

Corresponding Author:
Thomas Patton, PhD
Division of Infectious Diseases and Global Public Health
University of California San Diego
9500 Gilman Drive
La Jolla, CA, 92093
United States
Phone: 1 858 534 0830
Email: tepatton@health.ucsd.edu

Abstract

Background: There is no recognized gold standard method for estimating the number of individuals with substance use disorders (SUDs) seeking help within a given geographical area. This presents a challenge to policy makers in the effective deployment of resources for the treatment of SUDs. Internet search queries related to help seeking for SUDs using Google Trends may represent a low-cost, real-time, and data-driven infoveillance tool to address this shortfall in information.

Objective: This paper assesses the feasibility of using search query data related to help seeking for SUDs as an indicator of unmet treatment needs, demand for treatment, and predictor of the health harms related to unmet treatment needs. We explore a continuum of hypotheses to account for different outcomes that might be expected to occur depending on the demand for treatment relative to the system capacity and the timing of help seeking in relation to trajectories of substance use and behavior change.

Methods: We used negative binomial regression models to examine temporal trends in the annual SUD help-seeking internet search queries from Google Trends by US state for cocaine, methamphetamine, opioids, cannabis, and alcohol from 2010 to 2020. To validate the value of these data for surveillance purposes, we then used negative binomial regression models to investigate the relationship between SUD help-seeking searches and state-level outcomes across the continuum of care (including lack of care). We started by looking at associations with self-reported treatment need using data from the National Survey on Drug Use and Health, a national survey of the US general population. Next, we explored associations with treatment admission rates from the Treatment Episode Data Set, a national data system on SUD treatment facilities. Finally, we studied associations with state-level rates of people experiencing and dying from an opioid overdose, using data from the Agency for Healthcare Research and Quality and the CDC WONDER database.

Results: Statistically significant differences in help-seeking searches were observed over time between 2010 and 2020 (based on $P<.05$ for the corresponding Wald tests). We were able to identify outlier states for each drug over time (eg, West Virginia for both opioids and methamphetamine), indicating significantly higher help-seeking behaviors compared to national trends. Results from our validation analyses across different outcomes showed positive, statistically significant associations for the models relating to treatment need for alcohol use, treatment admissions for opioid and methamphetamine use, emergency department visits related to opioid use, and opioid overdose mortality data (based on regression coefficients having $P\leq.05$).

Conclusions: This study demonstrates the clear potential for using internet search queries from Google Trends as an infoveillance tool to predict the demand for substance use treatment spatially and temporally, especially for opioid use disorders.

(J Med Internet Res 2022;24(12):e41527) doi:10.2196/41527

KEYWORDS

internet; search; help-seeking; substance use treatment; surveillance; infoveillance; google trends

Introduction

Understanding help-seeking behavior for substance use treatment is critical for the effective deployment of resources. This presents a challenge to researchers and policy makers because there is no recognized gold standard method for estimating the number of individuals with substance use disorders (SUD) within a given geographical area [1]. A standard approach involves asking a sample of the general population questions about their substance use, either through surveys or in-depth interviews [2]. Unfortunately, these sources are subject to well-known limitations, such as low participation rates, lag time between data collection and published results, and data availability [3]. Additionally, survey scale-up is not always feasible given both costs and concerns about participant burden [4].

Indirect estimation approaches have been used, including capture-recapture [5], multiplier [6], and data triangulation methods [7], but these methods are also subject to limitations, either in the form of impractical data requirements or the potential for bias [7,8]. Finally, efforts have been made to collect data on drug-related harms, such as overdose statistics, although time-lags in their dissemination have meant that these initiatives have struggled to keep pace with the rapidly evolving opioid epidemic in the United States [9]. In view of this methodological backdrop, there has been limited scope to develop real-time surveillance mechanisms to guide policy responses.

A promising development has emerged in the use of internet search queries related to substance use [10]. One of the main benefits of this approach to substance use surveillance is that the data are publicly accessible and can be easily obtained in real time [11]. There is a growing body of research exploring the use of internet search data for the surveillance of substance use trends. A study in 2018 found strong and significant correlations between Google search data for novel psychoactive drugs and annual drug use prevalence, collected in a nationally representative US sample [10]. Two studies explored the relationship between drug-related internet search queries and opioid-related emergency department (ED) visits in the United States, and both demonstrated the predictive potential of internet search data [12,13]. Three further studies found strong associations between drug-related internet search queries and opioid-related overdose deaths at the national, state, and county levels [14-16]. Elsewhere, studies have demonstrated the potential use of opioid-related data from social media platforms, including Twitter and Reddit, to inform surveillance efforts [17-19].

While the previous literature has focused on the use of internet data as a proxy for real-time data on opioid-related health harms, this study provides new insights into the use of internet search data to explore SUD help seeking for a broad range of substances, including cocaine, methamphetamine, opioids, cannabis, and alcohol, and validate these against observed SUD indicators. These substances were chosen because they are the 5 most common types of substance that people are admitted to treatment for in the United States [20]. By validating surveillance of SUD help seeking as a methodological tool, it is our hope that key stakeholders, including local health departments, harm reduction organizations, and researchers, will be better able to proactively respond to need [21]. We first described help-seeking searches for cocaine, methamphetamine, opioids, cannabis, and alcohol at national and state level from 2010 to 2020 in the United States and characterized heterogeneity in these outcomes between states.

We sought to determine the feasibility of using search query data as a low-cost and real-time indicator of unmet treatment need, demand for treatment, and a predictor of the health harms related to unmet treatment needs. The exploratory nature of this study warrants a continuum of hypotheses to account for different outcomes that might be expected to occur depending on the relative demand versus capacity for treatment. If there is sufficient treatment capacity, one would expect to see a strong, positive association between help-seeking searches and treatment admissions. However, given the limited capacity for SUD treatment in the United States, it was important to consider additional hypotheses; if there is excess demand for treatment, we would expect to see a weaker association between help-seeking searches and treatment admissions but a stronger association with unmet treatment need and drug-related health harms. In addition, it is also key to acknowledge that treatment seeking for SUD is a complex process that involves moving, often nonlinearly, through different stages of behavior change [22]. Therefore, considering several outcomes also allows us to reflect the different situations that individuals with SUD, or those around them and trying to help, might be experiencing.

We tested 3 hypotheses, the first of which posits that treatment-seeking searches are positively associated with unmet treatment needs, as searching for help indicates that the person is struggling with their substance use and is considering treatment as an option but has not yet received help (ie, contemplation). Next, we tested the hypothesis that treatment-seeking searches are positively associated with treatment admissions, as searching for help is an indicator that the person is actively seeking to engage (ie, preparation/action) [22,23]. Finally, we tested the hypothesis that treatment-seeking searches are positively associated with nonfatal and fatal overdose, as expressing a treatment need often occurs in the latter stages of SUD, when symptoms are more severe, leading to an increased risk of overdose and death (ie, contemplation,

preparation, or relapse) [24]. We identified relevant variables across different state-level data sources to validate the models for each of these outcomes and determine whether internet searches for substance use help seeking can be used to enhance SUD treatment need surveillance and treatment linkage efforts.

Methods

Extraction of Google Search Query Data

We obtained Google queries in November 2020 originating from the United States that included the terms "quit," "stop," "rehab(s)," "rehabilitation," "treatment(s)," "help," or "detox" in combination with (A) alcohol ("alcohol," "alcoholic," or "alcoholism"), (B) cannabis ("cannabis" or "marijuana"), (C) cocaine ("cocaine"), (D) methamphetamine ("methamphetamine" or "meth"), or (E) opioids ("opioid(s)," "heroin," "fentanyl," "oxycontin," "oxycodone," "codeine," "hydrocodone," "morphine") from January 1, 2010, to November 1, 2020. For example, "Where can I get help for alcoholism" would be included in the alcohol help-seeking search category. These searches were specified without quotation marks, and the data were obtained by selecting the "search terms" option, as opposed to the "topics" option. The search terms for our drugs of interest corresponded to the standard dictionary term for each (eg, methamphetamine), alongside other commonly used terms if relevant (eg, meth) based on the authors' expertise in SUD and others' contributions in this field [12,13,16]. For opioids, we also included names of most frequently used street drugs (ie, heroin, fentanyl) and prescription drugs with their brand name if very commonly used (eg, oxycodone and OxyContin). For alcohol, we also included "alcoholic" and alcoholism," as these are part of the mainstream English lexicon used to describe alcohol use disorders. Despite the extensive range of slang terms used to describe drugs [25], these were not included, given that slang is ever evolving, its linguist survival is often short-lived, and it is typically context specific and limited in use within specific social settings [26]. Given our focus on treatment seeking (ie, a formal context), our broad geographical scale (ie, all US states), and our extended time scale (10 years), we opted to limit our search to the most standard terms to allow for consistency over time and space. The search query data were obtained for each calendar year between 2010 and 2020 from Google Trends using the Google Application Programming Interface (API) Client library in Python [27]. Trends in Google queries were measured in query fractions (QFs), which estimate the number of searches that mention substance-specific keywords, in combination with the help-seeking keywords, in the time frame and geography divided by the total number of searches in the same time frame and geography and expressed as a rate per 1 million searches. This approach facilitates comparability by adjusting for changes in Google usage over time, as well as differences across states and substance types.

Statistical Analysis of Google Search Query Data

Negative binomial regression models were fitted to the QF data to make inferences regarding the significance of temporal changes in help-seeking queries. Negative binomial regression is commonly used to analyze count and rate data exhibiting over-dispersion (ie, variance greater than the mean) [28]. The QF data in this study were found to be overdispersed, as shown in Table 1; therefore, the negative binomial model was chosen to analyze these data. The model specifications included a main fixed effect for year (ie, 2010 through 2020). Random effects were included for intercept terms to account for differences between states at the beginning of the study and for correlations between data points collected in the same states over different years. Moreover, autocorrelated error terms were specified to account for correlations in the data between successive time points. A Wald test was performed to confirm whether the variable "year" was statistically significant for each of the models [29]. We calculated Gini coefficients [30] to quantify the dispersion of help-seeking queries across states for each substance and each year.

Table 1. Descriptive statistics from 2010 to 2020 for annual search query fractions (QFs) by substance type[a].

Statistic	Alcohol	Cannabis	Cocaine	Meth[b]	Opioids	Composite[c]
Mean	27.3	8.4	2.5	4.2	8.0	12.7
Median	25.6	7.4	2.2	3.7	7.2	12.8
Minimum	10.3	2.7	1.0	0.8	2.4	9.1
Maximum	70.7	48.0	9.8	35.0	37.7	16.6
SD	7.2	3.7	1.0	3.3	3.5	1.8
IQR	7.3	3.2	0.7	2.9	2.7	2.8

[a]Query fractions (QFs) refer to queries per every 1 million total Google searches.

[b]Meth: methamphetamine.

[c]Variable estimated by combining QF statistics for opioid, methamphetamine, and cocaine use treatment seeking.

Validation of Google Search Queries as Indicators of Unmet Treatment Needs for Substance Use (Hypothesis 1)

First, an analysis exploring the number of people needing but not receiving treatment at a specialty facility for SUD in the past year was conducted using data from the National Survey on Drug Use and Health (NSDUH) for the years 2016 to 2019. The NSDUH is an annual state-level representative survey of the civilian, noninstitutionalized population aged 12 or older and is publicly accessible from the website of the Substance Abuse and Mental Health Services Administration (SAMHSA)

[31]. To produce state-level estimates for variables collected in this survey (rounded to the nearest thousand), the Research Triangle Institute conducted an analysis of the sample data for each year using survey-weighted hierarchical Bayes methods [32]. The NSDUH separately enquires about needing but not receiving treatment at a specialty facility for alcohol and illicit drug use in the past year. Therefore, a negative binomial model was utilized to regress NSDUH estimates specific to alcohol use on the variables of alcohol QF and year, and a second analysis was conducted exploring illicit drug use, also using a negative binomial regression model. The main fixed effects included in the second model were a composite QF statistic, estimated by combining QF statistics for opioid, methamphetamine, and cocaine use unmet treatment need and the year corresponding to the data points. For both sets of analyses involving NSDUH data, random effects were specified for intercept terms and the natural logarithm of states' population estimates from the US Census Bureau as offset terms [33], which reflect the number of times the event could have potentially occurred. Additionally, interactions between the main fixed effects were assessed to infer if and how the association between alcohol QF and treatment need as well as the association between the composite illicit drug QF and treatment need varied across the years.

Validation of Google Search Queries as Indicators of Treatment Seeking for Substance Use Disorders (Hypothesis 2)

We investigated whether there was a positive association between treatment-seeking searches and the receipt of treatments for SUD. For the latter, data were obtained from the Treatment Episode Data Set: Admissions (TEDS-A) data sets (years 2012 to 2018) on the number of admissions to substance use facilities by primary substance for which treatment was sought [34]. Observations from these data were only selected for admissions involving individual referrals for treatment (ie, excluding mandated treatment visits). Negative binomial regression models were fitted to the admissions data with separate analyses for the different types of substance use (alcohol, cannabis, opioids, cocaine, and methamphetamine). In each model, the year and corresponding help-seeking QF variable (ie, substance-specific) were included as main fixed effects, along with intercepts for the states as random effects and the natural logarithm of states' population estimates from the US Census Bureau as offset terms [33]. Additionally, interactions between the main fixed effects variables (ie, help-seeking QF and year) were assessed to infer if and how the association between treatment-seeking searches and admissions varied across the years.

Validation Of Google Search Queries as Predictors of Health Harms Related to Unmet Treatment Need (Hypothesis 3)

We investigated whether help-seeking searches were positively associated with nonfatal and fatal opioid overdose. Accordingly, data on the rate of ED encounters associated with opioid use per 100,000 people (mostly corresponding to nonfatal overdoses) were obtained from the Agency for Healthcare Quality and Research (AHRQ) [35], and data on the number of opioid-related overdose deaths were obtained from the CDC WONDER database (using the criteria set out in previous research [36]). For these analyses, we regressed state-specific opioid hospitalization rates and mortality count data, respectively, on the variables opioid QF and the year corresponding to the data points, using a negative binomial specification. Once again, random intercepts were included to account for correlations between repeated observations within states. The analysis of fatal opioid overdose data included an offset term corresponding to the log of the state-level population. This approach was not taken for the AHRQ data, as these data were obtained in the form of rates, rather than count data. The interaction terms between the main fixed effects (ie, opioid QF and year) were also assessed.

All analyses were conducted using R software (version 4.1.0.), and the negative binomial models were fitted using the glmmTMB package [28].

Ethical Considerations

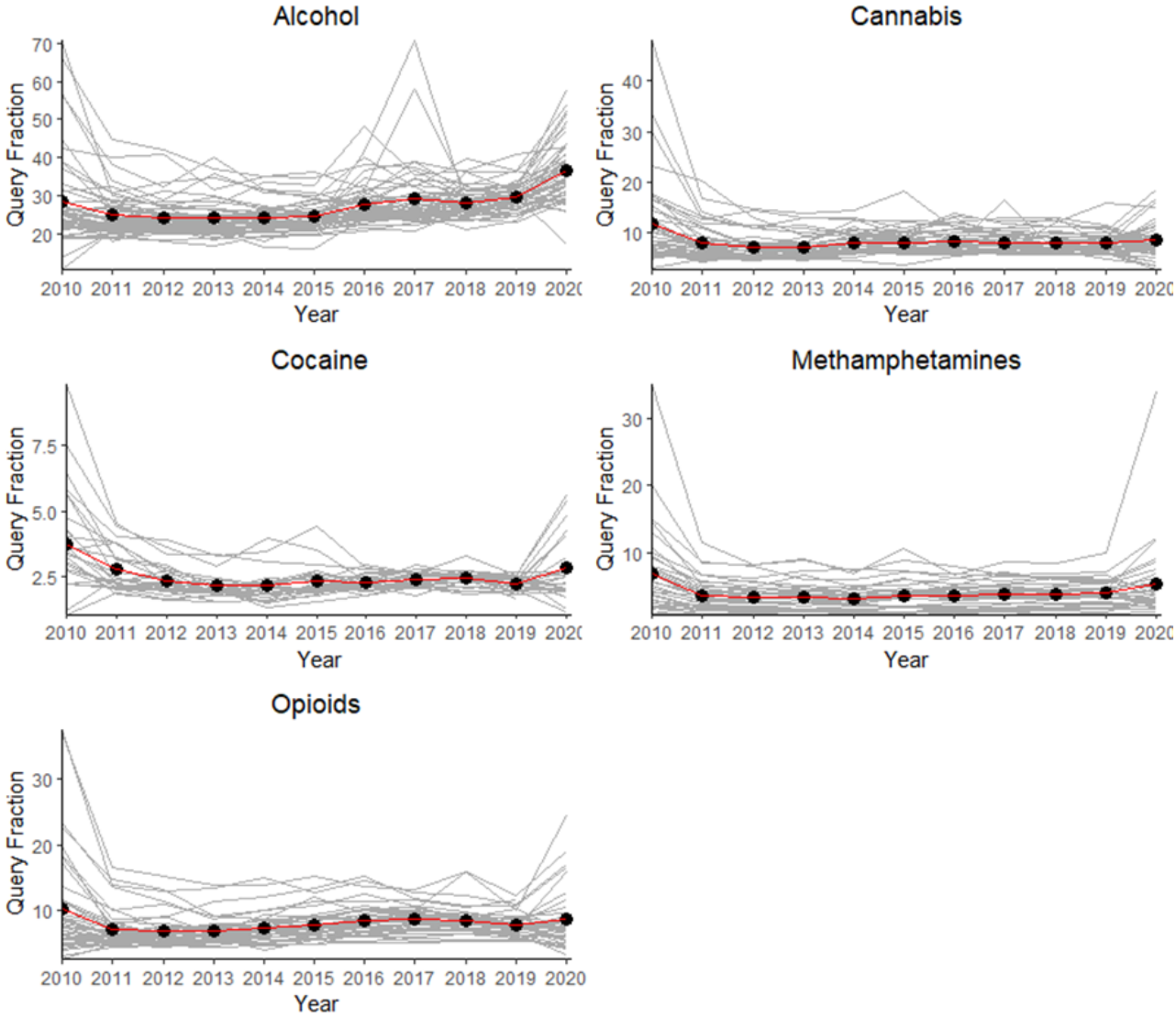
Ethical review was not required because the study relied on public, aggregated, and deidentified data. Given that this study relied on the use of secondary deidentified data (numbers were aggregated to the state level), the Institutional Review Board of the University of California San Diego determined that an ethics review was not required (Project #200332XX).

Results

Descriptive Analysis of Google Search Query Data for SUD Help Seeking

Figure 1 shows that QF values were highest on average in 2010 for all substances, except in the case of alcohol, where it was the second highest year for searching, and the highest levels were observed in 2020. Help-seeking searches were lowest in 2012 in the case of opioids, in 2013 in the case of alcohol and cannabis, and 2014 in the case of cocaine and methamphetamine. Figure 1 also shows the varying levels of completeness in the search query data across the different types of SUD. Missing data points can occur in cases involving very low search volumes [37].

Figure 1. Average help-seeking trends for substance use. Gray lines represent state-specific trends while black dots represent the mean estimates for states* with data across all time points. *Number of states (plus the District of Columbia) with nonmissing query data by substance: Alcohol=51, Opioid=41, Cannabis=44, Methamphetamine=32, Cocaine=25. Number of data points by substance: Alcohol=561, Opioid=461, Cannabis=484, Methamphetamine=382, Cocaine=285.



Statistical Analysis of Time and Geographic Trends in Google Search Query Data for SUD Help Seeking

The negative binomial regression analyses of QF data (results shown in Table S1 in Multimedia Appendix 1) showed that the variable "year" was statistically significant for all substance types based on the resulting Wald tests (P<.001). This indicates that there were important variations in help-seeking searches between the various years. Pairwise comparisons tests for significant differences in the help-seeking search counts over consecutive years were performed by applying Bonferroni corrections to the outputs of the negative binomial regression analyses (results shown in Table S2 in Multimedia Appendix 1). All substances showed significant decreases from 2010 to 2011 (alcohol: 10%, cannabis: 21%, cocaine: 25%, methamphetamine: 43%, opioids: 26%). Aside from this, significant differences across consecutive years were found for alcohol (13% increase from 2015 to 2016 and 21% increase from 2019 to 2020) and methamphetamine (23% increase from 2019 to 2020).

Table 1 provides descriptive statistics for each of the QF variables to facilitate the interpretation of all subsequent regression analyses where these were employed as independent variables. Inequalities in help-seeking searches between states, as measured by Gini coefficients, were highest for methamphetamine across all years (Figure 2). Inequalities were lowest for alcohol for all years except those between 2016 and 2018, when cocaine was the lowest. A consistent trend observed across all substances was that inequalities were highest in 2010 and then reduced over time before sharply increasing again in 2020. These results can also be further understood by looking at the box and whisker plots, which show the spread of data points across states by substance type and year (see Figure 3 for opioid use and Figure S1 in Multimedia Appendix 1 for other substances).

Figure 2. Gini coefficient estimates from query fractions (QF) variables across substances and years.

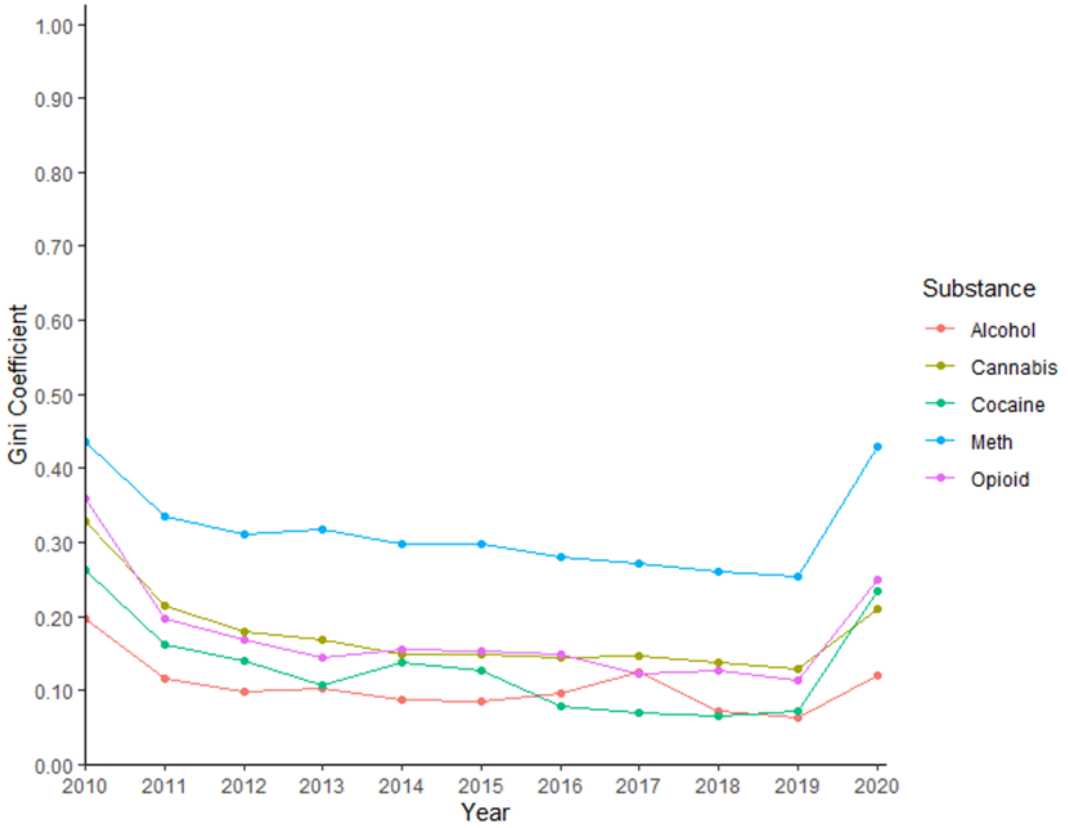
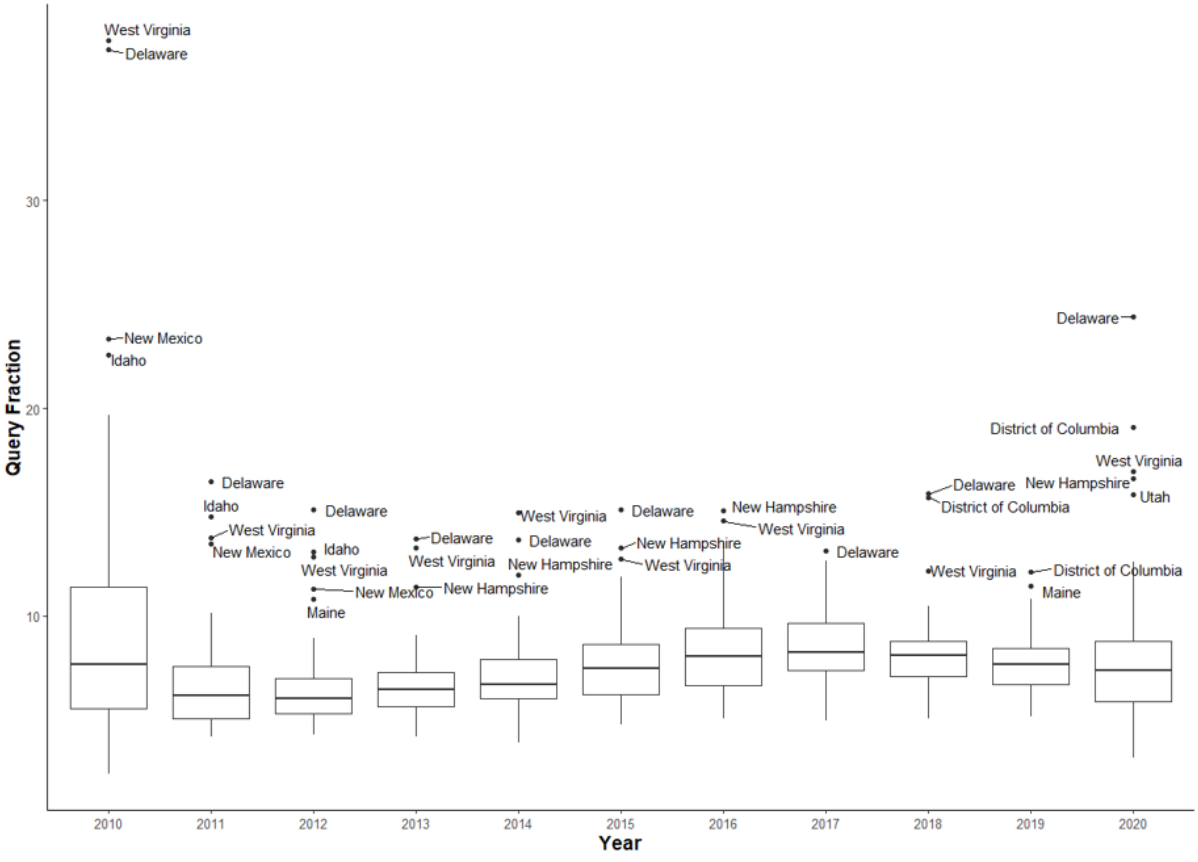


Figure 3. Box and whisker plot of help-seeking searches for opioid use.



Validation of Google Search Queries for SUD Help Seeking as Indicators of Unmet Treatment Need for SUD (Hypothesis 1)

The analysis of NSDUH data showed a statistically significant (*P*=.004) association between QF and the number of people needing but not receiving treatment for alcohol use at a specialty facility (rate ratio changes are shown in Table 2, and regression outputs can be found in Tables S3 and S4 in Multimedia Appendix 1). The estimates were not significantly different from 0 in the case of illicit drug use (*P*=.26). The coefficient estimates for alcohol use and illicit drug use confirmed our expectation of a positive association between the QF variables and the number of people needing but not receiving treatment for their substance use. After adjusting for variations across years, both the analyses for alcohol use and illicit drug use showed that a 1-unit increase (ie, 1 additional search per million searches) in the composite QF variable approximately corresponded to a 1% increase in the expected rate of people needing but not receiving treatment for alcohol use and illicit drug use, respectively. Neither analysis showed statistically significant interactions between QF variables and the variable "year."

Predictions were made for the model analyzing the number of individuals needing but not receiving treatment for alcohol use on account of the significant QF finding. No evaluation of predictive performance was conducted for the illicit drug use model because its association with the QF variable was not statistically significant. Comparisons between the predicted and observed rates of people needing but not receiving treatment for alcohol use are presented in Figure S2 in Multimedia Appendix 1. The predictive performance was also quantified by calculating root mean squared errors (RMSE), comparing observed and predicted rates. The mean RMSE was 697, which is low when compared to the mean rate of 5284 per 100,000 people needing but not receiving treatment for alcohol use. The resulting scatter index of 13%, which is calculated by dividing the mean RMSE by the mean rate and then multiplied by 100, suggests a reasonable predictive performance based on previously used benchmarks [38]. Predictive performance was also examined over time and across states/territories (Tables S5-S6 in Multimedia Appendix 1). It was shown to be best in 2019 for the states of Idaho, Virginia, Michigan, New York, and Kansas and worse in 2018 for the District of Columbia, Colorado, Oregon, Montana, Vermont, compared to other years and states, respectively.

Table 2. Estimates of the rate ratio change in number of people needing but not receiving treatment (NSDUH) associated with a one unit increase in the query fractions (QFs) variable.

Variable	Estimate	95% CI	*P* value
Illicit drug use	1.01	0.99-1.03	.26
Alcohol use	1.01	1.00-1.01	.004

Validation of Google Search Queries for SUD Help Seeking as Indicators of Treatment Seeking for SUD (Hypothesis 2)

The analysis of TEDS-A data showed that the association between help-seeking searches and the receipt of treatments for SUD varied by substance type (rate ratios are shown in Table 3, and regression outputs can be found in Tables S7 to S11 in Multimedia Appendix 1). Statistically significant and positive associations between these variables were found for methamphetamine use (*P*<.001). Interactions between the methamphetamine QF and the year variable were nonsignificant and thus ruled out (*P*=.88 for Type III Wald test). The model outputs showed that a 1-unit increase in the methamphetamine QF variable approximately corresponded to a 26% increase in the expected rate of treatment episodes. Statistically significant and positive associations were also found for opioids (*P*<.001). Interactions between the opioid QF and the year variable were nonsignificant and thus ruled out (*P*=.26 for Type III Wald test). The outputs from the model showed that a 1-unit increase in the opioid QF variable approximately corresponded to a 12% increase in the expected rate of treatment episodes.

In the case of cannabis, the association was also positive and slightly above a 5% statistical significance criterion (*P*=.07). Although the data did not allow strong inferences to be drawn from the analysis of cannabis data, the outputs from the model showed that a 1-unit increase in the cannabis QF variable approximately corresponded to a 3% increase in the expected rate of treatment episodes. Findings for the analyses of alcohol and cocaine use showed both nonsignificant association between treatment-seeking searches and the receipt of treatments (*P*=.92 for alcohol use and *P*=.22 for cocaine use). Neither model exhibited significant interactions between the QF and the year variable (*P*=.23 for alcohol use and *P*=.88 for cocaine use for Type III Wald test for the models of treatment).

Table 3. Estimates of the rate ratio change in number of individual treatment referrals associated with a 1 unit increase in the query fractions (QFs) variable.

Variable	Estimate	95% CI	*P* value
Alcohol QF[a]	1.00	0.99-1.01	.92
Cannabis QF	1.03	1.00-1.07	.07
Cocaine QF	0.90	0.76-1.07	.22
Meth QF	1.26	1.17-1.36	<.001
Opioid QF	1.12	1.07-1.17	<.001

[a]QF: query fraction.

Predictions were made for the models analyzing admissions to treatment for methamphetamine and opioid use. Comparisons between the predicted and observed rates of admission to treatment are presented in Figures S3-S4 in Multimedia Appendix 1. The predictive performance of the models for opioid and methamphetamine use was also quantified by calculating RMSE. The mean RMSE for methamphetamine was 11.7, which indicates a poor predictive performance, given that the mean admission rate was 15.2 per 100,000 people. The predictive performance was also shown to be weak for opioids based on comparisons between the mean RMSE (77.9) and the mean admission rate (102.4 per 100,000 people). Predictive performance was also examined over time and across states (Tables S12 and S13 in Multimedia Appendix 1). For both substances, predictive performance was found to be best in 2011 compared to other years for both substances, and in the states of Indiana, Texas, New York, and Michigan for methamphetamine use and in Illinois, Ohio, Missouri, Utah for opioid use. It was the worst in 2018 for both substances, compared to other years, and generally worse among states with higher admission rates.

Validation Of Google Search Queries for SUD Help Seeking as Predictors of Health Harms Related to Unmet Treatment Need (Hypothesis 3)

The analysis investigating the relationship between treatment-seeking searches for opioid use and opioid-related emergency department visits using AHRQ data showed a positive and statistically significant association (*P*<.001, see Tables S14-S15 in Multimedia Appendix 1). However, statistically significant interactions between the opioid QF and the variable year were identified, indicating that the relationship was not stable over time (Type III Wald test *P*=.005). An evaluation of the simple main effects of the opioid QF by year showed a decreasing trend over time (Table S14 in Multimedia Appendix 1). In 2011, a 1-unit increase in the opioid QF variables was associated with a 6% increase in the expected rate of opioid-related emergency department visits, but by 2018, there was a nonsignificant association between these variables. No evaluation of predictive performance was conducted for this model, as the association with the QF variable was found to vary over time.

The analysis investigating the relationship between treatment-seeking searches for opioid use and opioid overdose mortality counts using CDC WONDER data showed a positive and statistically significant association (*P*<.001, see Table 4 and Table S16 in Multimedia Appendix 1). Interactions between the opioid QF and the year variable were nonsignificant and thus ruled out (Type III Wald test *P*=.11). The outputs from the model showed that a 1-unit increase in the opioid QF variable corresponded to a 11% increase in the expected overdose mortality count (Table 4). The predictive performance for the model was determined by estimating the RMSE. The relative difference between the mean RMSE (4.3) and the mean admission rate per 100,000 people (12.2) indicated a better predictive performance, on average, when compared to the models predicting treatment admission rates. Figure S5 in Multimedia Appendix 1 illustrates the differences between predicted and observed mortality rates across states. Predictive performance was best in 2013 compared to other years (Table S17 in Multimedia Appendix 1) and in the states of New York, Florida, Virginia, and Wisconsin compared to other states (Table S18 in Multimedia Appendix 1). It was worst in 2017 and in West Virginia, Ohio, Idaho, and Maryland, where opioid overdose mortality was very high (with the exception of Idaho).

Table 4. Estimates of the rate ratio change in number of overdose deaths associated with a 1-unit increase in the opioid query fractions (QFs) variable.

Variable	Estimates	95% CI	*P* value
Opioid QF[a]	1.11	1.09-1.14	<.001

[a]QF: query fraction.

Discussion

Principal Findings

To our knowledge, this is the first study to retrospectively describe spatial and temporal changes in substance use searches in the United States and rigorously investigate their association with outcomes along the continuum of care (and absence of care) for SUD. In the future, monitoring of Google search queries with validated metrics may allow the prospective identification of variations by substance and state indicating specific SUD treatment information and linkage needs in the population, providing useful near real-time insights to public health organizations developing and delivering campaigns for

SUD treatment. Key stakeholders (local health departments, harm reduction organizations, etc) could then better allocate resources to target SUD treatment needs (eg, a digital intervention in real time) for each substance in specific states. For instance, between 2010 and 2020, West Virginia (methamphetamine and opioids), New Mexico (methamphetamine), Delaware (opioids), and Connecticut (cocaine) were repeatedly found to exhibit high levels of demand for information on SUD treatments that were potentially unmet.

Importantly, the positive and significant associations we identified between help-seeking searches for opioid and methamphetamine use and admissions to substance use treatment facilities suggests that, at least for these 2 substances, internet search data represents a valuable resource to assess treatment seeking. Interpreting the magnitude of these associations should be considered in the context of the baseline rate of treatment admissions and the overall population size for a given state. For instance, the implications of a 1-unit increase in the rate of help-seeking searches for methamphetamine use in California differs vastly from that in Virginia. The average rate of treatment admissions per 100,000 across all years was 37.43 for California and 1.12 for Virginia. Given that a 1-unit increase in the rate of help-seeking searches is associated with a 26% increase in treatment admissions, this corresponds to 9.73 additional admissions per 100,000 for the average rate in California and 0.29 additional admissions per 100,000 for the average rate in Virginia. In absolute terms, this equates to over 3800 additional admissions for California and only 5 additional admissions for Virginia.

Further analyses showed significant associations between help-seeking searches for opioid use and data on health harms related to unmet treatment need. These findings have implications for both surveillance and treatment, as they demonstrate the clear potential of search query monitoring to fill existing gaps and indicate that the internet likely represents a strategic platform to link people in need of treatment to services. This is especially important given that there are well-documented challenges in estimating the prevalence and incidence of SUD [5,8]. As such, leveraging internet search platforms could make health agencies more responsive to both information and treatment referral needs. This potential can be realized by developing a surveillance platform for real-time monitoring and linkage to services that can allow users to rapidly evaluate fluctuating patterns in SUD help seeking and implement strategic outreach. To realize this potential, search data need to be measured in terms of QFs to ensure that data points are comparable over time and across states. This approach was achieved in this study by extracting search data using the Google API Client library. It is important to highlight that this is not achieved when data are extracted directly through the Google Trends website but rather when data are normalized according to the selected time frame and geographical region [37].

Limitations

Our study is not without limitations. Several states were missing search query data for SUD help-seeking behavior because Google Trends will only report search queries if they are above a minimum threshold. There was variation in the predictive performance of our models over time and across states. In particular, performance was lowest in states where rates of treatment admissions or overdose mortality rates were very high, which is expected when using RMSE as the performance indicator since it penalizes large errors. Using search data may be subject to selection bias, as not all people access the internet equally. Although some queries may reflect general curiosity rather than help seeking, it is well known that internet search trends mirror many health-related behaviors [39], and in the specific case of SUD, that of family members, partners, and friends trying to help their loved one [39]. Another potential confounding factor is the fact that the Google search algorithm is nonstatic. Search patterns change over time due to the thousands of decisions being made by Google's programmers as the company strives to test and improve its search algorithm [40]. This could lead to temporal changes in the likelihood of individuals successfully finding treatment following an online search. As such, this phenomenon could distort the association between search trends and treatment admissions.

While our approach may overcome many of the ongoing limitations in substance use surveillance (ie, a lack of timely, substance specific, and publicly available data), the finest granularity of aggregate Google search data is limited to designated marketing areas [41], so it does not necessarily align with the jurisdictional level of public health departments. The approach taken in this paper also assumes that search queries are made using standard terms for SUD in the context of treatment seeking, which disregards instances where people might use slang terms. It is also possible that the predictive value of specific terms varies between states and over time. However, given the nonpunitive nature of online help seeking for SUD (as compared with that of purchasing or selling drugs), we expect this to be limited. It is important to recognize the potential limitations of using data on the number of treatment-seeking visits from TEDS-A. Given that these data are collected from facilities receiving public funding, the findings from analyses using this data potentially misrepresent associations for states with greater reliance on private funding or nonspecialty settings such as office-based outpatient treatment. Other potential confounding factors include geographical and temporal variations in the number of help-seeking queries in other languages, the proportion of queries coming from surrogate seekers [42], and the use of alternative search engines. In particular, including searches using Spanish terms would have a heterogeneous impact across states, and the relationship between searches and health outcomes might be different depending on the policies and interventions in place to facilitate healthcare access among non-English speakers and those who are undocumented [43,44]. This warrants a separate study focusing on Spanish language terms and SUD-related health outcomes among Hispanic individuals.

Importantly, the strength and significance of associations between searches and outcomes along the continuum of care varied depending on the substance and outcome, as well as between states and through time. This is expected, given that there have been heterogeneities in drug policies over time and across States. Between 2010 and 2020, cannabis was made legal

in 19 states for medical use, in 8 states for recreational use, and in 3 states for medical use first and then later for recreational use [45]. Given that the impacts of legalization on the social acceptability of treatment-seeking behaviors are still poorly understood [46], it is difficult to surmise whether changes in the legal status of cannabis across states and over the duration of the study may have had a distorting impact on the results. Another key policy area is the state adoption of naloxone access laws (NAL), which increased rapidly from 2013 onward [47]. By 2020, all 50 states and the District of Columbia had some form of NAL in place, although the laws varied significantly across states [48]. Despite the proven clinical benefits of naloxone for the reversal of opioid overdoses, its population-level impact depends on the effectiveness of distribution programs alongside multiple contextual factors [49,50]. For this reason, the impact NALs may have had on our results is unclear.

While the goal of this study was to validate the use of help-seeking queries as a surveillance tool across states, our findings call for further investigation within states to contextualize and interpret the results. The inclusion of additional covariates could potentially help to improve the predictive performance of the models developed in this paper and elucidate factors that determine variations in outcomes across years. A key challenge in this regard was the limited sample size, in that there was insufficient statistical power to include additional predictors. One potential remedy to this problem would be to obtain data with more granularity in terms of the time intervals between observations (eg, monthly data) or the geographical level under investigation (eg, county-level data) to increase the number of observations. Finally, while this study retrospectively analyzes SUD help-seeking internet search data to validate their value for surveillance and linkage to treatment, real-time analysis would be the most useful for informing public health agencies, as indicated by some examples investigating mental health–related outcomes during the COVID-19 pandemic [51,52].

Conclusions

This study examined temporal and spatial trends in the annual fractions of substance use help-seeking internet search queries by US state for cocaine, methamphetamine, opioids, cannabis, and alcohol. Our investigations showed positive, statistically significant associations for the models relating to treatment need (but not receiving treatment) for alcohol use, treatment admissions for opioid and methamphetamine use, and overdose mortality data. In the wake of current substance use trends, it is critical that public health professionals learn from and respond to the millions of individuals searching for help online. The field should invest in and prioritize automated surveillance, including extensions of our approach, to understand evolving public health needs.

Acknowledgments

TP and DJ acknowledge support from the National Institute on Drug Abuse (NIDA) through a T32 training grant (T32 DA023356). TP and AB acknowledge support from NIDA through the Avenir grant (DP2DA049295). SS and JA acknowledge support from a California HIV/AIDS research program grant.

Conflicts of Interest

TC reports an equity interest in Data Science Solutions, a research consulting firm, outside of this work. No other authors have conflicts to declare.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 1742 KB - jmir_v24i12e41527_app1.docx]

References

1. Degenhardt L, Bucello C, Calabria B, Nelson P, Roberts A, Hall W, GBD Illicit Drug Use Writing Group, et al. What data are available on the extent of illicit drug use and dependence globally? Results of four systematic reviews. Drug Alcohol Depend 2011 Sep 01;117(2-3):85-101. [doi: 10.1016/j.drugalcdep.2010.11.032] [Medline: 21377813]
2. Key substance use and mental health indicators in the United States: Results from the 2020 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2021. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf [accessed 2022-11-13]
3. Goyder J. Surveys on surveys: Limitations and potentialities. POQ 1986;50(1):27-41. [doi: 10.1086/268957]
4. Hamilton J, Breen N, Klabunde C, Moser R, Leyva B, Breslau E, et al. Opportunities and challenges for the use of large-scale surveys in public health research: a comparison of the assessment of cancer screening behaviors. Cancer Epidem Biomar 2015;24(1):3-14. [doi: 10.1158/1055-9965.epi-14-0568]
5. Barocas JA, White LF, Wang J, Walley AY, LaRochelle MR, Bernson D, et al. Estimated prevalence of opioid use disorder in Massachusetts, 2011–2015: a capture–recapture analysis. Am J Public Health 2018 Dec;108(12):1675-1681. [doi: 10.2105/ajph.2018.304673]

6. Mallow PJ, Sathe N, Topmiller M, Chubinski J, Carr D, Christopher R. Estimating the Prevalence of Opioid use Disorder in the Cincinnati Region using Probabilistic Multiplier Methods and Model Averaging. J Health Econ Outcomes Res 2019 Apr 3;6(2):61-69 [FREE Full text] [doi: 10.36469/9729] [Medline: 32685580]
7. Edwards J, Georgiades K. Reading between the lines: a pursuit of estimating the population prevalence of mental illness using multiple data sources. Can J Psychiatry 2022 Feb;67(2):101-103 [FREE Full text] [doi: 10.1177/07067437211016255] [Medline: 33969716]
8. Jones HE, Harris RJ, Downing BC, Pierce M, Millar T, Ades AE, et al. Estimating the prevalence of problem drug use from drug-related mortality data. Addiction 2020 Dec 09;115(12):2393-2404. [doi: 10.1111/add.15111] [Medline: 32392631]
9. Blanco C, Wall MM, Olfson M. Data needs and models for the opioid epidemic. Mol Psychiatry 2022 Feb 29;27(2):787-792 [FREE Full text] [doi: 10.1038/s41380-021-01356-y] [Medline: 34716409]
10. Perdue RT, Hawdon J, Thames KM. Can big data predict the rise of novel drug abuse? J Drug Issues 2018 May 28;48(4):508-518. [doi: 10.1177/0022042618772294]
11. Trevisan F. Search engines: From social science objects to academic inquiry tools. FM 2014 Oct 23. [doi: 10.5210/fm.v19i11.5237]
12. Young SD, Zheng K, Chu LF, Humphreys K. Internet searches for opioids predict future emergency department heroin admissions. Drug Alcohol Depend 2018 Sep 01;190:166-169 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.05.009] [Medline: 30036853]
13. Young SD, Zhang Q, Zhou J, Pacula RL. Internet search and medicaid prescription drug data as predictors of opioid emergency department visits. NPJ Digit Med 2021 Feb 11;4(1):21 [FREE Full text] [doi: 10.1038/s41746-021-00392-w] [Medline: 33574500]
14. Sumner SA, Bowen D, Holland K, Zwald ML, Vivolo-Kantor A, Guy GP, et al. Estimating weekly national opioid overdose deaths in near real time using multiple proxy data sources. JAMA Netw Open 2022 Jul 21;5(7):e2223033 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.23033] [Medline: 35862045]
15. Arendt F. The opioid-overdose crisis and fentanyl: the role of online information seeking via internet search engines. Health Commun 2021 Sep 14;36(10):1148-1154. [doi: 10.1080/10410236.2020.1748820] [Medline: 32285691]
16. Campo DS, Gussler JW, Sue A, Skums P, Khudyakov Y. Accurate spatiotemporal mapping of drug overdose deaths by machine learning of drug-related web-searches. PLoS One 2020 Dec 7;15(12):e0243622 [FREE Full text] [doi: 10.1371/journal.pone.0243622] [Medline: 33284864]
17. Flores L, Young SD. Regional variation in discussion of opioids on social media. J Addict Dis 2021 Feb 11;39(3):316-321 [FREE Full text] [doi: 10.1080/10550887.2021.1874804] [Medline: 33570000]
18. Bowen DA, O'Donnell J, Sumner SA. Increases in online posts about synthetic opioids preceding increases in synthetic opioid death rates: a retrospective observational study. J Gen Intern Med 2019 Dec 29;34(12):2702-2704 [FREE Full text] [doi: 10.1007/s11606-019-05255-5] [Medline: 31468343]
19. Anwar M, Khoury D, Aldridge AP, Parker SJ, Conway KP. Using Twitter to surveil the opioid epidemic in North Carolina: an exploratory study. JMIR Public Health Surveill 2020 Jun 24;6(2):e17574 [FREE Full text] [doi: 10.2196/17574] [Medline: 32469322]
20. Admissions to and discharges from publicly funded substance use treatment. Substance Abuse and Mental Health Services Administration. Rockville, MD; 2021. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35314/2019_TEDS_Proof.pdf [accessed 2022-11-15]
21. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? JAMA 2014 Apr 09;311(14):1399-1400 [FREE Full text] [doi: 10.1001/jama.2014.1505] [Medline: 24577162]
22. 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]
23. Peteet JR, Brenner S, Curtiss D, Ferrigno M, Kauffman J. A stage of change approach to addiction in the medical setting. Gen Hosp Psychiatry 1998 Sep;20(5):267-273. [doi: 10.1016/s0163-8343(98)00033-4]
24. Willoughby FW, Edens JF. Construct validity and predictive utility of the stages of change scale for alcoholics. J Subst Abuse 1996 Jan;8(3):275-291. [doi: 10.1016/s0899-3289(96)90152-2]
25. Slang terms and code words: A reference for law enforcement personnel. Drug Enforcement Administration. 2018. URL: https://www.dea.gov/sites/default/files/2018-07/DIR-022-18.pdf [accessed 2022-11-14]
26. Coleman J. The Life of Slang. Oxford, UK: OUP; 2012.
27. Google API Python client library. GitHub. URL: https://github.com/googleapis/google-api-python-client [accessed 2022-11-14]
28. Brooks M, Kristensen K, Benthem K, Magnusson A, Berg C, Nielsen A, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J 2017;9(2):378. [doi: 10.32614/rj-2017-066]
29. Bolker B. Linear and generalized linear mixed models. In: Fox G, Negrete-Yankelevich S, Sosa V, editors. Ecological Statistics: Contemporary Theory and Application. Oxford, UK: Oxford Academic; 2015:309-333.
30. Gini C. Concentration and dependency ratios. Rivista di Politica Economica 1997;87:769-789.
31. National drug use and health surveys. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/data/ [accessed 2022-11-14]

32. Wright D. State estimates of substance use from the 2001 National Household Survey on Drug Abuse: Volume II. Individual state tables and technical appendices. Substance Abuse and Mental Health Services Administration. Rockville, MD; 2003. URL: https://webharvest.gov/peth04/20041028054530/http://www.oas.samhsa.gov/nhsda/2k1State/PDF/2k1SAEv2.pdf [accessed 2022-11-15]
33. American community survey. US Census Bureau. URL: https://data.census.gov/cedsci/ [accessed 2022-11-14]
34. Treatment episode data sets (TEDS). Substance Abuse and Mental Health Services Administration. Rockville, MD URL: https://www.datafiles.samhsa.gov/data-sources [accessed 2022-11-14]
35. Healthcare cost and utilization project (HCUP). Agency for Healthcare Research and Quality. Rockville, MD URL: https://www.hcup-us.ahrq.gov/faststats/OpioidUseServlet [accessed 2022-11-14]
36. Haffajee RL, Lin LA, Bohnert ASB, Goldstick JE. Characteristics of US counties with high opioid overdose mortality and low capacity to deliver medications for opioid use disorder. JAMA Netw Open 2019 Jun 05;2(6):e196373 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.6373] [Medline: 31251376]
37. Mavragani A, Ochoa G. Google trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
38. Oyeleye M, Chen T, Titarenko S, Antoniou G. A predictive analysis of heart rates using machine learning techniques. Int J Environ Res Public Health 2022 Feb 19;19(4):2417 [FREE Full text] [doi: 10.3390/ijerph19042417] [Medline: 35206603]
39. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with Web search. Proc Natl Acad Sci U S A 2010 Oct 12;107(41):17486-17490 [FREE Full text] [doi: 10.1073/pnas.1005962107] [Medline: 20876140]
40. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506]
41. DMA regions. The Nielsen Company. URL: https://www.nielsen.com/us/en/contact-us/intl-campaigns/dma-maps/ [accessed 2022-11-14]
42. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". J Cancer Educ 2015 Mar 4;30(1):12-19 [FREE Full text] [doi: 10.1007/s13187-014-0701-3] [Medline: 24989816]
43. Hammig B, Henry J, Davis D. Disparities in health care coverage among US born and Mexican/Central American born labor workers in the US. J Immigr Minor Health 2019 Feb 31;21(1):66-72. [doi: 10.1007/s10903-018-0697-6] [Medline: 29388095]
44. Chang CD. Social determinants of health and health disparities among immigrants and their children. Curr Probl Pediatr Adolesc Health Care 2019 Jan;49(1):23-30. [doi: 10.1016/j.cppeds.2018.11.009] [Medline: 30595524]
45. Marijuana Policy Project. URL: https://www.mpp.org/ [accessed 2022-11-15]
46. Mennis J, Stahler GJ. Adolescent treatment admissions for marijuana following recreational legalization in Colorado and Washington. Drug Alcohol Depend 2020 May 01;210:107960. [doi: 10.1016/j.drugalcdep.2020.107960] [Medline: 32222560]
47. Bohler RM, Hodgkin D, Kreiner PW, Green TC. Predictors of US states' adoption of naloxone access laws, 2001-2017. Drug Alcohol Depend 2021 Aug 01;225:108772 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108772] [Medline: 34052687]
48. Legislative Analysis and Public Policy Association. Naloxone access: summary of state laws. Legislative Analysis. URL: https://legislativeanalysis.org/naloxone-summary-of-state-laws/ [accessed 2022-11-15]
49. Abouk R, Pacula RL, Powell D. Association between state laws facilitating pharmacy distribution of naloxone and risk of fatal overdose. JAMA Intern Med 2019 Jun 01;179(6):805-811 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0272] [Medline: 31058922]
50. Duska M, Rhoads JM, Saunders EC, Onega T. State naloxone co-prescribing laws show mixed effects on overdose mortality rates. Drug Sci Policy Law 2022 Jul 25;8:205032452211125. [doi: 10.1177/20503245221112575]
51. Ayers JW, Leas EC, Johnson DC, Poliak A, Althouse BM, Dredze M, et al. Internet searches for acute anxiety during the early stages of the COVID-19 pandemic. JAMA Intern Med 2020 Dec 01;180(12):1706-1707 [FREE Full text] [doi: 10.1001/jamainternmed.2020.3305] [Medline: 32832984]
52. Silverio-Murillo A, Hoehn-Velasco L, Rodriguez Tirado A, Balmori de la Miyar JR. COVID-19 blues: Lockdowns and mental health-related google searches in Latin America. Soc Sci Med 2021 Jul;281:114040 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114040] [Medline: 34144481]

Abbreviations

AHRQ: Agency for Healthcare Quality and Research
API: application programming interface
CRF: case report form
ED: emergency department
NAL: naloxone access law
NSDUH: National Survey on Drug Use and Health
QF: query fraction

RMSE: root mean squared error
SAMHSA: Substance Abuse and Mental Health Services Administration
SUD: substance use disorder
TEDS-A: Treatment Episode Data Set: Admissions

Edited by G Eysenbach; submitted 29.07.22; peer-reviewed by Z Zhang, A Zheluk; comments to author 11.10.22; revised version received 01.11.22; accepted 04.11.22; published 01.12.22.

Please cite as:
Patton T, Abramovitz D, Johnson D, Leas E, Nobles A, Caputi T, Ayers J, Strathdee S, Bórquez A
Characterizing Help-Seeking Searches for Substance Use Treatment From Google Trends and Assessing Their Use for Infoveillance: Longitudinal Descriptive and Validation Statistical Analysis
J Med Internet Res 2022;24(12):e41527
URL: https://www.jmir.org/2022/12/e41527
doi:10.2196/41527
PMID:36454620

Original Paper

Topics and Sentiment Surrounding Vaping on Twitter and Reddit During the 2019 e-Cigarette and Vaping Use–Associated Lung Injury Outbreak: Comparative Study

Dezhi Wu[1], PhD; Erin Kasson[2], MSc; Avineet Kumar Singh[1], MSc; Yang Ren[1], MSc; Nina Kaiser[2], BA; Ming Huang[3], PhD; Patricia A Cavazos-Rehg[2], PhD

[1]Department of Integrated Information Technology, University of South Carolina, Columbia, SC, United States

[2]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, United States

[3]Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Ming Huang, PhD
Department of Artificial Intelligence and Informatics
Mayo Clinic
200 1st St SW
Rochester, MN, 55902
United States
Phone: 1 507 538 3287
Email: Huang.Ming@mayo.edu

Abstract

Background: Vaping or e-cigarette use has become dramatically more popular in the United States in recent years. e-Cigarette and vaping use–associated lung injury (EVALI) cases caused an increase in hospitalizations and deaths in 2019, and many instances were later linked to unregulated products. Previous literature has leveraged social media data for surveillance of health topics. Individuals are willing to share mental health experiences and other personal stories on social media platforms where they feel a sense of community, reduced stigma, and empowerment.

Objective: This study aimed to compare vaping-related content on 2 popular social media platforms (ie, Twitter and Reddit) to explore the context surrounding vaping during the 2019 EVALI outbreak and to support the feasibility of using data from both social platforms to develop in-depth and intelligent vaping detection models on social media.

Methods: Data were extracted from both Twitter (316,620 tweets) and Reddit (17,320 posts) from July 2019 to September 2019 at the peak of the EVALI crisis. High-throughput computational analyses (sentiment analysis and topic analysis) were conducted. In addition, in-depth manual content analyses were performed and compared with computational analyses of content on both platforms (577 tweets and 613 posts).

Results: Vaping-related posts and unique users on Twitter and Reddit increased from July 2019 to September 2019, with the average post per user increasing from 1.68 to 1.81 on Twitter and 1.19 to 1.21 on Reddit. Computational analyses found the number of positive sentiment posts to be higher on Reddit (P<.001, 95% CI 0.4305-0.4475) and the number of negative posts to be higher on Twitter (P<.001, 95% CI –0.4289 to −0.4111). These results were consistent with the clinical content analyses results indicating that negative sentiment posts were higher on Twitter (273/577, 47.3%) than Reddit (184/613, 30%). Furthermore, topics prevalent on both platforms by keywords and based on manual post reviews included mentions of youth, marketing or regulation, marijuana, and interest in quitting.

Conclusions: Post content and trending topics overlapped on both Twitter and Reddit during the EVALI period in 2019. However, crucial differences in user type and content keywords were also found, including more frequent mentions of health-related keywords on Twitter and more negative health outcomes from vaping mentioned on both Reddit and Twitter. Use of both computational and clinical content analyses is critical to not only identify signals of public health trends among vaping-related social media content but also to provide context for vaping risks and behaviors. By leveraging the strengths of both Twitter and Reddit as publicly available data sources, this research may provide technical and clinical insights to inform automatic detection of social media users who are vaping and may benefit from digital intervention and proactive outreach strategies on these platforms.

(J Med Internet Res 2022;24(12):e39460) doi:10.2196/39460

KEYWORDS

vaping; e-cigarette; social media; Twitter; Reddit; e-cigarette and vaping use–associated lung injury; EVALI; sentiment analysis; topic analysis

Introduction

Background

In the United States, vaping has become dramatically more popular in recent years, with 1 in every 20 American adults using vaping devices and >2 million middle- and high-school students in the United States using e-cigarettes in 2021 [1,2]. Vaping places individuals at risk for several negative health consequences including diminished lung function and cardiac performance, susceptibility to nicotine dependence, and impacted neurological development, particularly among youth [3,4]. However, despite these negative health consequences, youth and young adults have been found to report limited understanding of the dangers of vaping [5,6] and high perceived ability to quit vaping if desired [7]. Of further concern, e-cigarette and vaping use–associated lung injury (EVALI) resulted in hospitalizations and deaths in 2019, and many of these cases were later linked to vitamin E acetate (a filler substance in unregulated products) [8]. In the context of these risks and negative health outcomes, the United States Food and Drug Administration–labeled vaping among teens as a national epidemic in 2018 and continues to release policies to regulate vaping products more effectively [9]. Given the deleterious health effects of vaping and increased risks for EVALI, future research on publicly available, larger-scale data from sources such as social media are necessary to monitor this growing public health concern and to inform outreach interventions for vaping cessation. Previous literature has leveraged social media data for surveillance of health topics, including illicit drug use [10], mental well-being [11,12], public health [13,14], and other health-related experiences [15]. Twitter is a social media site that is used by approximately 22% (1/5) of American adults [16] as a source of information as well as information sharing [17]. Individuals on web-based platforms such as Twitter may be more willing to openly share experiences and personal stories about mental health or substance misuse with reduced fears of judgment or legal action, allowing them to access social support and advice and share this advice with others who are going through similar experiences [18]. For example, a study of 1200 tweets during mental health awareness week found that awareness, stigma, and personal experiences were central themes of discourse among Twitter users [19]. As such, Twitter has been used as a mass data source of information for public health monitoring and can be used to better understand attitudes and behaviors of individuals in relation to vaping [20-22]. For instance, during the COVID-19 pandemic, Twitter data were used to better understand sentiment and reactions to smoking in relation to the virus [23] as well as individual perspective of global-scale events and US-related lifestyle changes [24]. Although Twitter has several strengths related to surveillance and public health monitoring, other social media platforms such as Reddit may have complementary strengths to provide data on individual-level user vaping behaviors.

Reddit is a similar pseudonymous social media platform used by the public to discuss personal experiences that may be stigmatizing [25-27], including young adults who may disclose personal information with less fear of offline harm or consequences [28]. Reddit data have been used to investigate attitudes and behaviors of individuals who use illicit substances [29-31], and similar research has been conducted among those who vape. One analysis of Reddit threads indicated that primary motivations for vaping among individuals with mental illness include self-medication, freedom and control, vaping as a hobby, social connectedness, as well as vaping to quit smoking [32]. Other studies have used Reddit data to analyze public responses and concerns about vape bans [33], communities supporting e-cigarette cessation [34], and attitudes and reviews toward e-cigarette products [35].

Both Twitter and Reddit are popular social media platforms, but they differ in multiple ways that impact users' posting behaviors and post content. Twitter, with >300 million monthly active users [36], only allows short 280-character tweets for breaking news, trends, and opinions, often leading to incomplete or misleading statements [37]. In contrast, Reddit, with >430 million actively monthly users, has no character posting limit, is anonymous, and comprises network of communities, namely, subreddits, dedicated to specific topics, allowing users to relate to other individuals with similar backgrounds, views, and lived experiences. With Reddit's anonymity, people can honestly voice their own opinions with in-depth text and content to spread awareness and important news [38,39]. Thus, posts about the same topic during the same period (ie, posts about vaping in 2019) are expected to vary with regard to the type of content shared and the level of impact on public perception based on the platform on which they are shared.

Large-scale evaluations using computer science (CS) strategies, including those using natural language processing and machine learning for text mining, have been conducted previously on vaping content from social media [40-42]. For example, Visweswaran et al [41] developed machine learning classifiers to identify vaping-relevant tweets toward the development of a vaping surveillance system. Results demonstrate that social media content can be used for overall infoveillance, and such data could inform future, individual-level detection models to identify at-risk posts and users. A systematic review conducted by Kwon and Park [32] found that sentiment regarding vaping tended to be more positive across social media sites, and previous research on Twitter has demonstrated that those who smoke are more likely to engage with misinformation about vaping [43]. Studies conducted on Reddit posts have illustrated health symptoms associated with vaping [44] and highlighted communities aimed to support those wanting to quit vaping [45].

Objectives

Studying the EVALI public health crisis specifically could aid in the identification of content and keywords related to both

acute and long-term health outcomes associated with vaping shared on social media, as such signals of vaping risk may have been amplified during this period. By leveraging the strengths of both Twitter and Reddit as publicly available data sources as well as using an interdisciplinary approach to analyze complex social media content, technical and clinical insights may be garnered to inform the future development of an automatic detection model to connect with vaping users who may benefit from digital intervention on social media platforms. However, to date, there are few studies comparing insights from both Twitter and Reddit for substance misuse within the same time frame [46,47], and no known studies related to vaping have been conducted to analyze the 2019 EVALI outbreak at both the individual user level, and population level. As such, this paper examined vaping-related content on Twitter and Reddit to better understand the (1) sentiment and keywords associated with vaping-related content during the 2019 EVALI time frame, (2) differences in sentiment and keywords between content on Twitter and Reddit, and (3) similarities or differences between statistical analyses and clinical coding of vaping-related content.

Methods

Data Collection

In this study, we focused on comparing vaping-related keyword frequencies and sentiment on Twitter and Reddit during the EVALI outbreak period using data from both platforms from July 2019 to September 2019, as our previous work had identified this as a time frame during which vaping-related social media content increased [48]. To define the criteria for large-scale data extraction, our team first conducted a manual analysis of 200 randomly selected vaping-related tweets across the 2019 time line to generate a list of clinically relevant keywords. Our primary research questions guided the creation of this keyword list, which included *vaping*, *vape*, and 60 other specifying terms (Multimedia Appendix 1). Using this set of keywords, a random sample of 316,620 vaping-related tweets with an average of 27 words per tweet was extracted during the EVALI outbreak period (July, August, and September 2019). For comparison purposes, we used the same set of keywords to randomly extract Reddit data, resulting in 17,320 Reddit posts with an average of 211 words per post associated with vaping during the EVALI outbreak.

GetOldTweets [49] is an open-source python library that allowed our team to extract a random sample of tweets with our identified vaping keywords. This module permitted access to and extraction of historical tweets of any date and topic. The benefit of using this application program interface (API) is that it had no restrictions on size and provided access to historical tweets [49,50]. We used Pushshift Reddit API (version 4.0) [51], which provided rich features for searching and extraction and flexible ways to aggregate publicly available Reddit posts and comments.

Data Cleaning

After we extracted posts from Twitter and Reddit based on the keywords, we cleaned our data sets before further analysis. As we only focused on English-language posts in this study, we first removed the posts that contain non-English languages. We also removed invalid Reddit posts marked as “removed” or “deleted.” After that, the number of Twitter posts reduced from 316,620 to 286,703, and the number of Reddit posts reduced from 17,320 to 12,069.

For the text in the posts, we first converted all the characters to lowercases to avoid the case-sensitive process. Then, we removed all special characters non–American Standard Code for Information Interchange from the text. For text contractions, we expanded them into multiple individual words. Next, we removed the stop words that have no significant contributions to the meaning of the text from the text (eg, is, a, the, and of). After that, we removed the special terms from the tweet text, including mentions, hashtags, links, ticks, punctuations, numbers, and over spaces. Then, we applied the word lemmatization function to convert the words to their base forms.

Sentiment Analysis

Sentiment analysis is a common computer technique to measure the subjectivity, opinions, attitudes, and emotions in texts [52]. Sentiment analysis quantifies the sentiment contents in a given text along a continuum scale, for example, from −1 to 1 [41,53]. We applied Valence Aware Dictionary and sEntiment Reasoner (VADER) as the tool to analyze the sentiment of tweets and Reddit posts, as VADER is a lexicon and rule-based sentiment analysis tool [54] that recent studies [24,55,56] have found to effectively calculate sentiment social media analysis. More specifically, VADER has been attuned to social media sentiments and pretrained by a gold standard sentiment lexicon, which was developed based on mature sentiment word-banks, popular sentiment expression, and common slang with sentiment value in social media. To determine the sentiment, VADER maps lexical features to emotion intensities known as a sentiment score, which can be obtained by summing up the intensity of each word in the text. The score is then normalized to −1 (most extreme negative) and +1 (most extreme positive). In our study, if the text sentiment score was >0, then the text was classified as positive. The text was classified as negative if the sentiment score was <0. The neutral text’s sentiment score was 0. Our study further classified posts into positive, negative, and neutral sentiment toward vaping using this sentiment score, calculating the distribution of the posts in terms of the 3 sentiment types per month.

Keyword Analysis

In addition to the sentiment analysis described earlier, we used chi-square tests to compare differences between the frequency of keywords in Twitter and Reddit posts during each month across the following topics: (1) sentiment, (2) emotion-related keywords, (3) health-related keywords, (4) age-related keywords, (5) marketing-related keywords, (6) product-related keywords, (7) addiction-related keywords, and (8) quitting-related keywords.

Term Frequency–Inverse Document Frequency

Term frequency–inverse document frequency (TF-IDF) is a statistical measurement that can represent the word relevant in a corpus [57]. The TF-IDF score is calculated based on the term frequency and inverse document frequency. Using this method helps us find the common words on Twitter and on Reddit. On

the basis of the TF-IDF scores, we can identify the most important words on both the platforms. The formulas are as follows:

TF = number of a word in the document / number of words in the document **(1)**

IDF = log(number of documents / number of documents with the word) **(2)**

TF-IDF = TF × IDF **(3)**

Clinical Coding Comparison

During the EVALI outbreak, July, August, and September 2019 were identified as months during and just before the dramatic increase in vaping-related discussions on Twitter based on both the content and sentiment analyses outlined earlier. As such, a random sample of 200 posts per month from the Twitter and Reddit data sets described earlier were extracted for in-depth human coding toward contextual content analysis. Specifically, members of our clinical team with experience in substance use research (students in psychology, social work, or public health at the graduate level and with relevant experience coding qualitative social media data led by author PCR, a clinical psychologist) used inductive and deductive methods to construct a codebook based on a review of sample tweets and informed by previous literature [58,59]. Three primary coding categories were used: (1) type of post, including personal, marketing, or media or news or other [60]; (2) sentiment toward vaping [61]; and (3) health outcomes mentioned, including both positive (eg, quitting combustible smoking) and negative (eg, lung injury, death, and addiction or dependence) [60,62]. Secondary concepts that were coded as either present or not present included (1) mentions teens or adolescents or young adults [63] and (2) mentions marijuana or weed or cannabidiol or tetrahydrocannabinol [64,65]. Two independent human coders reviewed each post and assigned applicable codes based on text content, and agreement among coders was substantial as reflected by an average κ score of 0.62 [66]. A third coder then reviewed the coding from each preliminary coder and provided final codes for those tweets on which there was disagreement [67], which is a third-party resolution method used in previous qualitative analysis literature [68]. Both frequency and qualitative themes were then compared with the preliminary results from the CS analyses to aid in the conceptualization of the clinical themes reflected in the data set.

Total frequency of each theme mentioned on both Twitter and Reddit was compared across the months of July, August, and September 2019 (sum of 3 months) to demonstrate relative weight of each topic on the respective platforms.

Ethics Approval

The Washington University Institutional Review Board (202101009) reviewed the methods of data extraction and analysis for this study. Given that the data are publicly available on social media, the study was determined to be nonhuman subjects research and exempt from review.

Results

Data Set Summary and Unique Users

This section presents the results from the high-throughput computational analyses. In total, we collected 286,703 tweets and 12,096 posts on Reddit. The sample size differences between Twitter and Reddit were related to the amount of information included in each Reddit post and in a tweet. The word limit for each tweet is 280 characters, whereas the word limit for each Reddit post is 40,000 characters. Thus, each Reddit post included much richer information than a tweet. To analyze the data set at the word level and further content analysis, the number of extracted Reddit posts was significantly smaller than the number of tweets. Table 1 presents the number of unique users and posts per user on both platforms. Overall, the number of vaping-related posts and unique users on Twitter and Reddit had an increasing trend from July 2019 to September 2019. In particular, the number of posts and unique users on Twitter increased by approximately 4 times from August 2019 to September 2019. The number of posts per user on Twitter and Reddit increased from 1.68 to 1.81 and 1.19 to 1.21, respectively.

Table 1. Number of unique users and posts per user on Twitter and Reddit mentioning vaping during the e-cigarette and vaping use–associated lung injury outbreak.

Month in 2019	Unique users, n (%)		Posts per user, n	
	Twitter	Reddit	Twitter	Reddit
July	17,904 (11.06)	2893 (28.75)	1.68	1.19
August	28,604 (17.67)	3066 (30.47)	1.66	1.2
September	115,373 (71.27)	4105 (40.79)	1.81	1.21

Sentiment Analysis Results

CS pattern analysis of sentiment found that overall posts with positive sentiment about vaping were more common than negative posts on Reddit (8905/12,096, 73.62%), and negative sentiment was dominant on Twitter (174,448/286,703, 60.86%) during the EVALI period (Table 2). Clinical results based on a small random sample during this period were similar to the results using CS methods, still demonstrating that Reddit had a higher number of positive sentiment posts and also reflecting that Twitter had a higher number of negative sentiment posts based on manual review of post content.

The results of monthly sentiment trends indicated that the percentage of posts with positive sentiment was higher than that with negative sentiment in July both on Twitter and on Reddit. In August and September, the percentage of negative posts was higher than that of the positive ones on Twitter. Moreover, there was a significant decrease in the percentage of positive sentiment

from July to September on Twitter, whereas positive posts were dominant on Reddit in August and September.

The chi-square tests (Table S1 in Multimedia Appendix 2) found an overall significant difference in sentiment between platforms. Twitter contained significantly more negative postings (174,488/286,703, 60.86%) than Reddit (2281/12,096, 18.86%), and Reddit contained significantly more positive posts (8905/12,095, 73.62%) than Twitter (85,209/286,703, 29.72%).

In addition to the sentiment analysis and trends, we also ran chi-square tests to compare emotion expression–related posting differences on Twitter and Reddit. We selected common emotional words from the list of most frequent words on both Twitter and Reddit. Positive keywords included *safe*, *good*, and *love*, and negative keywords included *kill*, *bad*, *dangerous*, *concern*, and *serious*. The statistical results indicated significant posting differences between the 2 platforms as a whole, based on their frequency percentages. We found that positive emotion expressions were much more significant on Reddit than on Twitter in all 3 months during the EVALI outbreak period (Table S2 in Multimedia Appendix 2).

Table 2. Sentiment analysis and clinical coding on Twitter and Reddit.

	Sentiment analysis, n (%)		Clinical coding, n (%)	
	Twitter (n=286,703)	Reddit (n=12,096)	Twitter (n=577)	Reddit (n=613)
Positive	85,209 (29.72)	8905 (73.62)	201 (34.8)	291 (47.5)
Negative	174,488 (60.86)	2281 (18.86)	273 (47.3)	184 (30)
Neutral	27,006 (9.42)	910 (7.52)	103 (17.9)	138 (22.5)

Keyword Analysis by Topic

Health-Related Keyword Analysis

The distributions and percentages of the posts that contained vaping health-related keywords are shown in Table 3. Figure 1 presents the frequency of the top 6 words associated with health issues in July, August, and September 2019. The top 6 words were commonly shared between Twitter and Reddit. On the basis of the TF-IDF scores as shown in Multimedia Appendix 3, we found that the most important health-related keywords often mentioned on Twitter included *death*, *lung*, *quit*, *smoking*, *disease*, and *harm*, whereas the most important words in the Reddit posts included *death*, *lung*, *quit*, *smoking*, *cough*, and *doctor*.

We performed a chi-square test to compare health-related keywords, including *death*, *lung*, *disease*, *risk*, *crisis*, *sick*, *doctor*, *cancer*, *injury*, *epidemic*, *research*, *damage*, *harm*, *harmful*, *patient*, *cough*, *chest*, *prevention*, *smoking*, and *quit* based on the posts in July, August, and September 2019. The chi-square test results (Table S3 in Multimedia Appendix 2) showed significant differences between health-related keywords posting on Twitter and Reddit for each of the 3 months and as a whole. However, owing to the significant differences between the size of posts on Twitter and Reddit, the overall effect size was small. On the basis of the percentages, more health-related keywords were discussed on Twitter than on Reddit, and negative health outcomes were highly discussed on both Reddit and Twitter.

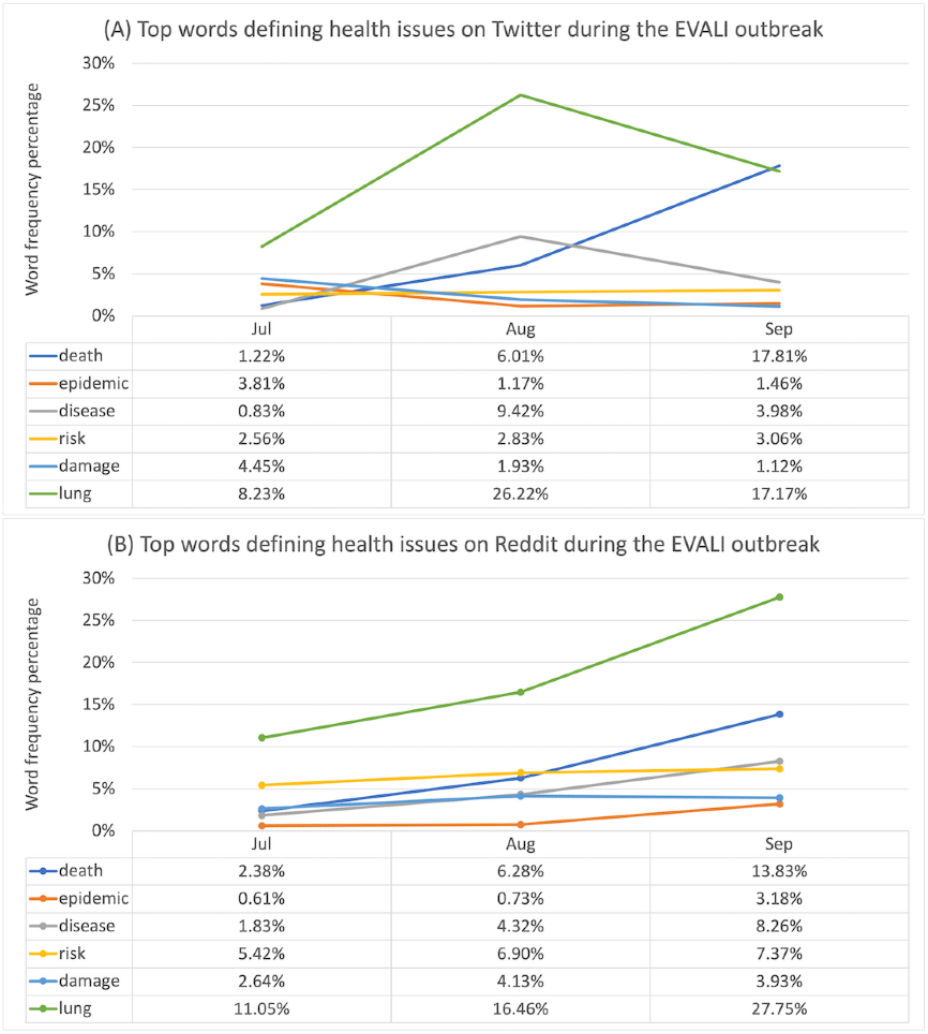
In addition to investigating the sentiment of health-related keywords, chi-square tests associated with addiction-related keywords (Table S4 in Multimedia Appendix 2) showed significant differences and small effect sizes between platforms in each month and the entire EVALI outbreak period. On the basis of percentages, the addiction-related keywords were mentioned more significantly on Twitter than on Reddit.

Within the in-depth clinical coding, negative health outcomes were mentioned much more frequently on both the platforms (Twitter: 230/577, 39.9% and Reddit: 227/578, 39.3%) than positive health outcomes (Twitter: 134/577, 23.2% and Reddit: 182/578, 31.5%). Additional topic mentioned within these negative health outcomes included EVALI/hospitalization, which was more prevalent on Twitter (Twitter: 176/577, 30.5% and Reddit: 146/578, 25.3%), whereas addiction or dependence on vaping products was mentioned more often on Reddit (Twitter: 57/577, 9.9% and Reddit: 123/578, 21.3%). Those mentioning positive health outcomes related to vaping were more common on Reddit, consistent with the keyword analysis described earlier; further, clinical coding found that vaping as a means of quitting combustible smoking was more often mentioned on Reddit than on Twitter (Twitter: 118/577, 20.5% and Reddit: 177/578, 30.6%).

Table 3. Distribution and percentage of health-related keywords on Twitter and Reddit.

Health-related keywords	July 2019, n (%)		August 2019, n (%)		September 2019, n (%)		Total, n (%)	
	Twitter	Reddit	Twitter	Reddit	Twitter	Reddit	Twitter	Reddit
death	343 (1.1)	70 (2)	2701 (5.67)	129 (3.5)	32,971 (15.77)	493 (9.9)	36,015 (12.56)	692 (5.7)
lung	2305 (7.67)	281 (8.2)	11,612 (24.39)	411 (11.2)	33,394 (15.98)	870 (17.5)	47,311 (16.5)	1562 (12.91)
disease	235 (0.8)	56 (2)	4268 (8.96)	112 (3.0)	7969 (3.8)	297 (6.0)	12,472 (4.35)	465 (3.8)
risk	708 (2.4)	147 (4.3)	1231 (2.59)	197 (5.4)	6012 (2.88)	280 (5.6)	7951 (2.77)	624 (5.2)
crisis	81 (0.3)	9 (0.3)	216 (0.5)	10 (0.3)	6072 (2.90)	102 (2.1)	6369 (2.22)	121 (1)
sick	415 (1.4)	148 (4.3)	1054 (2.21)	182 (4.9)	5467 (2.62)	370 (7.5)	6936 (2.42)	700 (5.8)
doctor	682 (2.3)	183 (5.3)	2359 (4.95)	228 (6.2)	4623 (2.21)	323 (6.5)	7664 (2.67)	734 (6.1)
cancer	442 (1.5)	84 (2)	782 (1.64)	80 (2)	3691 (1.77)	139 (2.8)	4915 (1.71)	303 (2.5)
injury	96 (0.3)	31 (1)	1256 (2.64)	45 (1)	3990 (1.91)	104 (2.1)	5342 (1.86)	180 (1.5)
epidemic	1091 (3.63)	16 (0.5)	533 (1.12)	24 (0.7)	2920 (1.40)	139 (2.8)	4544 (1.58)	179 (1.5)
research	523 (1.7)	159 (4.6)	712 (1.49)	178 (4.8)	3006 (1.44)	264 (5.3)	4241 (1.48)	601 (5.0)
damage	1315 (4.37)	78 (2)	882 (1.85)	119 (3.2)	2237 (1.07)	156 (3.1)	4434 (1.55)	353 (2.9)
harm	1503 (5.00)	139 (4.0)	1888 (3.97)	180 (4.9)	8253 (3.95)	303 (6.1)	11,644 (4.06)	622 (5.1)
harmful	492 (1.6)	40 (1)	643 (1.35)	49 (1)	2892 (1.38)	143 (2.9)	4027 (1.40)	232 (1.9)
patient	162 (0.5)	61 (2)	1103 (2.31)	82 (2)	1741 (0.83)	122 (2.5)	3006 (1.05)	265 (2.2)
cough	262 (0.9)	141 (4.1)	415 (0.9)	163 (4.4)	1130 (0.54)	257 (5.2)	1807 (0.63)	561 (4.6)
chest	105 (0.4)	108 (3.1)	127 (0.3)	133 (3.6)	431 (0.2)	227 (4.6)	663 (0.2)	468 (3.9)
prevention	96 (0.3)	6 (0.2)	246 (0.5)	14 (0.4)	529 (0.3)	31 (1)	871 (0.3)	51 (0.4)
smoking	3486 (11.60)	430 (12.5)	4145 (8.71)	435 (11.8)	15,604 (7.47)	604 (12.2)	23,235 (8.10)	1469 (12.15)
quit	3017 (1036)	823 (23.9)	3567 (7.49)	853 (23.2)	17,365 (8.31)	1160 (23.36)	23,949 (8.35)	2836 (23.45)

Figure 1. Top words defining health issues on Twitter (A) and Reddit (B) during the e-cigarette and vaping use–associated lung injury (EVALI) outbreak.

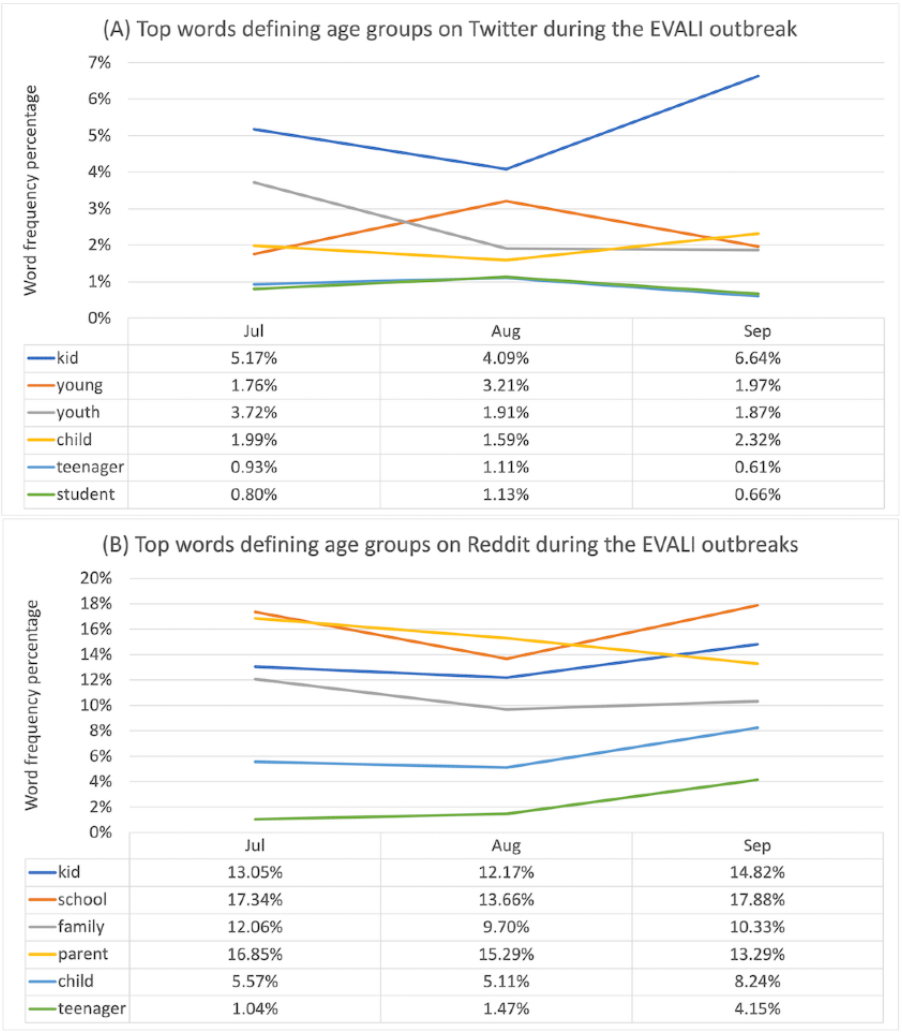


Age-Related Keyword Analysis

The top 6 words related to age groups in July, August, and September 2019 are presented in Figure 2 for Twitter and Reddit based on frequency. Among the age-related keywords, *kids* was the most used word on Twitter and Reddit after August 2019. Other frequently used words on Twitter included *youth, young, child,* and *teenager*. Reddit posts more often contained words such as *parent, school,* and *family*.

Age-related keywords in our data set included *kid, adult, child, young, old, youth, parent, school, age, student, family, teenager, minor, mother, husband, wife, adolescent, father,* and *aunt* in July, August, and September 2019 separately. The chi-square test results (Table S5 in Multimedia Appendix 2) showed significant differences and small effect sizes between age-related keywords on Twitter and Reddit for each of the 3 months and as a whole and indicated that age-related keywords were more frequently mentioned on Twitter than on Reddit. Clinical review of post content focused only on mentions of youth and young adults and found differing results, showing that Twitter had 22.9% (132/577) of tweets mentioning youth and Reddit had 28.5% (165/578) of posts mentioning this group.

Figure 2. Top words on age groups on Twitter (A) and Reddit (B) during the e-cigarette and vaping use–associated lung injury (EVALI) outbreak.

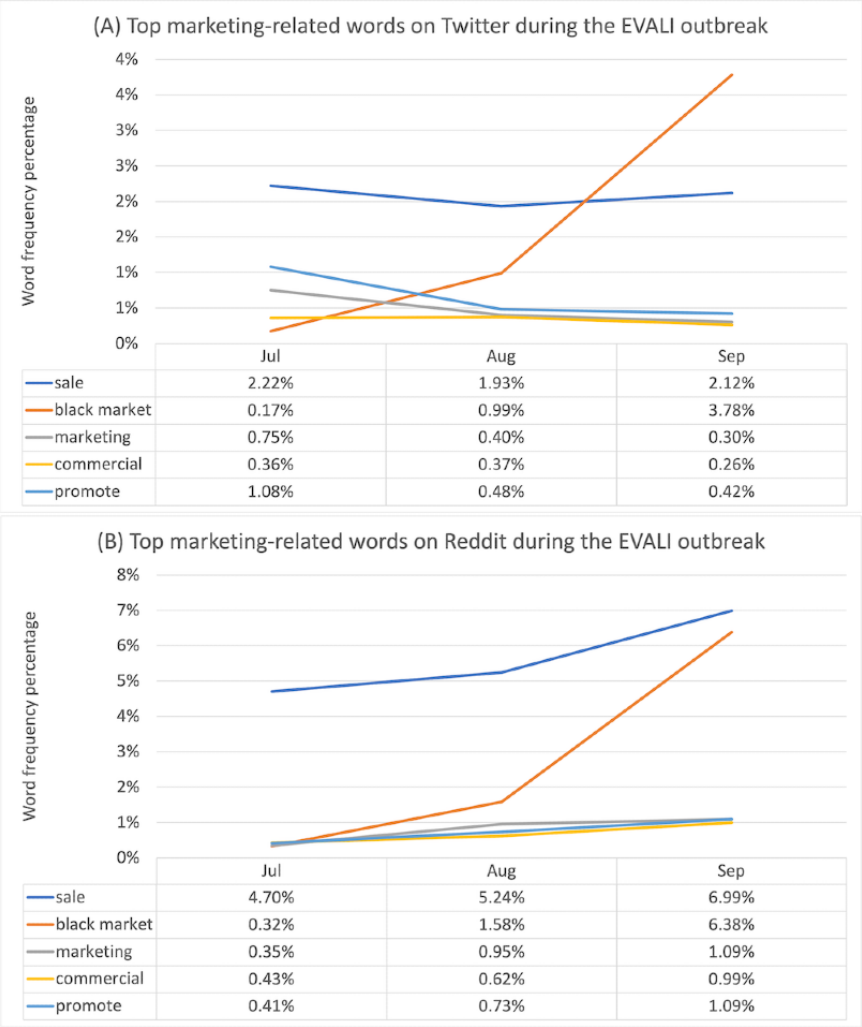


Marketing-Related Keyword Analysis

Frequently used words about vaping marketing were highly similar between Twitter and Reddit, including *sale, commercial, market, black market,* and *promote*. The trends of top 5 marketing-related words on Twitter and Reddit during the 3 months of EVALI outbreak are illustrated in Figure 3. Mentions of *sale, black market,* and *commercial* increased on Twitter and Reddit from July 2019 to September 2019.

Marketing-related keywords in our data set included *black market, black market, market, sale, news, promote, marketing, commercial, blackmarket,* and *media* for July, August, and September 2019. The chi-square test results (Table S6 in Multimedia Appendix 2) showed significant differences with small effect sizes between marketing-related keywords posting on Twitter and Reddit for all the 3 months, indicating that they were discussed more frequently on Twitter than on Reddit. Clinical analyses focused on marketing regulation and policies and had different results, showing that Twitter discussed policies 8.3% (48/577) of the time and Reddit discussed policies 20% (116/578) of the time.

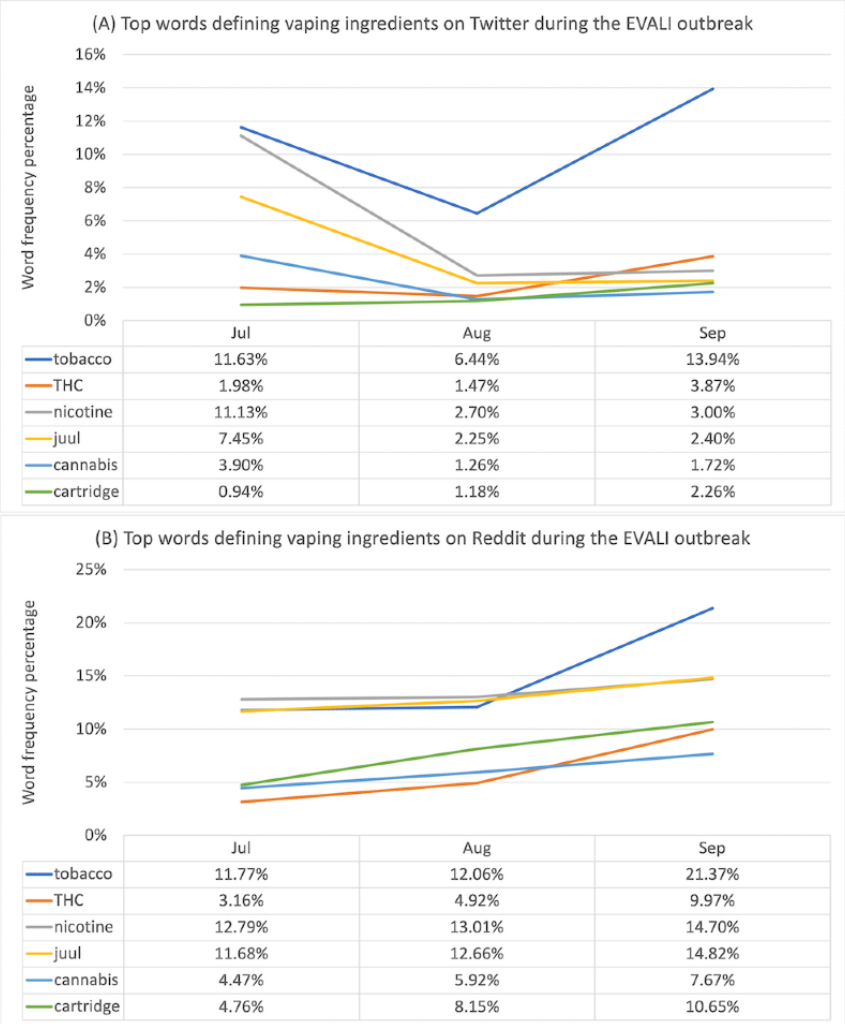
Figure 3. Top marketing-related words on Twitter (A) and Reddit (B) during the e-cigarette and vaping use–associated lung injury (EVALI) outbreak.



Vaping Product Keyword Analysis

The detailed distributions and percentages of the vaping product keywords are listed in Multimedia Appendix 3, and the top words related to vaping substances on Twitter and Reddit are illustrated in Figure 4. On both platforms, the most frequent word about vaping ingredients or products was *cigarette*, and mentions of marijuana-related keywords (*weed*, *CBD*, *THC*, and *cannabis*) and alcohol were also prevalent. On Reddit, specific keywords about product components, such as *juice*, *cartridge*, and *liquid*, were slightly more common. The most common words on Reddit included *cigarette*, *product*, and *juice*, which varied across months. The most common words on Twitter included *cigarette*, *tobacco*, and *product* and stayed consistent across August 2019 to September 2019. On the basis of the TF-IDF scores as shown in Multimedia Appendix 3, we found that the most important words in the posts from Twitter included *cig*, *cigarette*, *tobacco*, *product*, *thc*, and *nicotine*, whereas the most important words in the Reddit posts included *nicotine*, *cigarette*, *juice*, and *weed*.Vaping product-related keywords in our data set included *cigarette*, *tobacco*, *product*, *thc*, *cig*, *nicotine*, *juice*, *juul*, *cartridge*, *liquid*, *cannabis*, *chemical*, *alcohol*, *ecigarette*, *weed*, *cbd*, *flavour*, and *ingredient* based on the data sets in July, August, and September 2019. The chi-square test results (Table S7 in Multimedia Appendix 2) showed significant differences with small effect sizes between vaping product-related keywords posting on Twitter and Reddit for all the 3 months, finding that more vaping product-related keywords were mentioned on Twitter based on percentages. Clinical analyses found different results, showing that marijuana-related keywords were mentioned more than twice as often on Reddit (208/578, 35.9%) than on Twitter (77/577, 13.3%).

Figure 4. Top words defining vaping ingredients on Twitter (A) and Reddit (B) during the e-cigarette and vaping use–associated lung injury (EVALI) outbreak. THC: tetrahydrocannabinol.



Quitting Vaping

In addition, quit-related keywords in our data set included *quit*, *quitting*, *stop*, and *stopper* to compare pattern differences on Twitter and Reddit. The chi-square test results indicated significant posting differences with small effect sizes between the 2 platforms for all the 3 months and as a whole, showing quit-related words mentioned on Twitter more often based on percentages (Table S8 in Multimedia Appendix 2). Clinical analyses showed different results, with Reddit having 21.8% (126/578) of posts related to quitting and Twitter having 6.4% (37/577) of tweets related to quitting.

Discussion

Principal Findings and Implications

As vaping has become more popular in recent years, so have discussions about its direction, policies, and health connotations on social media platforms, and this study illustrated differences in sentiment and keyword content on Twitter and Reddit during the EVALI outbreak in 2019. According to the trends in the frequency of vaping-related posts during this time frame, vaping-related content increased slowly between July and August, with a dramatic spike from August to September. Moreover, there was a significant increase in the number of unique Twitter and Reddit users who participated in these discussions during the EVALI outbreak. The fact that increasing trends in the frequency of social media vaping-related content peaked in parallel with the EVALI outbreak and across both popular social media platforms supported the utility of social media as a surveillance system for exploring naturally occurring, real-time reactions and communications during a public health vaping-associated crisis.

Importantly and based on our content analysis, Twitter and Reddit content within posts about vaping were found to contain primarily positive sentiment about vaping. However, the 2 platforms were notably different based on the most prevalent type of content identified. Specifically, Reddit users tended to reveal personal vaping experiences and opinions about vaping benefits, policies, and products, including how potential restrictive vaping policies may have negative impacts on users who vape (ie, less access to vaping products that aid cigarette smoking cessation). Mentions of marijuana were also >2 times as high on Reddit as on Twitter and often included queries to other Reddit users about the safety of specific vaping products and which symptoms, if any, should warrant concern or medical care. In contrast, Twitter included more mainstream media content surrounding vaping, specifically related to the rise in

EVALI cases. We also observed that Twitter feeds contained attention-grabbing negative sentiment and higher use of negative emotional expressions, including *kill*, *bad*,*dangerous*, *concern*, and *serious*, as well as increased content on possible negative health outcomes of vaping, including addiction. Although both platforms had mentions of youth, Twitter highlighted headlines about the youth vaping epidemic and EVALI among teens and ways to limit vaping products for adults who use them as smoking cessation aids, whereas on Reddit, mentions related to youth mostly were individuals describing their own vaping behaviors, including initiating vaping behaviors as a teen.

In summary, we observed numerous and meaningful distinctions in the frequencies of content topics across both social media outlets. These differences may be owing to the way individuals socially network as well as their motive for discussion on each platform. For instance, information on Twitter is known as "the" social media platform for news coverage, and it is most often used by journalists and major news providers to broadcast news and update the public in real time as important events transpire [69]. This may explain why Twitter had a higher frequency of negative posts related to vaping, as journalists and their audience leveraged this platform for updates and interactions throughout the unfolding of the EVALI outbreak, especially as it evolved into a crisis that resulted in many hospitalizations and several deaths. In contrast, Reddit distinguishes itself from other social media platforms by facilitating more candid discussions, including exchanges about substance use behaviors, given its pseudonymous user system and generous character limit restrictions; this may be why we found a higher prevalence of content describing one's personal experiences with vaping.

Comparison With Previous Work

It might also be that the differences we found were owing to the distinctions between the users themselves. For instance, the demographic user base of Twitter is predominantly White adults, who have a higher degree of education and are more likely to be identified as Democrat than the general public, with 10% of users creating 80% of the tweets [70]. In contrast, Reddit users tend to span degrees of education attainment and live in urban or suburban areas [71]. The Centers for Disease Control and Prevention finds that within people of color, there are higher percentages of individuals who vape compared with the percentage of White people who vape [72], and another study shows that higher level of education attainment was linked to lower odds of e-cigarette use [73]. This suggests that users on Reddit may be more likely to vape than users on Twitter, explaining their different sharing patterns of personal vaping-related experiences and concerns over restrictive policy.

Limitations

The findings of this paper should be considered within the context of its limitations. First, we analyzed only text-based posts or messages on these platforms. Although this provided us with data-heavy information from each social media site, it did not include the multitude of multimedia content including photos, videos, and links that are available for further analysis. Second, owing to the character limits on Twitter and the unlimited length of Reddit posts, the differences between the number of words in each post could have impacted both the sentiment and keyword analyses in this study. Third, because of the timing of our data gathering, we did not garner information related to COVID-19 and its implications on those who vape and vaping policies, leaving us unable to discern more recent implications. Fourth, our original keyword list used to extract the vaping-related data sets from Twitter and Reddit may have contained more negative health-related keywords, and this could have impacted the results with regard to sentiment and health outcomes, causing a potential selection bias in our keyword list. As this study focused on the health issue regarding vaping-related topics on Twitter and Reddit during the outbreak period, the keyword list included multiple sentiment-related words. It will bring bias to our sentiment analysis results, but these words were the key to selecting the related posts and addressing our research questions. In addition, we applied the GetOldTweets and Pushshift APIs to extract the data based on the keyword list. As the extraction mechanism of these APIs is to find the posts with the same field as one of the keywords without further filtration to matched posts, the extracted data set might include the posts from bots instead of real users, which may introduce bias to our sentiment results. We plan to apply different methods to clean the posts generated by bots in our future studies. However, the use of this data set was in line with the larger aims of this study, which were to better understand the content and sentiment surrounding vaping on Twitter and Reddit to inform the development of potential identification and outreach methods on social media to those at risk of negative health outcomes to improve public health. The fifth limitation was that we applied an existing tool VADER to analyze the sentiment of the posts, and thus, it could bias to our sentiment analysis results, which are common issues for any sentiment analysis tool owing to the complex dynamics of human expressions, emotions, and contexts. In the future, we will also consider creating a sentiment analysis model optimization with social media posts to overcome the current disadvantages of not effectively identifying sarcastic sentences.

Future Directions

Overall, the results of this study revealed the strengths of both Twitter and Reddit as publicly available social media data sources as a public health crisis transpired and evolved. Health practitioners working with those who vape or who have interest in quitting vaping should be aware of the information and possible misinformation related to vaping and work to assess whether social media engagement on various platforms could impact continued use or be a barrier to cessation. The results shared in this manuscript could also inform social media companies and public health officials by alerting them to the marketing of vaping products on these sites and encouraging protections for communities such as those on Reddit aimed to support vaping cessation. In addition, to improve public health reach, future research could explore automatic detection mechanisms that leverage each platform's content and type of networking identified here, especially to study the potential for identifying users that are vaping and may want information and support to quit. This could help lead to efficient and timely social media informed proactive outreach strategies to distribute health education about vaping, including strategies for vaping cessation.

Acknowledgments

The authors would like to acknowledge the funding support provided by the University of South Carolina (USC), Columbia, South Carolina, United States (grant 80002838); partial support from the USC Big Data Health Science Center, a USC excellence initiative program (grant BDHSC-2021-14 and BDHSC-2022); and a research grant from the USC Advancing Chronic Care Outcome through Research and Innovation Center (ACORN-2022) and National Institutes of Health (grant K02 DA043657, UL1 TR002377, and R34 DA054725). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Data Availability

Access to the data set analyzed in this manuscript may be made available to researchers via reasonable request to the corresponding author.

Authors' Contributions

DW and EK contributed equally, and PAC-R served as a senior author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords and terms used for data extraction.
[DOCX File , 13 KB - jmir_v24i12e39460_app1.docx]

Multimedia Appendix 2
Statistical analysis results.
[DOCX File , 29 KB - jmir_v24i12e39460_app2.docx]

Multimedia Appendix 3
Distribution and term frequency–inverse document frequency score of vaping product–related keywords on Twitter and Reddit.
[DOCX File , 22 KB - jmir_v24i12e39460_app3.docx]

References

1. Mirbolouk M, Charkhchi P, Kianoush S, Uddin SM, Orimoloye OA, Jaber R, et al. Prevalence and distribution of e-cigarette use among U.S. adults: behavioral risk factor surveillance system, 2016. Ann Intern Med 2018 Oct 02;169(7):429-438. [doi: 10.7326/M17-3440] [Medline: 30167658]
2. Youth E-Cigarette Use Remains Serious Public Health Concern Amid COVID-19 Pandemic. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/media/releases/2021/p0930-e-cigarette.html [accessed 2021-12-07]
3. Singh S, Windle SB, Filion KB, Thombs BD, O'Loughlin JL, Grad R, et al. E-cigarettes and youth: patterns of use, potential harms, and recommendations. Prev Med 2020 Feb 03;133:106009. [doi: 10.1016/j.ypmed.2020.106009] [Medline: 32027913]
4. El-Mahdy MA, Mahgoup EM, Ewees MG, Eid MS, Abdelghany TM, Zweier JL. Long-term electronic cigarette exposure induces cardiovascular dysfunction similar to tobacco cigarettes: role of nicotine and exposure duration. Am J Physiol Heart Circ Physiol 2021 May 01;320(5):H2112-H2129 [FREE Full text] [doi: 10.1152/ajpheart.00997.2020] [Medline: 33606584]
5. Park E, Kwon M, Gaughan MR, Livingston JA, Chang Y. Listening to adolescents: their perceptions and information sources about e-cigarettes. J Pediatr Nurs 2019;48:82-91. [doi: 10.1016/j.pedn.2019.07.010] [Medline: 31362205]
6. Morean ME, Bold KW, Kong G, Gueorguieva R, Camenga DR, Simon P, et al. Adolescents' awareness of the nicotine strength and e-cigarette status of JUUL e-cigarettes. Drug Alcohol Depend 2019 Nov 01;204:107512 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.05.032] [Medline: 31487572]
7. Strombotne K, Sindelar J, Buckell J. Who, me? Optimism bias about US teenagers' ability to quit vaping. Addiction 2021 Nov;116(11):3180-3187 [FREE Full text] [doi: 10.1111/add.15525] [Medline: 33891783]
8. Outbreak of Lung Injury Associated with the Use of E-Cigarette, or Vaping, Products. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html#latest-information [accessed 2021-09-11]
9. Vaping Illness Update: FDA Warns Public to Stop Using Tetrahydrocannabinol (THC)-Containing Vaping Products and Any Vaping Products Obtained Off the Street. U.S. Food & Drug Administration. 2021. URL: https://www.fda.gov/

consumers/consumer-updates/vaping-illness-update-fda-warns-public-stop-using-tetrahydrocannabinol-thc-containing-vaping [accessed 2021-09-11]
10. Kazemi DM, Borsari B, Levine MJ, Dooley B. Systematic review of surveillance by social media platforms for illicit drug use. J Public Health (Oxf) 2017 Dec 01;39(4):763-776 [FREE Full text] [doi: 10.1093/pubmed/fdx020] [Medline: 28334848]
11. Skaik R, Inkpen D. Using social media for mental health surveillance: a review. ACM Comput Surv 2021 Nov 30;53(6):1-31. [doi: 10.1145/3422824]
12. Amir S, Dredze M, Ayers JW. Mental health surveillance over social media with digital cohorts. In: Proceedings of the 6th Workshop on Computational Linguistics and Clinical Psychology. 2019 Presented at: CLPsych '19; June 6, 2019; Minneapolis, MN, USA p. 114-120. [doi: 10.18653/v1/w19-3013]
13. Paul MJ, Sarker A, Brownstein JS, Nikfarjam A, Scotch M, Smith K, et al. Social media mining for public health monitoring and surveillance. In: Proceedings of the 2016 Pacific Symposium on Biocomputing. 2016 Presented at: PBS '16; January 4-8, 2016; Big Island, HI, USA p. 468-479. [doi: 10.1142/9789814749411_0043]
14. Aiello AE, Renson A, Zivich PN. Social media- and Internet-based disease surveillance for public health. Annu Rev Public Health 2020 Apr 02;41:101-118 [FREE Full text] [doi: 10.1146/annurev-publhealth-040119-094402] [Medline: 31905322]
15. Jiang K, Feng S, Song Q, Calix RA, Gupta M, Bernard GR. Identifying tweets of personal health experience through word embedding and LSTM neural network. BMC Bioinformatics 2018 Jun 13;19(Suppl 8):210 [FREE Full text] [doi: 10.1186/s12859-018-2198-y] [Medline: 29897323]
16. Odabaş M. 10 facts about Americans and Twitter. Pew Research Center. 2022 May 5. URL: https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/ [accessed 2022-03-30]
17. 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 Human Behav 2017 Jul;72:115-122. [doi: 10.1016/j.chb.2017.02.041]
18. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. J Med Internet Res 2017 Apr 05;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
19. Makita M, Mas-Bleda A, Morris S, Thelwall M. Mental health discourses on Twitter during mental health awareness week. Issues Ment Health Nurs 2021 May;42(5):437-450. [doi: 10.1080/01612840.2020.1814914] [Medline: 32926796]
20. Jordan SE, Hovet SE, Fung IC, Liang H, Fu KW, Tse ZT. Using Twitter for public health surveillance from monitoring and prediction to public response. Data 2019;4(1):6. [doi: 10.3390/data4010006]
21. Dai X, Bikdash M, Meyer B. From social media to public health surveillance: word embedding based clustering method for Twitter classification. In: Proceedings of the 2017 SoutheastCon. 2017 Presented at: SECON '17; March 30-April 2, 2017; Concord, NC, USA p. 1-7. [doi: 10.1109/secon.2017.7925400]
22. Allem JP, Majmundar A, Dharmapuri L, Cruz TB, Unger JB. E-liquid-related posts to Twitter in 2018: thematic analysis. Addict Behav Rep 2019 Dec;10:100196 [FREE Full text] [doi: 10.1016/j.abrep.2019.100196] [Medline: 31431917]
23. McCausland K, Maycock B, Leaver T, Wolf K, Freeman B, Jancey J. E-cigarette advocates on Twitter: content analysis of vaping-related tweets. JMIR Public Health Surveill 2020 Oct 14;6(4):e17543 [FREE Full text] [doi: 10.2196/17543] [Medline: 33052130]
24. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: longitudinal analysis of Twitter data. J Med Internet Res 2020 Dec 14;22(12):e21418 [FREE Full text] [doi: 10.2196/21418] [Medline: 33284783]
25. Rhidenour KB, Blackburn K, Barrett AK, Taylor S. Mediating medical marijuana: exploring how veterans discuss their stigmatized substance use on Reddit. Health Commun 2022 Sep;37(10):1305-1315. [doi: 10.1080/10410236.2021.1886411] [Medline: 33602000]
26. Pestana J, Beccaria F, Petrilli E. Psychedelic substance use in the Reddit psychonaut community. A qualitative study on motives and modalities. Drugs Alcohol Today 2021;21(2):112-123. [doi: 10.1108/dat-03-2020-0016]
27. Green JL, Vosburg SK, Robbins R, Faraone SV, Antshel KM. Substance use trajectories: nonmedical use (NMU) of prescription stimulants via non-oral routes of administration among adults recruited from Reddit. CNS Spectr 2021 May 10;26(2):166-167 [FREE Full text] [doi: 10.1017/S1092852920002692]
28. Andalibi N, Haimson OL, De Choudhury M, Forte A. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 May Presented at: CHI '16; May 7-12, 2016; San Jose, CA, USA p. 3906-3918. [doi: 10.1145/2858036.2858096]
29. Kaufman MR, Bazell AT, Collaco A, Sedoc J. "This show hits really close to home on so many levels": an analysis of Reddit comments about HBO's Euphoria to understand viewers' experiences of and reactions to substance use and mental illness. Drug Alcohol Depend 2021 Mar 01;220:108468 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108468] [Medline: 33540349]
30. Vosburg SK, Robbins RS, Antshel KM, Faraone SV, Green JL. Characterizing pathways of non-oral prescription stimulant non-medical use among adults recruited from Reddit. Front Psychiatry 2021 Jan 25;11:631792 [FREE Full text] [doi: 10.3389/fpsyt.2020.631792] [Medline: 33597899]

31. Lu J, Sridhar S, Pandey R, Al Hasan M, Mohler G. Investigate transitions into drug addiction through text mining of Reddit data. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019 Jul Presented at: KDD '19; August 4-8, 2019; Anchorage, AK, USA p. 2367-2375. [doi: 10.1145/3292500.3330737]
32. 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]
33. Xu YA, Kim HM, Wang Y, Wu J, Hong T, McLaughlin M. Public responses and concerns regarding vape bans on Reddit: a longitudinal topic modeling approach. In: Proceedings of the 13th International Conference on Social Computing and Social Media: Experience Design and Social Network Analysis. 2021 Presented at: SCSM '21; July 24–29, 2021; Virtual p. 391-403. [doi: 10.1007/978-3-030-77626-8_26]
34. Shaw K, Watkins SL. E-cigarette Cessation on Reddit:How users discuss their cessation experience, social support, and recommendations to quit. College of Public Health, The University of Iowa. 2020. URL: https://www.public-health.uiowa.edu/wp-content/uploads/2020/12/Kamara-Shaw.pdf [accessed 2022-03-30]
35. Allem JP, Majmundar A, Dharmapuri L, Unger JB, Cruz TB. Insights on electronic cigarette products from reviews on the Reddit forum. Tob Prev Cessat 2019 Sep 12;5:29 [FREE Full text] [doi: 10.18332/tpc/111533] [Medline: 32411892]
36. Iqbal M. Twitter revenue and usage statistics. Business of Apps. 2022. URL: https://www.businessofapps.com/data/twitter-statistics/ [accessed 2022-01-29]
37. Seema. 3 reasons why Twitter is so toxic – and how you can protect yourself. Notifier. 2021. URL: https://notifier.so/docs/3-reasons-why-twitter-is-so-toxic-and-how-you-can-protect-yourself/ [accessed 2022-01-29]
38. Sattleberg W. The Demographics Of Reddit: Who Uses The Site? Alphr. 2021 Apr 6. URL: https://www.alphr.com/demographics-reddit/ [accessed 2022-01-29]
39. Rutnik M. Change my mind: Reddit is a social network and it's the best one out there. Android Authority. 2021 Oct 17. URL: https://www.androidauthority.com/reddit-vs-facebook-vs-twitter-3037642/ [accessed 2022-01-29]
40. Myslín M, Zhu SH, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
41. Visweswaran S, Colditz JB, O'Halloran P, Han NR, Taneja SB, Welling J, et al. Machine learning classifiers for Twitter surveillance of vaping: comparative machine learning study. J Med Internet Res 2020 Aug 12;22(8):e17478 [FREE Full text] [doi: 10.2196/17478] [Medline: 32784184]
42. Janmohamed K, Soale AN, Forastiere L, Tang W, Sha Y, Demant J, et al. Intersection of the web-based vaping narrative with COVID-19: topic modeling study. J Med Internet Res 2020 Oct 30;22(10):e21743 [FREE Full text] [doi: 10.2196/21743] [Medline: 33001829]
43. Liu J, Wright C, Williams P, Elizarova O, Dahne J, Bian J, et al. Smokers' likelihood to engage with information and misinformation on twitter about the relative harms of e-cigarette use: results from a randomized controlled trial. JMIR Public Health Surveill 2021 Dec 21;7(12):e27183 [FREE Full text] [doi: 10.2196/27183] [Medline: 34931999]
44. Chen L, Lu X, Yuan J, Luo J, Luo J, Xie Z, et al. A social media study on the associations of flavored electronic cigarettes with health symptoms: observational study. J Med Internet Res 2020 Jun 22;22(6):e17496 [FREE Full text] [doi: 10.2196/17496] [Medline: 32568093]
45. Struik L, Yang Y. e-Cigarette cessation: content analysis of a quit vaping community on Reddit. J Med Internet Res 2021 Oct 25;23(10):e28303 [FREE Full text] [doi: 10.2196/28303] [Medline: 34694229]
46. Priya S, Sequeira R, Chandra J, Dandapat SK. Where should one get news updates: Twitter or Reddit. Online Soc Netw Media 2019 Jan;9:17-29. [doi: 10.1016/j.osnem.2018.11.001]
47. Dulam RV, Murthy M, Luo J. Seeing through the smoke: a world-wide comparative study of e-cigarette flavors, brands and markets using data from Reddit and Twitter. arXiv 2020 Feb 4:1-7 [FREE Full text]
48. Kasson E, Singh AK, Huang M, Wu D, Cavazos-Rehg P. Using a mixed methods approach to identify public perception of vaping risks and overall health outcomes on Twitter during the 2019 EVALI outbreak. Int J Med Inform 2021 Nov;155:104574 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104574] [Medline: 34592539]
49. GetOldTweets3 0.0.11. Python Package Index. 2019. URL: https://pypi.org/project/GetOldTweets3/ [accessed 2022-01-21]
50. 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]
51. Pushshift Reddit API v4.0 Documentation. Pushshift. 2018. URL: https://reddit-api.readthedocs.io/en/latest/#searching-submissions [accessed 2021-10-01]
52. Cambria E, Das D, Bandyopadhyay S, Feraco A. A Practical Guide to Sentiment Analysis. Cham, Switzerland: Springer; 2017.
53. Huang M, ElTayeby O, Zolnoori M, Yao L. Public opinions toward diseases: infodemiological study on news media data. J Med Internet Res 2018 May 08;20(5):e10047 [FREE Full text] [doi: 10.2196/10047] [Medline: 29739741]
54. Hutto C, Gilbert E. Vader: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. 2014 May 16 Presented at: ICWSM '14; May 27-29, 2015; Ann Arbor, MI, USA p. 216-225. [doi: 10.1609/icwsm.v8i1.14550]

55. 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 Dec 20;23(12):e27307 [FREE Full text] [doi: 10.2196/27307] [Medline: 34932014]
56. Krawczyk K, Chelkowski T, Laydon DJ, Mishra S, Xifara D, Gibert B, et al. Quantifying online news media coverage of the COVID-19 pandemic: text mining study and resource. J Med Internet Res 2021 Jun 02;23(6):e28253 [FREE Full text] [doi: 10.2196/28253] [Medline: 33900934]
57. Qaiser S, Ali R. Text mining: use of TF-IDF to examine the relevance of words to documents. Int J Comput Appl 2018 Jul 16;181(1):25-29. [doi: 10.5120/ijca2018917395]
58. 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]
59. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
60. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. Int J Public Health 2016 Mar;61(2):249-256. [doi: 10.1007/s00038-016-0791-2] [Medline: 26841895]
61. Martinez LS, Hughes S, Walsh-Buhi ER, Tsou MH. "Okay, we get it. You vape": an analysis of geocoded content, context, and sentiment regarding e-cigarettes on Twitter. J Health Commun 2018;23(6):550-562. [doi: 10.1080/10810730.2018.1493057] [Medline: 29979920]
62. 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]
63. Miech R, Patrick ME, O'Malley PM, Johnston LD. What are kids vaping? Results from a national survey of US adolescents. Tob Control 2017 Jul;26(4):386-391 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053014] [Medline: 27562412]
64. Budney AJ, Sargent JD, Lee DC. Vaping cannabis (marijuana): parallel concerns to e-cigs? Addiction 2015 Nov;110(11):1699-1704 [FREE Full text] [doi: 10.1111/add.13036] [Medline: 26264448]
65. Sowles SJ, Krauss MJ, Connolly S, Cavazos-Rehg PA. A content analysis of vaping advertisements on Twitter, November 2014. Prev Chronic Dis 2016 Sep 29;13:E139 [FREE Full text] [doi: 10.5888/pcd13.160274] [Medline: 27685432]
66. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [doi: 10.11613/bm.2012.031] [Medline: 23092060]
67. Syed M, Nelson SC. Guidelines for establishing reliability when coding narrative data. Emerg Adulthood 2015 May 26;3(6):375-387. [doi: 10.1177/2167696815587648]
68. Lee JL, DeCamp M, Dredze M, Chisolm MS, Berger ZD. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on Twitter. J Med Internet Res 2014 Oct 15;16(10):e237 [FREE Full text] [doi: 10.2196/jmir.3765] [Medline: 25591063]
69. Orellana-Rodriguez C, Keane MT. Attention to news and its dissemination on Twitter: a survey. Comput Sci Rev 2018 Aug;29:74-94. [doi: 10.1016/j.cosrev.2018.07.001]
70. Wojcik J, Hughes A. Sizing up Twitter Users. Pew Research Center. 2019 Apr 24. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2022-03-22]
71. Duggan M, Smith A. 6% of Online Adults are reddit Users. Pew Research Center. 2013 Jul 3. URL: https://www.pewresearch.org/internet/2013/07/03/6-of-online-adults-are-reddit-users/ [accessed 2022-03-22]
72. Characteristics of Electronic Cigarette Use Among Middle and High School Students—United States, 2015. Centers for Disease Control and Prevention. 2016. URL: https://www.cdc.gov/tobacco/data_statistics/mmwrs/byyear/2016/mm655051a2/index.html [accessed 2022-03-22]
73. Assari S, Mistry R, Bazargan M. Race, educational attainment, and e-cigarette use. J Med Res Innov 2020;4(1):e000185 [FREE Full text] [doi: 10.32892/jmri.185] [Medline: 32090188]

Abbreviations

API: application program interface
CS: computer science
EVALI: e-cigarette and vaping use–associated lung injury
TF-IDF: term frequency–inverse document frequency
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by T Leung; submitted 11.05.22; peer-reviewed by T Deng, H Zhang, R Chew, N Ezike, W Ceron; comments to author 10.08.22; revised version received 16.09.22; accepted 29.10.22; published 13.12.22.

Please cite as:
Wu D, Kasson E, Singh AK, Ren Y, Kaiser N, Huang M, Cavazos-Rehg PA
Topics and Sentiment Surrounding Vaping on Twitter and Reddit During the 2019 e-Cigarette and Vaping Use–Associated Lung Injury Outbreak: Comparative Study
J Med Internet Res 2022;24(12):e39460
URL: https://www.jmir.org/2022/12/e39460
doi:10.2196/39460
PMID:36512403

Original Paper

Predicting Smoking Prevalence in Japan Using Search Volumes in an Internet Search Engine: Infodemiology Study

Kazuya Taira[1], PhD; Takahiro Itaya[2], DrPH; Sumio Fujita[3]

[1]Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Healthcare Epidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan

[3]Yahoo Japan Corporation, Tokyo, Japan

Corresponding Author:
Kazuya Taira, PhD
Department of Human Health Sciences
Graduate School of Medicine
Kyoto University
53, Shogoinkawara-cho
Sakyo-ku
Kyoto, 606-8507
Japan
Phone: 81 075 751 3927
Email: taira.kazuya.5m@kyoto-u.ac.jp

Abstract

Background: Tobacco smoking is an important public health issue and a core indicator of public health policy worldwide. However, global pandemics and natural disasters have prevented surveys from being conducted.

Objective: The purpose of this study was to predict smoking prevalence by prefecture and sex in Japan using Internet search trends.

Methods: This study used the infodemiology approach. The outcome variable was smoking prevalence by prefecture, obtained from national surveys. The predictor variables were the search volumes on Yahoo! Japan Search. We collected the search volumes for queries related to terms from the thesaurus of the Japanese medical article database Ichu-shi. Predictor variables were converted to per capita values and standardized as *z* scores. For smoking prevalence, the values for 2016 and 2019 were used, and for search volume, the values for the April 1 to March 31 fiscal year (FY) 1 year prior to the survey (ie, FY 2015 and FY 2018) were used. Partial correlation coefficients, adjusted for data year, were calculated between smoking prevalence and search volume, and a regression analysis using a generalized linear mixed model with random effects was conducted for each prefecture. Several models were tested, including a model that included all search queries, a variable reduction method, and one that excluded cigarette product names. The best model was selected with the Akaike information criterion corrected (AICC) for small sample size and the Bayesian information criterion (BIC). We compared the predicted and actual smoking prevalence in 2016 and 2019 based on the best model and predicted the smoking prevalence in 2022.

Results: The partial correlation coefficients for men showed that 9 search queries had significant correlations with smoking prevalence, including *cigarette* (r=–0.417, P<.001), *cigar* in kanji (r=–0.412, P<.001), and *cigar* in katakana (r=-0.399, P<.001). For women, five search queries had significant correlations, including *vape* (r=0.335, P=.001), *quitting smoking* (r=0.288, P=.005), and *cigar* (r=0.286, P=.006). The models with all search queries were the best models for both AICC and BIC scores. Scatter plots of actual and estimated smoking prevalence in 2016 and 2019 confirmed a relatively high degree of agreement. The average estimated smoking prevalence in 2022 in the 47 prefectures for the total sample was 23.492% (95% CI 21.617%-25.367%), showing an increasing trend, with an average of 29.024% (95% CI 27.218%-30.830%) for men and 8.793% (95% CI 7.531%-10.054%) for women.

Conclusions: This study suggests that the search volume of tobacco-related queries in internet search engines can predict smoking prevalence by prefecture and sex in Japan. These findings will enable the development of low-cost, timely, and crisis-resistant health indicators that will enable the evaluation of health measures and contribute to improved public health.

(J Med Internet Res 2022;24(12):e42619) doi:10.2196/42619

KEYWORDS

health policy; internet use; quality indicators; search engine; smoking; tobacco use; public health; infodemiology; smoking trend; health indicator; health promotion

Introduction

Smoking Prevalence as a Health Policy Indicator

Tobacco smoking is a cause of many types of cancer, respiratory disease, and coronary artery disease [1-4]. Since 2013, various electronic nicotine delivery systems (ENDS) and heat-not-burn tobacco products have been introduced as alternatives to cigarettes [5], with 3 brands available in Japan: Iqos, Glo, and Ploom Tech. These ENDS and heat-not-burn tobacco products have been reported to be potentially harmful and alerts have been issued [6,7].

Smoking prevalence is included as an indicator in the Health Japan 21 (the Second Term), a guideline for health measures in Japan, and policy measures are being implemented based on target values [8]. Currently, smoking prevalence is ascertained through national surveys, such as the Comprehensive Survey of Living Conditions [9] and the National Health and Nutrition Survey [10]. The Comprehensive Survey of Living Conditions is conducted every 3 years, and the National Health and Nutrition Survey is conducted every 5 years on a large scale. There are only a limited number of surveys, which require enormous effort and cost in the millions of dollars, that have a large enough sample size to be tabulated by prefecture. Furthermore, these surveys were not conducted in years with major natural disasters, such as the 2016 Kumamoto earthquake; moreover, after 2020, due to the COVID-19 pandemic, the Comprehensive Survey of Living Conditions in 2020 and the National Health and Nutrition Survey in 2020 and 2021 were cancelled. The discontinuation of these large-scale national surveys monitoring health indicators has prevented the evaluation of policies and hindered evidence-based policy making.

Trends in Smoking Prevalence in Japan

Smoking prevalence in Japan has decreased dramatically, from 82.3% in 1965 to 27.8% in 2018 for men over 20 years old and from 16.5% in 1965 to 8.7% in 2018 for women [11]. However, Japan has a slightly higher smoking prevalence than other high-income countries, indicating that tobacco control measures have not yet reached the level of best practice [12]. In addition, it has been pointed out that socioeconomic disparities may be a factor associated with Japan's persistent decline in smoking prevalence [13], which is an important public health issue.

Internet Search Engine and Tobacco Use

Significant correlations have been reported between internet search trends and the prevalence of tobacco and smokeless tobacco use by state in the United States [14]. Moreover, youth are consistently exposed to tobacco-related content on the internet [15], and exposure to tobacco-related content on social media has been reported to be a risk for smoking behavior [16]. Therefore, we hypothesized that it would be possible to predict smoking prevalence at the regional level based on search trends on the internet. In addition, it has been reported that internet search trends can track users' interest in ENDS [17] and heat-not-burn tobacco [18]. Thus, the purpose of this study was to predict smoking prevalence by prefecture and sex in Japan based on internet search trends.

Methods

This study used the infodemiology approach to monitoring smoking prevalence in Japan based on internet search engine trends.

Outcome Variable

The outcome variable was smoking prevalence. Smoking prevalence was obtained for each prefecture from the Comprehensive Survey of Living Conditions [9], a national survey conducted by the Japanese government. This is a national survey of households that are randomly selected and stratified by region from all over Japan. A simple survey (distributed to approximately 55,000 households, comprising 138,000 people) is conducted annually, and a large-scale survey (distributed to approximately 277,000 households, comprising 688,000 people) is conducted every 3 years. Because smoking prevalence is included only in the larger-scale survey, we obtained smoking prevalence for each prefecture in 3-year periods from 2001 to 2019 (Multimedia Appendix 1). However, in 2016, data from Kumamoto prefecture were missing due to a natural disaster caused by an earthquake. Since the volume of data was too small to impute missing values, and they could not be properly estimated, complete case analyses were conducted in this study.

Predictors

The predictor variable was search volumes on Yahoo! Japan Search, one of the largest search engines in Japan. This study was conducted in collaboration with Yahoo Japan Corporation, and the researchers were authorized to access search log data from Yahoo! Japan Search. The necessary data were extracted with a tabulation program that accessed the Yahoo Japan Corporation server via a virtual private network connection. The search queries for which search volumes were extracted were words that were listed as synonyms for "cigarette," "smoking," and "e-cigarette" in the thesaurus of Ichu-shi Web, the largest Japanese medical article search database. The monthly number of searches per prefecture for each search query was obtained, and the total number of searches per fiscal year (FY, running from April 1 to March 31) was calculated.

Because the Yahoo! Japan search log data were available starting for the year 2014, to predict smoking prevalence by search volume in the FYs before the FY in which we wanted to predict smoking prevalence, we obtained search volumes for FYs 2015, 2018, and 2021. These were all years during which the triennial large-scale survey of smoking prevalence was conducted, making retrospective data available. The reason for predicting smoking prevalence in FYs is that Japanese local governments evaluate their projects every FY. Queries that were not written in Japanese or that had a month in which they were never

searched for were excluded; 18 queries were thus used in the analysis (Textbox 1).

Textbox 1. Search queries used in the analysis, with original Japanese-language terms. The multiple entries for some terms reflect the multiple Japanese writing systems, including hiragana, katakana, kanji, and the Latin alphabet.

Search terms (with Japanese text)

- *Cigarette* in katakana (シガレット)
- *Tobacco* in hiragana (たばこ), *tobacco* in katakana (タバコ), *tobacco* in kanji (煙草)
- *Cigar* in kanji (葉巻), *cigar* in katakana (シガー)
- *Glo* (glo), *glo* in katakana (グロー)
- *Vape* (vape)
- *Iqos* (iqos), *iqos* in katakana (アイコス)
- *Ploom Tech* in katakana (プルームテック)
- *Electronic cigarette* in katakana (電子タバコ), *electronic cigarette* in hiragana (電子たばこ)
- *Heat-not-burn tobacco* in hiragana (加熱式たばこ), *heat-not-burn tobacco* in katakana (加熱式タバコ)
- *Smoking* in kanji (喫煙)
- *Quitting smoking* in kanji (禁煙)

Standardization of Predictors

Since the male to female ratio for each search query can be obtained based on registration information from user IDs, search volumes by sex were calculated by prorating the search volumes by sex. Since search volumes are affected by the size of the prefectural population, the search volumes per capita were calculated by dividing the total or male/female prefectural population and then converting the results to a *z* score to standardize the results. The *z* scores were calculated with the following formula ("query A" refers to each search query used in the analysis):

The value for the prefectural populations used to calculate the search volumes per capita was the value on October 1 (the median day of the FY) of the respective year, obtained from open government data on population estimate statistics [19].

Statistical Analysis

Partial Pearson correlation coefficients, adjusted for data year, between smoking prevalence per prefecture and the *z* scores of the search volumes for each query for each prefecture were calculated. Regression analysis using a generalized linear mixed model (GLMM) was conducted with smoking prevalence as the outcome variable, *z* scores for each search query and survey year as predictor variables, and random effects for each prefecture. For the regression analysis, we used data from 2016 and 2019 for both smoking prevalence and number of searches. In the regression analyses, model 0 included only the survey year and intercept as predictors; model 1 included the survey year, intercept, and queries, but excluded the names of tobacco-related products; model 2 included the survey year, intercept, and queries selected by the backward selection method; and model 3 included the survey year, intercept, and all search queries in the total sample for men and women separately. The selection of the best model was determined using the Akaike information criterion corrected (AICC) for small sample size and the Bayesian information criterion (BIC). The AICC and BIC are model fit indices that focus on prediction accuracy, and both refer to better fit with smaller values relative to other models. The analyses were conducted using SPSS Statistics (version 27; IBM). The significance level was set at 1% for all analyses.

Evaluation of the GLMM Model and Prediction of Smoking Prevalence in 2022

Search volumes were substituted into the selected best model to predict smoking prevalence for 2016, 2019, and 2022. For 2016 and 2019, scatter plots of actual and estimated smoking prevalence were drawn to confirm agreement. The smoking prevalence in 2022 was plotted as a line graph along with actual values from 2001 to 2019 to confirm the trend.

Ethical Considerations

This study involved secondary analysis of public statistics and anonymized existing data; therefore, ethical review was waived by the Kyoto University Graduate School and Faculty of Medicine Ethics Committee.

Results

Correlation Coefficients Between Smoking Prevalence and Each Search Query

The results of the partial correlation analyses are presented in Table 1. For men, the following 9 search queries had significant correlations with smoking prevalence: *cigar* in katakana (シガー; r=–0.399, P<.001), *cigarette* in katakana (シガレット; r=–0.417, P<.001), *tobacco* in katakana (タバコ; r=–0.388, P<.001), *tobacco* in hiragana (たばこ; r=–0.334, P=.001), *tobacco* in kanji (煙草; r=–0.370, P<.001), *smoking* in kanji (喫煙; r=–0.346, P=.001), *electronic cigarette* in katakana (電子タバコ; r=–0.303, P=.003), *electronic cigarette* in hiragana (電子たばこ; r=–0.271, P=.009), and *cigar* in kanji (葉巻;

r=–0.412, *P*<.001). For women, the following 5 search queries had significant correlations: *vape*, (*r*=0.335, *P*=.001), *cigar* in katakana (シガー; *r*=0.286, *P*=.006), *quitting smoking* in kanji (禁煙; *r*=0.288, *P*=.005), *electronic cigarette* in katakana (電子タバコ; *r*=0.271, *P*=.009), and *electronic cigarette* in hiragana (電子たばこ; *r*=0.271, *P*=.009) (Table 1).

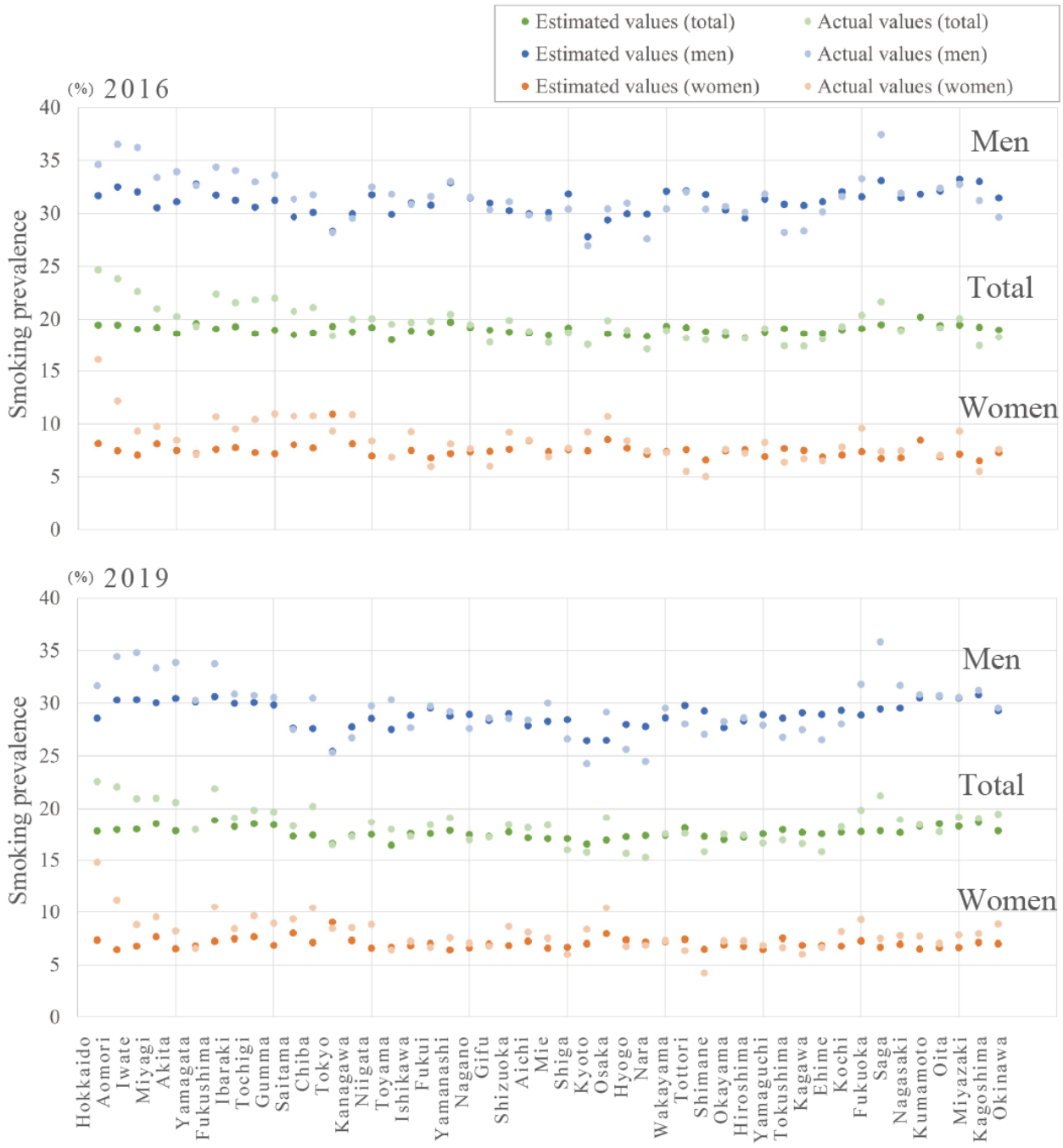
Table 1. Partial Pearson correlation coefficient (r) between prefectural smoking prevalence and the search volumes of each query by sex.

Search term	Total		Men		Women	
	r	*P* value	*r*	*P* value	*r*	*P* value
Glo	0.02	*P*=.89	–0.16	*P*=.14	0.22	*P*=.04
Iqos	–0.01	*P*=.96	–0.20	*P*=.06	0.18	*P*=.08
Vape	0.01	*P*=.89	–0.25	*P*=.02	0.34	*P*=.001
Iqos (アイコス)	0.01	*P*=.92	–0.21	*P*=.04	0.26	*P*=.01
Glo (グロー)	0.03	*P*=.77	–0.09	*P*=.39	0.19	*P*=.07
Cigar (シガー)	–0.12	*P*=.25	–0.40	*P*<.001	0.29	*P*=.006
Cigarette (シガレット)	–0.20	*P*=.06	–0.42	*P*<.001	0.18	*P*=.09
Tobacco (タバコ)	–0.11	*P*=.31	–0.39	*P*<.001	0.25	*P*=.02
Tobacco (たばこ)	–0.10	*P*=.36	–0.33	*P*=.001	0.23	*P*=.03
Ploomtech (プルームテック)	0.001	*P*=.99	–0.16	*P*=.14	0.20	*P*=.06
Tobacco (煙草)	–0.11	*P*=.30	–0.37	*P*<.001	0.24	*P*=.02
Heat-not-burn tobacco (加熱式タバコ)	–0.07	*P*=.46	–0.23	*P*=.03	0.16	*P*=.13
Heat-not-burn tobacco (加熱式たばこ)	0.01	*P*=.93	–0.15	*P*=.14	0.20	*P*=.06
Smoking (喫煙)	–0.11	*P*=.31	–0.35	*P*=.001	0.19	*P*=.07
Quitting smoking (禁煙)	0.02	*P*=.85	–0.21	*P*=.05	0.29	*P*=.005
Electronic cigarette (電子タバコ)	–0.05	*P*=.67	–0.30	*P*=.003	0.27	*P*=.009
Electronic cigarette (電子たばこ)	–0.03	*P*=.79	–0.27	*P*=.009	0.27	*P*=.009
Cigar (葉巻)	–0.15	*P*=.16	–0.41	*P*<.001	0.21	*P*=.046

Results of the Regression Analyses and Evaluation of the Best Models

Model 3 was the best model for both AICC and BIC scores in all regression analyses for the total sample (AICC 308.043 and BIC 312.452), men (AICC 338.656 and BIC 343.066), and women (AICC 302.225 and BIC 306.635). Details are provided in Tables S1, S2, and S3 in Multimedia Appendix 2. Moreover, no search queries with significant regression coefficients were found in model 3 for either sex or the total sample. Scatter plots of actual and estimated smoking prevalence in 2016 and 2019 confirmed a relatively high degree of agreement (Figure 1).

Figure 1. Comparison of actual and modeled estimates of smoking prevalence by prefecture in 2016 and 2019.



Estimates of Smoking Prevalence by Prefecture in 2022

The average estimated smoking prevalence in 2022 for the 47 prefectures for the total sample, including men and women (Figure 2), was 23.492% (95% CI 21.617%-25.367%), showing an increasing trend, with a prevalence of 29.024% (95% CI 27.218%-30.830%) for men (Figure 3) and 8.793% (95% CI 7.531%-10.054%) for women (Figure 4).

Figure 2. Total trends in smoking prevalence and predicted smoking prevalence in 2022.

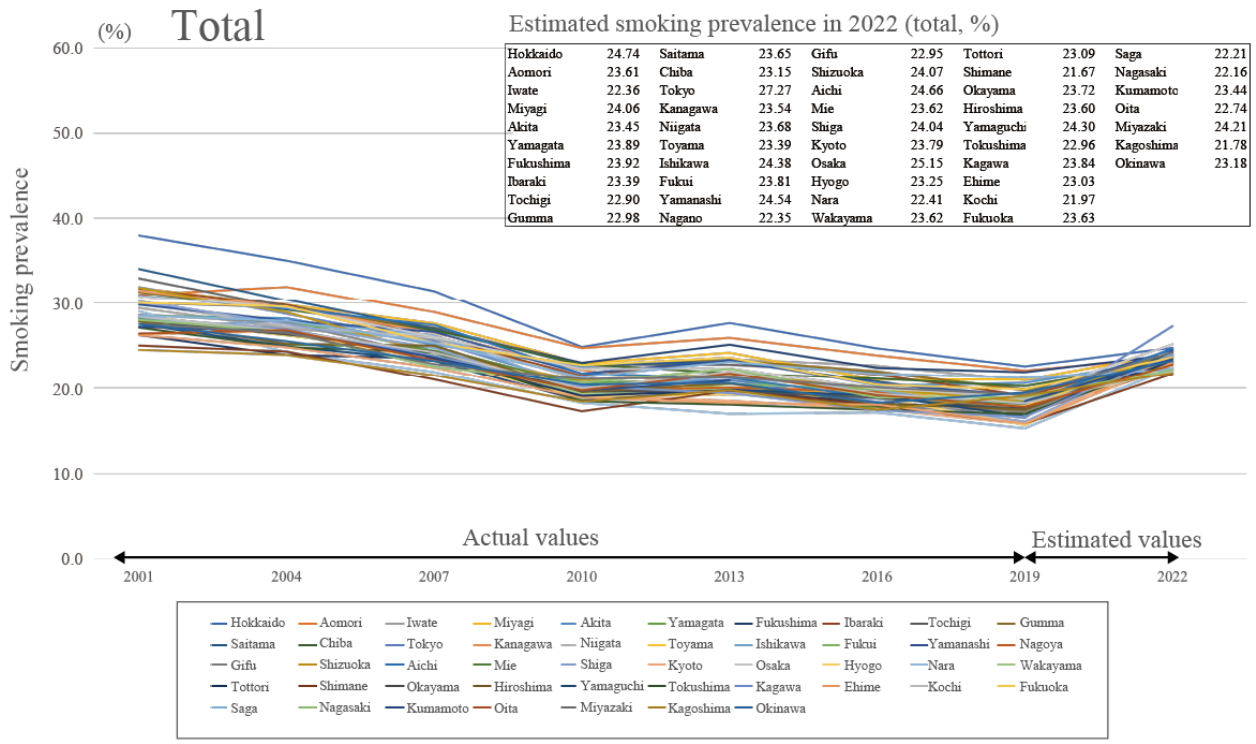


Figure 3. Trends in smoking prevalence and predicted smoking prevalence in 2022 for men.

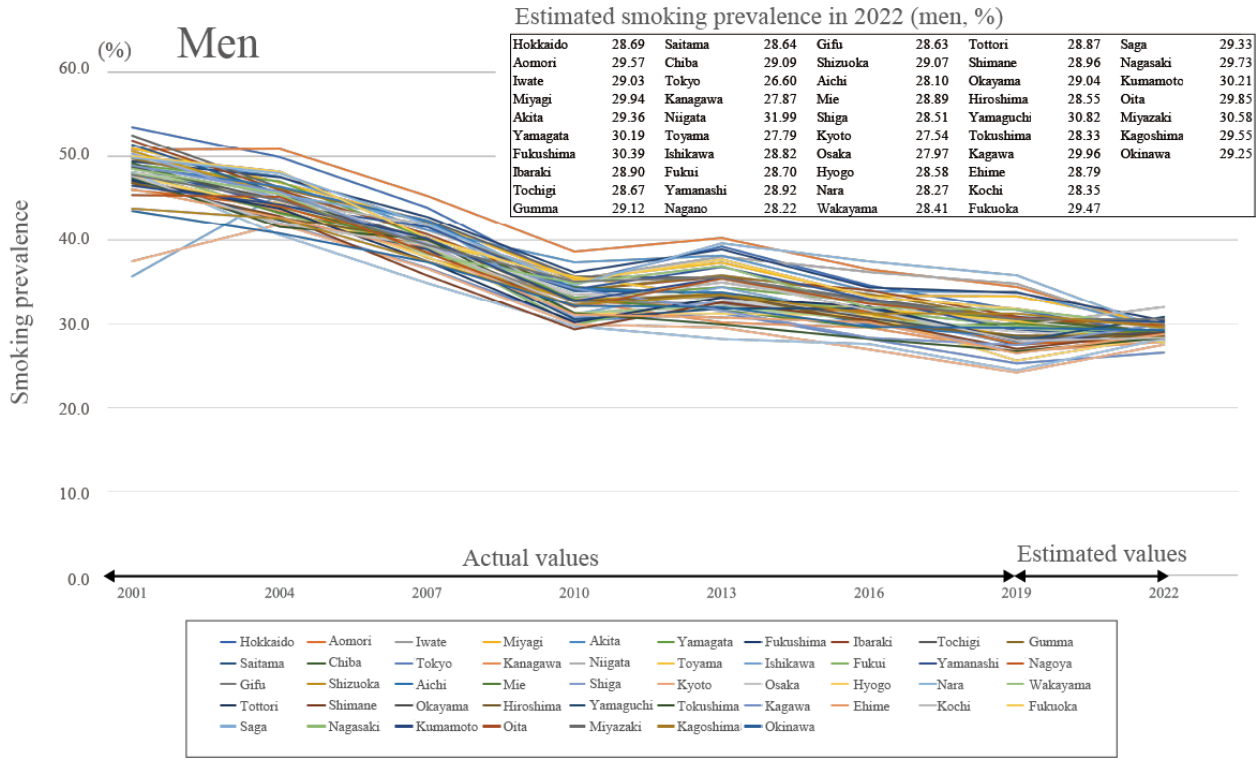
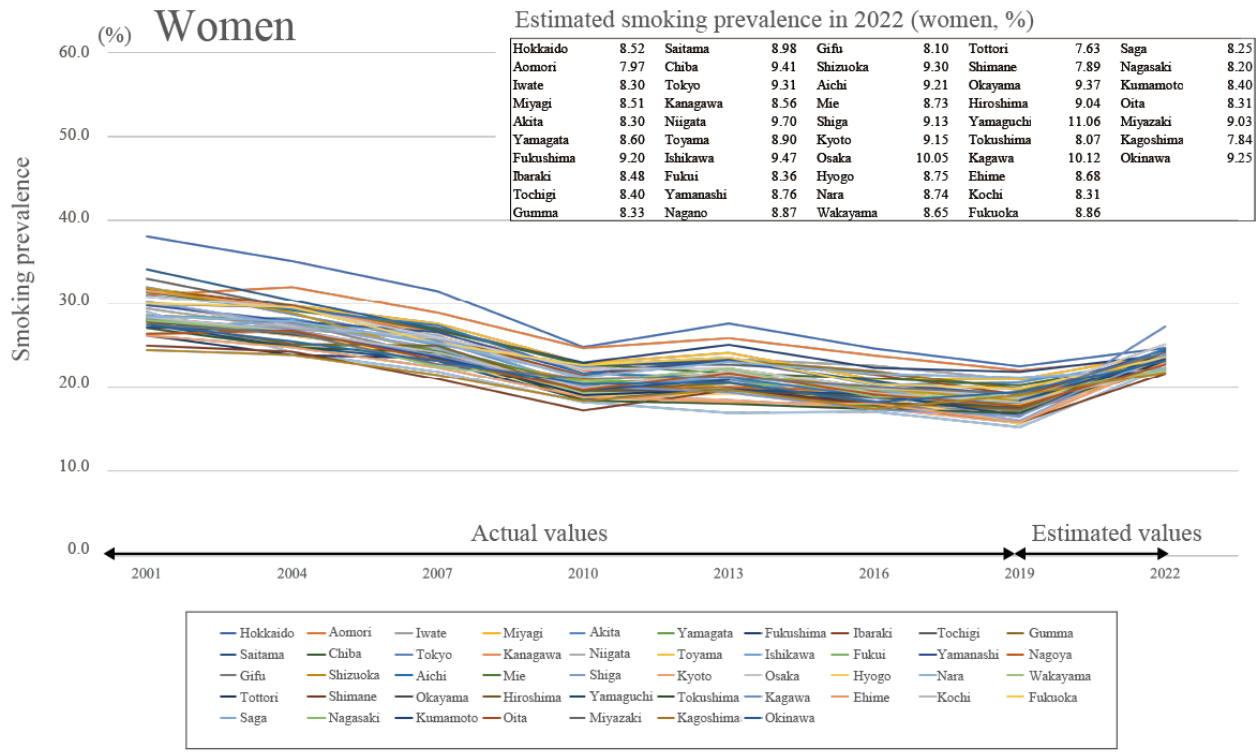


Figure 4. Trends in smoking prevalence and predicted smoking prevalence in 2022 for women.



Discussion

Principal Results

We found that smoking prevalence could be predicted with a moderate degree of accuracy using a GLMM based on search volume for tobacco-related queries in internet search engines. A univariate analysis of the partial correlation coefficients between smoking prevalence by prefecture and each search query revealed significant variables only for sex, but the GLMM predictions had approximately the same predictive accuracy for the total sample as for men and women. The GLMM models were judged from the AICC and BIC scores, and the models that included all search queries as predictors were adopted as the best models for all sexes and the total sample. These results suggest that rather than specific words being strong predictors of smoking prevalence, it is more likely that different words play a role in revealing prefectural-specific characteristics of monotonic smoking prevalence trends.

When actual and estimated smoking prevalence were compared, the estimated values tended to show less variation, and this may reduce the accuracy of predicting prefectures with smoking prevalence that deviates from the mean. However, this allows us to follow trends in smoking prevalence for Japan overall. Smoking prevalence has continued to show a consistent decreasing trend from 2001 to 2019, but estimates for 2022 show a flat trend or an increasing trend among women. After 2020, the aftereffects of the COVID-19 pandemic have caused behavioral restrictions and stagnating economic activity around the world, and Japan is in the midst of the seventh wave as of 2022, with the highest number of cases in the world [20]. This could be one contributor to the trends seen in this study. Economic disparity has also been identified as a contributing factor to the still-high prevalence of smoking in Japan [13]. Therefore, economic deprivation attributed to COVID-19 may be related to the increase in the estimated smoking prevalence in 2022.

More interestingly, the correlation coefficients between the search volume of tobacco-related queries and smoking prevalence were negative for men, positive for women, and uncorrelated in the total sample. Smoking prevalence in Japan has traditionally differed significantly between men and women, which could lead to relatively higher search volumes for smoking cessation behaviors in men and higher search volumes for smoking behaviors in women. Notably, the COVID-19 pandemic has disproportionately affected women's employment, a phenomenon called the "she-cession" [21]. Several studies have also reported an increase in suicide and mental health problems among women in Japan during the COVID-19 pandemic [22,23]. Therefore, it is possible that smoking prevalence among women increased as a coping behavior for stress [24].

Hence, when considering health policies based on the search volume of tobacco-related queries in internet search engines, men and women require different approaches. It would be desirable for future studies to not only predict smoking prevalence, but also investigate smokers' search behavior patterns and gender differences in more detail, clarifying the relationship between smokers' actual search behavior and smoking prevalence.

Limitations

One of the limitations of our study relates to the flexibility of the model, as both smoking prevalence and tobacco-related query retrieval volume statistics were only available for 2 years

(2016 and 2019). In particular, for Kumamoto prefecture, data were missing due to a major earthquake in 2016, so care must be taken in interpreting the estimates. The lack of statistically significant variables in the GLMM regression analysis may also be due to the small sample size and multicollinearity in the highly correlated relationships among search volumes for the tobacco-related queries. This study did not address this issue, because it focused on the accuracy of predicting smoking prevalence as an outcome. However, if these limitations can be resolved in future analyses using more sophisticated statistical methods, we could obtain insights into factors that influence changes in smoking prevalence. In addition, internet usage in Japan differs by age group: more than 90% of people between their teenage years and the sixth decade of life use the internet, while 74.2% of those in their sixties and 57.5% of those in their eighties use the internet [25], suggesting that the influence of the elderly may have been underestimated. However, as of 2019, the highest smoking prevalence in men was among those in their forties, at 36.5%, decreasing as people reached their sixties; the prevalence was 31.1% among those in their sixties and 15.1% among those aged 70 or older. The same trend was observed in women, with the highest smoking prevalence being 12.9% among those in their fifties, compared to 8.6% of those in their sixties and 3% in women aged 70 and older. Therefore, we believe that the impact of lower internet use among the elderly was limited. Finally, although this study used separate analyses for men and women, it is possible that gender bias was present in the internet search behavior. One previous study in Australia [26] suggests that female smokers who are highly socially disadvantaged seem to use the internet more frequently; however, there is no evidence of gender differences in internet search behavior in the general population in Japan. Research on gender bias in smokers' internet search behavior is also an issue for future study.

Comparison With Prior Work

Most previous studies in the field of infoveillance have focused on predicting the prevalence of infectious diseases, such as influenza [27,28]. This study suggests that smoking prevalence can also be predicted with high accuracy by the search volume of tobacco-related queries. Many previous studies using search engines have used Google Trends data, which does not provide search volume by sex or prefecture, unlike the data from Yahoo searches, making the latter more appropriate to conduct our research. Although we found no studies that predicted smoking prevalence by internet search volume, a moderate correlation between smoking prevalence and search volumes for tobacco-related queries was reported in the United States [14]. In this study, the same trend was observed for Japanese men, with a slightly weaker correlation for Japanese women. The US smoking prevalence in 2020 was 14.1% for men and 11% for women [29], with no significant difference. However, Japan has a very large gap in smoking prevalence between men and women, which may be the reason for the slightly weaker correlation for women.

Our findings may be useful in the evaluation of public health measures when large-scale nationwide surveys are not possible due to epidemics or natural disasters such as major earthquakes. In fact, Japan has a history of missing statistics, including statistics related to COVID-19 and the Kumamoto earthquake, and has a high probability of large-scale earthquakes and volcanic eruptions in the future.

Conclusions

This study suggests that internet search volume for tobacco-related queries can predict smoking prevalence by prefecture. Our findings may facilitate the development of low-cost, timely, and crisis-resistant health indicators that will enable the evaluation of health measures and contribute to improved public health.

Acknowledgments

We would like to thank Editage for English language editing. This work was supported by the Japan Society for the Promotion of Science KAKENHI (grant 22K17549).

Conflicts of Interest

SF is an employee of Yahoo Japan Corporation, a Japanese internet service company that provides the Yahoo! Japan search services analyzed in this paper. KT and TI declare no conflicts of interest.

Multimedia Appendix 1
Smoking prevalence by prefecture, 2001-2019.
[DOCX File , 50 KB - jmir_v24i12e42619_app1.docx]

Multimedia Appendix 2
Generalized linear mixed models for both sexes, men, and women.
[DOCX File , 52 KB - jmir_v24i12e42619_app2.docx]

References

1. Smoking cessation: a report of the Surgeon General. US Department of Health and Human Services. URL: https://www.hhs.gov/surgeongeneral/reports-and-publications/tobacco/index.html [accessed 2022-12-05]

2. Human cancer: known causes and prevention by organ site. International Agency for Research on Cancer. URL: https://handbooks.iarc.fr/publications/presentations/ [accessed 2022-12-05]
3. Smoking and health report of the Study Group on the Health Effects of Smoking [Japanese]. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/shingi2/0000135586.html [accessed 2022-12-05]
4. GBD 2019 Cancer Risk Factors Collaborators. The global burden of cancer attributable to risk factors, 2010-19: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2022 Aug 20;400(10352):563-591 [FREE Full text] [doi: 10.1016/S0140-6736(22)01438-6] [Medline: 35988567]
5. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul;23 Suppl 3(Suppl 3):iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]
6. The Japanese Respiratory Society's Views and Recommendations on Heated Cigarettes and E-Cigarettes (Revised-12-11) [Japanese]. Japanese Respiratory Society. 2019. URL: https://www.jrs.or.jp/information/file/hikanetsu_kenkai_kaitei.pdf [accessed 2022-12-05]
7. Auer R, Concha-Lozano N, Jacot-Sadowski I, Cornuz J, Berthet A. Heat-not-burn tobacco cigarettes: smoke by any other name. JAMA Intern Med 2017 Jul 01;177(7):1050-1052 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1419] [Medline: 28531246]
8. Health Japan 21 (the second term). National Institute of Health and Nutrition. URL: https://www.nibiohn.go.jp/eiken/kenkounippon21/en/kenkounippon21/ [accessed 2022-12-05]
9. Comprehensive Survey of Living Conditions. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/english/database/db-hss/cslc.html [accessed 2022-12-05]
10. National Health and Nutrition Survey. National Institute of Health and Nutrition. URL: https://www.nibiohn.go.jp/eiken/english/research/project_nhns.html [accessed 2022-12-05]
11. Adult Smoking Rate (JT National Smoker Rate Survey) [Japanese]. Japan Health Promotion and Fitness Foundation. URL: https://www.health-net.or.jp/tobacco/statistics/jt.html [accessed 2022-12-05]
12. WHO report on the global tobacco epidemic 2021: addressing new and emerging products. World Health Organization. URL: https://www.who.int/publications/i/item/9789240032095 [accessed 2022-12-05]
13. Tanaka H, Mackenbach JP, Kobayashi Y. Widening socioeconomic inequalities in smoking in Japan, 2001-2016. J Epidemiol 2021 Jun 05;31(6):369-377. [doi: 10.2188/jea.JE20200025] [Medline: 32595181]
14. Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Lowery A, Grucza RA, Chaloupka FJ, et al. Monitoring of non-cigarette tobacco use using Google Trends. Tob Control 2015 May;24(3):249-255. [doi: 10.1136/tobaccocontrol-2013-051276] [Medline: 24500269]
15. Jenssen BP, Klein JD, Salazar LF, Daluga NA, DiClemente RJ. Exposure to tobacco on the internet: content analysis of adolescents' internet use. Pediatrics 2009 Aug;124(2):e180-e186 [FREE Full text] [doi: 10.1542/peds.2008-3838] [Medline: 19620193]
16. Donaldson SI, Dormanesh A, Perez C, Majmundar A, Allem J. Association between exposure to tobacco content on social media and tobacco use: a systematic review and meta-analysis. JAMA Pediatr 2022 Sep 01;176(9):878-885 [FREE Full text] [doi: 10.1001/jamapediatrics.2022.2223] [Medline: 35816331]
17. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. Am J Prev Med 2011 Apr;40(4):448-453. [doi: 10.1016/j.amepre.2010.12.007] [Medline: 21406279]
18. Caputi TL, Leas E, Dredze M, Cohen JE, Ayers JW. They're heating up: Internet search query trends reveal significant public interest in heat-not-burn tobacco products. PLoS One 2017;12(10):e0185735 [FREE Full text] [doi: 10.1371/journal.pone.0185735] [Medline: 29020019]
19. Japan Statistical Yearbook. Statistics Bureau of Japan. URL: https://www.stat.go.jp/english/data/nenkan/index.html [accessed 2022-12-05]
20. Situation Report. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/covid-19/kokunainohasseijoukyou_00006.html [accessed 2022-12-05]
21. Bluedorn J, Caselli F, Hansen N, Shibata I, Mendes TM. Gender and Employment in the COVID-19 Recession: Evidence on “She-cessions”. International Monetary Fund. 2021. URL: https://www.elibrary.imf.org/view/journals/001/2021/095/001.2021.issue-095-en.xml [accessed 2022-12-05]
22. Ueda M, Nordström R, Matsubayashi T. Suicide and mental health during the COVID-19 pandemic in Japan. J Public Health (Oxf) 2022 Aug 25;44(3):541-548 [FREE Full text] [doi: 10.1093/pubmed/fdab113] [Medline: 33855451]
23. Nomura S, Kawashima T, Harada N, Yoneoka D, Tanoue Y, Eguchi A, et al. Trends in suicide in Japan by gender during the COVID-19 pandemic, through December 2020. Psychiatry Res 2021 Jun;300:113913 [FREE Full text] [doi: 10.1016/j.psychres.2021.113913] [Medline: 33839422]
24. Dugan S, Lloyd B, Lucas K. Stress and coping as determinants of adolescent smoking behavior. J Appl Social Pyschol 1999 Apr;29(4):870-886. [doi: 10.1111/j.1559-1816.1999.tb02030.x]
25. Communications Usage Trend Survey. Ministry of Internal Affairs and Communications. URL: https://www.soumu.go.jp/johotsusintokei/english/ [accessed 2022-12-05]

26. McCrabb S, Twyman L, Palazzi K, Guillaumier A, Paul C, Bonevski B. A cross sectional survey of internet use among a highly socially disadvantaged population of tobacco smokers. Addict Sci Clin Pract 2019 Oct 15;14(1):38 [FREE Full text] [doi: 10.1186/s13722-019-0168-y] [Medline: 31610808]
27. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 01;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
28. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
29. Current cigarette smoking among adults in the United States. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm#:~:text=In%202020%2C%20nearly%2013%20of,with%20a%20smoking%2Drelated%20disease [accessed 2022-12-05]

Abbreviations

AICC: Akaike information criterion corrected for small sample size
BIC: Bayesian information criterion
ENDS: electronic nicotine delivery system
FY: fiscal year
GLMM: generalized linear mixed model

Edited by G Eysenbach; submitted 14.09.22; peer-reviewed by A Zink, A Fittler; comments to author 26.10.22; revised version received 05.11.22; accepted 30.11.22; published 14.12.22.

Please cite as:
Taira K, Itaya T, Fujita S
Predicting Smoking Prevalence in Japan Using Search Volumes in an Internet Search Engine: Infodemiology Study
J Med Internet Res 2022;24(12):e42619
URL: https://www.jmir.org/2022/12/e42619
doi:10.2196/42619
PMID:36515993

Original Paper

Characterizing the Prevalence of Obesity Misinformation, Factual Content, Stigma, and Positivity on the Social Media Platform Reddit Between 2011 and 2019: Infodemiology Study

Catherine C Pollack[1,2], PhD; Jennifer A Emond[1,3], PhD; A James O'Malley[1,4], PhD; Anna Byrd[2*]; Peter Green[2*], BA; Katherine E Miller[2*], BA; Soroush Vosoughi[5], PhD; Diane Gilbert-Diamond[2,3,6], ScD; Tracy Onega[7], PhD

[1]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
[2]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
[3]Department of Pediatrics, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
[4]The Dartmouth Institute for Health Policy and Clinical Practice, Hanover, NH, United States
[5]Department of Computer Science, Dartmouth College, Hanover, NH, United States
[6]Department of Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States
[7]Department of Population Health Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, United States
[*]these authors contributed equally

Corresponding Author:
Catherine C Pollack, PhD
Department of Biomedical Data Science
Geisel School of Medicine at Dartmouth
Rubin 833
1 Medical Center Drive
Lebanon, NH, 03756
United States
Phone: 1 540 497 3419
Email: Catherine.c.pollack.gr@dartmouth.edu

Abstract

Background: Reddit is a popular social media platform that has faced scrutiny for inflammatory language against those with obesity, yet there has been no comprehensive analysis of its obesity-related content.

Objective: We aimed to quantify the presence of 4 types of obesity-related content on Reddit (misinformation, facts, stigma, and positivity) and identify psycholinguistic features that may be enriched within each one.

Methods: All sentences (N=764,179) containing “obese” or “obesity” from top-level comments (n=689,447) made on non–age-restricted subreddits (ie, smaller communities within Reddit) between 2011 and 2019 that contained one of a series of keywords were evaluated. Four types of common natural language processing features were extracted: bigram term frequency–inverse document frequency, word embeddings derived from Bidirectional Encoder Representations from Transformers, sentiment from the Valence Aware Dictionary for Sentiment Reasoning, and psycholinguistic features from the Linguistic Inquiry and Word Count Program. These features were used to train an Extreme Gradient Boosting machine learning classifier to label each sentence as 1 of the 4 content categories or other. Two-part hurdle models for semicontinuous data (which use logistic regression to assess the odds of a 0 result and linear regression for continuous data) were used to evaluate whether select psycholinguistic features presented differently in misinformation (compared with facts) or stigma (compared with positivity).

Results: After removing ambiguous sentences, 0.47% (3610/764,179) of the sentences were labeled as misinformation, 1.88% (14,366/764,179) were labeled as stigma, 1.94% (14,799/764,179) were labeled as positivity, and 8.93% (68,276/764,179) were labeled as facts. Each category had markers that distinguished it from other categories within the data as well as an external corpus. For example, misinformation had a higher average percent of negations (β=3.71, 95% CI 3.53-3.90; *P*<.001) but a lower average number of words >6 letters (β=−1.47, 95% CI −1.85 to −1.10; *P*<.001) relative to facts. Stigma had a higher proportion of swear words (β=1.83, 95% CI 1.62-2.04; *P*<.001) but a lower proportion of first-person singular pronouns (β=−5.30, 95% CI −5.44 to −5.16; *P*<.001) relative to positivity.

Conclusions: There are distinct psycholinguistic properties between types of obesity-related content on Reddit that can be leveraged to rapidly identify deleterious content with minimal human intervention and provide insights into how the Reddit population perceives patients with obesity. Future work should assess whether these properties are shared across languages and other social media platforms.

(J Med Internet Res 2022;24(12):e36729) doi:10.2196/36729

KEYWORDS

obesity; misinformation; social stigma; social media; Reddit; natural language processing

Introduction

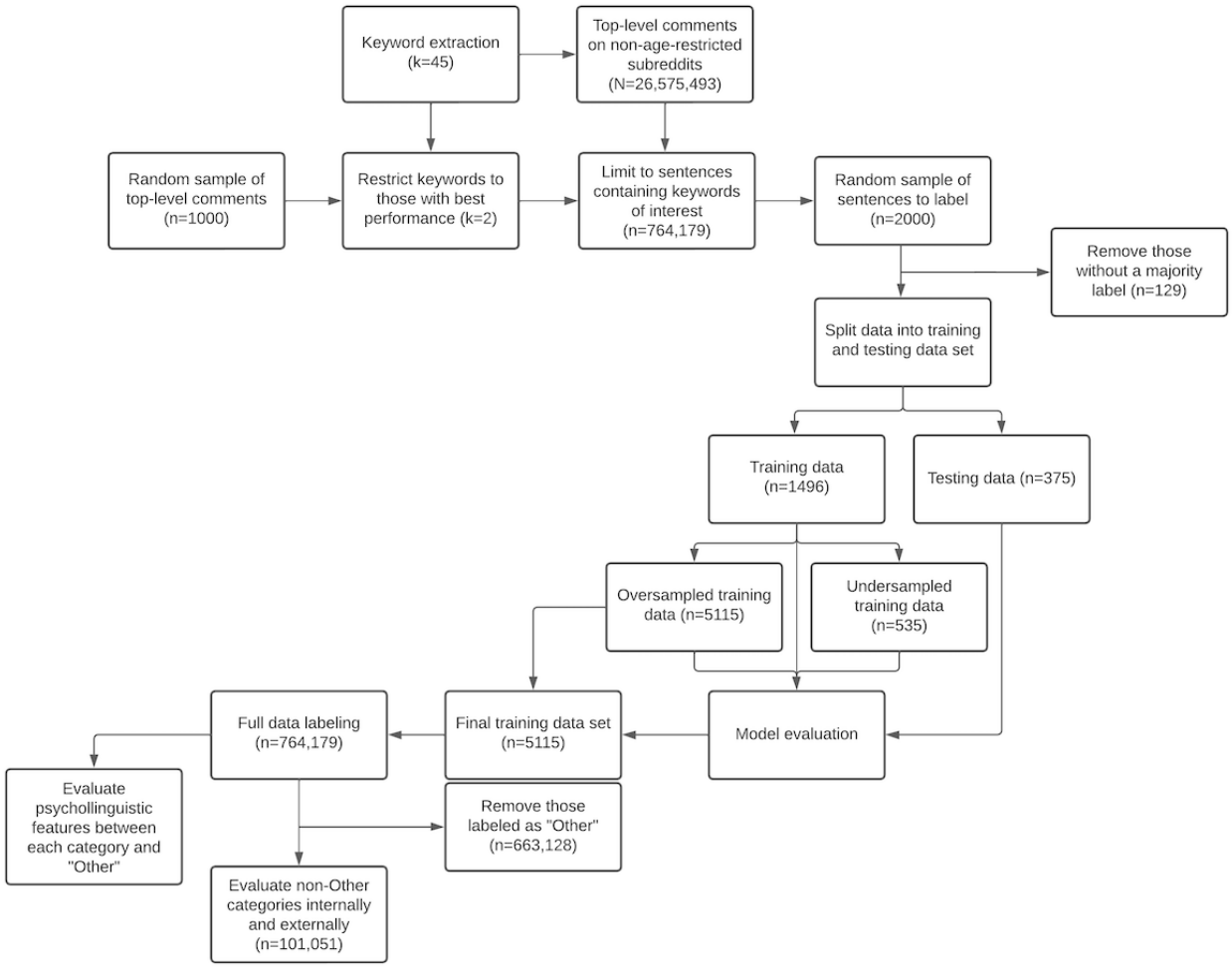
Social media has become a ubiquitous component of everyday life. A recent study suggested that 72% of Americans use social media, including 84% of those aged between 18 and 29 years, 81% of those aged between 30 and 49 years, and 73% of those aged between 50 and 64 years [1]. Although social media promises to foster meaningful connections between individuals around the world, it has been exploited to spread misinformation and disinformation on a variety of topics, from the 2016 presidential election to the COVID-19 pandemic [2,3]. The widespread use of social media presents an ideal medium to study the discourse surrounding these geopolitical and health topics, as well as other topics of concern. However, much of this work has focused on international infectious disease outbreaks and vaccines, with less research dedicated to understanding misinformation regarding chronic diseases such as obesity [4].

Over 42% of adults in the United States have obesity, with rates steadily increasing since the early 2000s [5]. Individuals with obesity may experience weight-related stigma, a phenomenon whereby individuals are ascribed negative traits (such as laziness) "due" to their weight [6-8]. Exposure to stigma has been associated with adverse physical, mental, and emotional health outcomes, and this effect is amplified by social media; for example, studies have shown that social media use is correlated with concerns over body image [9,10]. Social media–based studies on stigma have predominately focused on Facebook, Twitter, and Instagram, leaving a gap in understanding how stigma (and other deleterious content, such as misinformation) manifests on other platforms [11-13]. Furthermore, the aforementioned platforms are not typically used anonymously, which may differentially affect behaviors. In contrast, Reddit is an anonymized content aggregation platform with over 52 million daily active users as of January 2021 [14]. The platform consists of over 100,000 different "subreddits," which are smaller communities that are themed around a given topic. Subreddits range from more general communities for news and science to highly specific subreddits such as r/bodyweightfitness (a subreddit dedicated to sharing workout routines with 2.3 million subscribers as of October 2021) or r/eatcheapandhealthy (a subreddit dedicated to how to eat healthy foods on a budget with 3.6 million subscribers as of October 2021) [15]. Users can engage with these subreddits either by creating a post or by commenting on another user's post. These comments can also be commented upon, creating a branching comment section.

Previous research has leveraged content on specific subreddits to model predictors of successful weight loss [16,17]. These communities (such as r/loseit, which has 3 million subscribers as of October 2021) are designed to provide a space for individuals to seek motivation, to ask questions about weight loss, or to share their experiences. Other parts of the platform are not as supportive, and Reddit has historically faced scrutiny for the abundance of stigma on the platform. Most notable were concerns over the former subreddit r/fatpeoplehate, which was dedicated to ridiculing individuals with obesity. This subreddit was banned in 2015 during a push by the company to remove hate speech from the platform. An evaluation of this ban found that it was effective in reducing the amount of hate speech on Reddit that was directed toward individuals with obesity, including those who were previously active members of the banned subreddit [18]. However, this study was limited in scope, and there are no studies that comprehensively evaluate the presence of weight-related stigma or misinformation on the platform. Similarly, there are no studies that evaluate the presence of body positivity or factual content across the entire platform.

The purpose of this study was to characterize obesity-related content on Reddit. To do this, a semiautonomous pipeline was created that leveraged a set of psycholinguistic and semantic features to differentiate the 4 categories of interest: misinformation, factual content, stigma, and positivity (Figure 1). Briefly, this pipeline involved extracting a series of sentences containing "obese" or "obesity" from a broader pool of comments; manually assigning a category to a small, random subset; extracting a candidate set of features; identifying the best model and feature set based on performance on the labeled subset; and using the identified model and feature set to automatically label the entire set of obesity-specific data. Following this, a statistical analysis was performed to evaluate whether there were distinct types of features that were either enriched or underrepresented within each category of interest. All categories were compared against an external benchmark and sentences labeled as other, and each "pair" of categories (ie, fact vs misinformation and positivity vs stigma) was internally compared.

Figure 1. Flow diagram of process described in the Methods section. k: keyword.



Methods

Keyword Generation, Evaluation, and Data Collection

Keywords were derived from “obesity” and “diet” via colloquial terminology (eg, “chubby”), diet trends (eg, “keto”), and the National Institutes of Health National Library of Medicine Medical Subject Headings database (Multimedia Appendix 1) [19]. This process generated an initial list of 45 keywords that was used to curate all top-level Reddit comments (ie, comments that were made directly on a post and not another user’s comment) via Pushshift, a Reddit archive updated monthly [20]. Posts were excluded because they often contain video or image content that could not be evaluated in this study. Comments were limited to those made on non–age-restricted subreddits between 2011 and 2019, resulting in an initial corpus of 26,575,493 comments.

To evaluate keyword selection, a separate set of sentences from 1000 top-level Reddit comments were randomly collected and labeled by 1 member of the research team as to whether one of the keywords was included in the comment and, if so, whether the use of the word had a connotation related to obesity, nutrition, or weight loss. Of the 1000 sentences, 7 were “true positives” (ie, contained a keyword in a sentence that was related to obesity), 8 were “false positives” (ie, contained a keyword in a sentence unrelated to obesity), 7 were “false negatives” (ie, the sentence was related to obesity but did not include one of the keywords of interest), and the remaining were “true negatives” (ie, the sentence was not related to obesity and did not contain a keyword of interest). This resulted in high accuracy and specificity but low precision (ie, positive predictive value) and recall (ie, sensitivity; Multimedia Appendix 2). Consequentially, all analysis was restricted to only sentences containing the words “obese” or “obesity” (n=764,179 sentences across 689,447 comments after the removal of duplicates) to ensure that sentences included in the final analysis were related to the topic of interest. To assess the validity of this restriction, 100 random comments were selected for each keyword and evaluated as to whether they were related to the research question. Although 96% of comments containing obese and obesity were related to the research question, this was only true for 38% of comments for all keywords. Consequently, this restriction was considered sufficient.

Data Labeling

Three research assistants (RAs) trained by researchers with expertise in obesity medicine independently labeled 2000 random sentences in the analytic data set as misinformation, fact, stigma, positivity, or other. Misinformation and factual content were distinguished in accordance with peer-reviewed guidelines identified by the American Board of Obesity

Medicine and aligned with scientific literature [21]. Stigma was defined as derogatory language about individuals with obesity, including words such as "stupid," "lazy," or "dirty" [22]. Positivity was defined as affirmative language toward individuals with obesity or encouragement toward healthy weight loss (Multimedia Appendix 3).

In a form of hierarchical classification, the RAs were asked to assume a default label of fact or positivity unless the sentence contained misinformation (in which case they should label it as misinformation) or stigmatizing language (in which case they should label it as stigma). Sentences containing a mix of misinformation and stigmatizing language were labeled as stigma, while sentences containing both fact and positivity were labeled as fact. If the sentence was ambiguous or did not contain any type of information of interest, the RAs were instructed to label the sentence as other. A sentence was assigned a final label via an automated majority vote system if at least 2 RAs independently agreed on the label. If all 3 RAs disagreed on a label, the sentence was considered to not have a majority label. Using this method, 94% of sentences received a majority label. Fleiss κ was low (0.36), although this can occur even in instances where the agreement is high [23,24]. There were no significant psycholinguistic differences between posts that reached consensus and those that did not (Multimedia Appendix 3). Of the sentences receiving a majority label, 64% were labeled as other, 12% were labeled as fact, 9% were labeled as stigma, 5% were labeled as positivity, and 4% were labeled as misinformation.

Feature Extraction

Four feature categories were extracted from sentences in the analytic data set. Basic word context was extracted using term frequency–inverse document frequency (TF-IDF), which weighs the number of times a word or phrase appears in a sentence by its commonality within all analyzed sentences [25]. For this implementation of TF-IDF, only bigrams (ie, 2-word phrases) were retained if they were in the training data and had a document frequency >1% and <75%. These thresholds were chosen to limit computational complexity by excluding highly rare terms (ie, those with a frequency <1%) and highly common terms (ie, those with a frequency >75%) that may not produce informative features. To extract further information on the text, a pretrained, case-sensitive Bidirectional Encoder Representations from Transformers (BERT) model was used to generate dense numerical vector representations of the input sentences (ie, sentence embeddings). BERT-based models expand on traditional natural language processing (NLP) models such as Word2Vec by preserving the context of the input sentence in addition to basic word choice [26]. Sentence sentiment was extracted using the Valence Aware Dictionary and Sentiment Reasoner (VADER). Designed specifically for social media data, VADER expands on traditional lexicon-based approaches by incorporating grammatical rules into its analysis, including the use of capitalization, punctuation, negation, and emojis. The output of VADER is the ratio of text that is characterized as positive, neutral, and negative. These values are then used to generate a normalized, weighted composite score that aims to capture the overall sentiment of the text within a single number. It is calculated by summing the adjusted valence of each word within the text and normalizing it such that it falls on a scale from −1 to 1. After this normalization, values ≤−0.05 were considered "negative," values ≥0.05 were considered "positive," and values between −0.05 and 0.05 were considered "neutral" [27,28]. Specific psycholinguistic features were evaluated through the Linguistic Inquiry and Word Count (LIWC) program, which identifies the percentage of the sentence that can be ascribed to 80 different categories, including functional words (eg, pronouns, adjectives, and numbers), social words (eg, female and male referents), and informal speech (eg, swear words and punctuation). The LIWC also includes 4 proprietary metrics related to analytic thinking (characterized as logical and hierarchical thought), clout (characterized as displaying social status or confidence), authenticity (characterized as displaying humbleness or vulnerability), and emotional tone (whereby lower values represent more negative emotion). Additional metrics can be found in other studies [29,30]. The psycholinguistic features included in the LIWC can provide valuable insights into the writer's attitudes and perceptions toward the main topic of interest (such as obesity) [30]. Its utility in NLP tasks in the health domain on Reddit has been validated in other studies, including one that examined depression-related content on Reddit [31].

Model Development and Evaluation

A series of machine learning classification models were repeatedly trained and tested on the subset of 2000 labeled sentences to determine which would perform best at labeling the full data set. Five initial models were selected based on their innate ability to perform multiclass classification, capture nonlinearity in data, and generally achieve successful out-of-the-box performance. These models were random forest, Extreme Gradient Boosting (XGBoost), support vector machine with a radial basis function kernel, multinomial naive Bayes, and multilayer perceptron. Two "dummy" classifiers were also used so that model performance metrics for the 5 candidate classifiers could be compared with classifiers that use basic rules to assign the final label. One dummy classifier ("stratified" model) predicted the final label based on the class distribution of the training set, while the other ("most frequent" model) assigned every sentence the most frequent label within the training set. If a more complex candidate model could not outperform both dummy classifiers, this would suggest that it could not identify an underlying pattern within the data that could inform label selection; in other words, it could not "learn" how to distinguish misinformation, facts, stigma, and positivity from the extracted linguistic and semantic features. All models were deployed using the default hyperparameters.

Each model was evaluated using 3 versions of an 80:20 train-test split of the original data that received a majority label. In the first version, the training data remained unchanged (train: n=1496; test: n=375). In the second version, all minority classes were oversampled to match the size of the majority class (train: n=5115). This oversampling involved random sampling with replacement from each minority class to create a balanced data set. In the third version, the majority class was downsampled to match the size of the smallest minority class (train: n=535). This process involved randomly selecting sentences from the majority class until a pool of sentences the same size as the

smallest minority class was obtained. Both the second and third versions were performed to account for the unbalanced nature of the data; in all scenarios, the test data set remained unmodified so that an unbiased estimate of model performance could be calculated. All the features generated in *Feature Extraction* were first used in the models. The model and data set with the best performance were then subjected to forward variable selection to determine if a more parsimonious model could be generated without performance loss. The models were first trained using only 1 category of features (ie, TF-IDF, BERT, LIWC, or VADER). The feature with the best individual performance was then carried through to a secondary test that added additional features. This process was conducted iteratively until a model with all features was calculated or no future improvement was achieved. Model performance was also evaluated while varying the size of the training data to determine whether similar performance could be obtained with fewer labeled data. In all instances, model performance was evaluated using accuracy (ie, the proportion of sentences correctly assigned to their labeled class), average weight precision (ie, the weighted average across classes of the percentage of sentences assigned a given label that truly belong to that label), average weighted recall (ie, the weighted average across classes of the percentage of sentences of a given class assigned their correct label), and F_1-score (ie, the harmonic mean of precision and recall). Figure 1 shows a flow diagram depicting this process.

Statistical Analysis

Model feature importance was assessed by evaluating both the top 10 most frequent features across all trees used to split variables, as well as the top 10 features with the highest average information gain. Two sets of analysis were performed to characterize the fully labeled data set. First, for each content category (ie, misinformation, fact, stigma, and positivity), the mean of each LIWC variable in the Reddit data was compared with each LIWC variable's grand mean (ie, the weighted average of individual means). The LIWC grand mean was generated from the data used to create the LIWC software, which included 37,295 blog posts, 6179 pieces of expressive writing, 875 novels, 3232 transcripts of "natural speech," 34,929 articles from the *New York Times*, and 35,269 Twitter posts [29]. Comparisons were made using Cohen *d*, which measures the standardized difference between 2 means. This standardized difference is considered "large" if above 0.8, "medium" if between 0.5 and 0.8, "small" if it was between 0.2 and 0.5, and "negligible" if below 0.2 [32]. This analysis was conducted to quantify the magnitude of the difference between the Reddit data and a set of heterogeneous texts, which could help identify uniquely enriched features in various types of obesity-related content on the platform.

In the second analysis, individual hurdle models were constructed for each variable derived from TF-IDF, VADER, and LIWC to evaluate the variations relative to each pair of labeled data. In the first step, a logistic regression model assessed the log-odds of obtaining a 0 result (eg, the log-odds of no first-person singular pronoun) for either misinformation (referenced against fact) or stigma (referenced against positivity). In the second step of the process, data were truncated at 0 before a standard linear regression model was constructed to model the difference in the mean value of each LIWC variable for either misinformation (referenced against fact) or stigma (referenced against positivity). In the standard linear regression model, this translates to the percentage of the sentence that comprises a given variable. For example, a β coefficient of 10 in the misinformation model for swear words implies that, on average, misinformation comments contained 10% more swear words (absolute difference) than factual content. This procedure was chosen given the 0-inflated, semicontinuous nature of the data and the ability to separately model the conditional presence or absence of a feature. This allows for the nature of the effects to vary; for example, a feature with a negative coefficient in the binary outcome model but a positive coefficient in the semicontinuous model suggests that the feature is rare but, when present, exists in large quantities [33,34]. This process was also repeated in a single model that compared each label category to the "Other" category as a reference. In both scenarios, log-odds and 95% CIs were used to report the findings from the logistic portion of the model, whereas β coefficients with 95% CIs were used to report the findings from the linear regression portion of the model. To account for false discovery, *P* values within this post hoc analysis were recomputed using a reformulation of the Benjamini-Hochberg procedure that adjusts the actual *P* values themselves while allowing for the false discovery rate (in this case, 5%) to parallel the traditional significance threshold [35,36]. *P* values were recomputed for each category comparison and for each step of the hurdle model (eg, logistic model for fact vs misinformation, continuous model for fact vs misinformation, and logistic model for stigma vs positivity) for a total of 570 tests. In a sensitivity analysis, this process was repeated only for the training data. In all cases, a significance threshold of 0.05 was used. Analyses were conducted in R (version 4.1.0) and Python (version 3.7.4) using packages such as *imblearn*, *nltk*, *sentence_transformers*, *sklearn*, and *xgboost* [37-41].

Ethics Approval

No additional ethics approval was required for this study given the public nature of the relevant social media data [42]. The code is available at GitHub [43], while the data are available upon request.

Results

Data Description

The 764,179 sentences containing "obese" or "obesity" were derived from 689,447 top-level comments (1.11 sentences per comment on average) made between 2011 and 2019. Comments had an average score (ie, the difference between "upvotes" and "downvotes") of 14.3 (SD 167). These comments were generated by a maximum of 375,053 unique authors (1.84 comments per author minimum), of which 22,418 (5.98%) were "deleted." Of the 13,123 subreddits present within the data, the most frequent include r/AskReddit (n=97,540, a subreddit whereby users can "ask and answer thought-provoking questions"), r/fatlogic (n=77,417, a subreddit to "learn about or promote health eating habits, and dispel 'fatlogic' [ie, anything that deviates from the scientific facts of body weight management...]"), r/loseit (n=27,649, a subreddit "to discuss

healthy and sustainable methods of weight loss"), r/fatpeoplehate (n=14,992, a now-banned subreddit that was dedicated to sharing derogatory content about individuals with obesity), and r/Fitness (n=13,107, a subreddit to discuss "physical fitness goals").

Model Selection

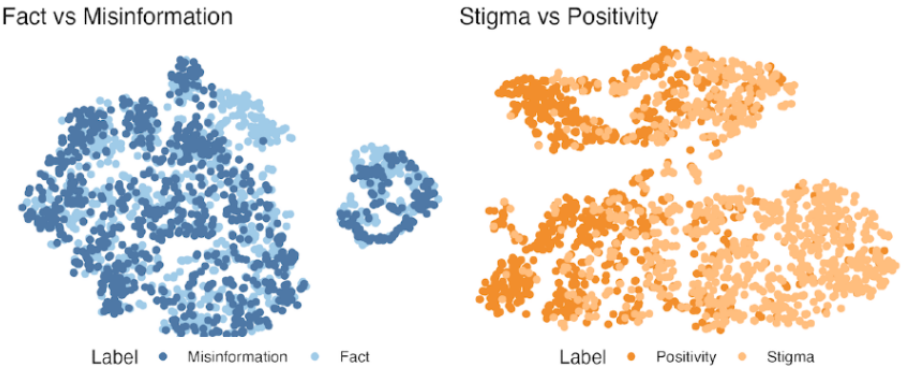
During the training process, XGBoost with oversampling achieved the best overall performance relative to all other models, including dummy classifiers (accuracy=0.69; weighted average precision=0.69; weighted average recall=0.69; weighted F_1-score=0.63; Multimedia Appendix 4). XGBoost also performed the best under the original data conditions (accuracy=0.69; weighted average recall=0.69; weighted average precision=0.61; weighted F_1-score=0.61), while naive Bayes performed best with the downsampled data (accuracy=0.30; weighted average precision=0.68; weighted average recall=0.30; weighted F_1-score=0.25). All features were retained after performing forward variable selection, although the BERT-only model had identical accuracy, weighted average recall, and weighted average F_1-score values compared with the full model; weighted average precision was 3 percentage points less (0.64 vs 0.69; Multimedia Appendix 5). Finally, the analysis of training size versus performance did not yield a plateau before the maximum value was reached, so all labeled posts were retained for training (Multimedia Appendix 6). The confusion matrix of the final XGBoost model with oversampling can be found in Multimedia Appendix 7.

Model Implementation and Feature Importance

After running the final XGBoost model with an oversampled training data set on the entire set of sentences, 3610 (0.47%) sentences were labeled as misinformation, 14,366 (1.88%) sentences were labeled as weight-related stigma, 14,799 (1.94%) were labeled as positivity, 68,276 (8.93%) were labeled as factual content, and the rest (663,128/764,179, 86.78%) were labeled as other. While misinformation and factual content appeared to cluster together in a 2D space (suggesting the presence of some similarity between the categories), stigma and positivity were more distinct (suggesting greater differentiation between the labeled categories; Figure 2).

There were 875 candidate features (11 from TF-IDF, 4 from VADER, 767 from BERT, and 93 from LIWC) for the model to evaluate. When frequency was used to assess which features contributed the most to the labeling process, the top 10 features were all dimensions of BERT, meaning that they did not have a direct interpretation. In contrast, 4 of the top-10 features with the highest information gain were non-BERT. These were first-person singular pronouns (eg, "I," "me," and "mine"), anger-tagged terms (eg, "hate," "kill," and "annoyed"), word count, and swear words. None of the BERT dimensions that contributed the most to information gain were in the top 10 of the frequent feature analysis.

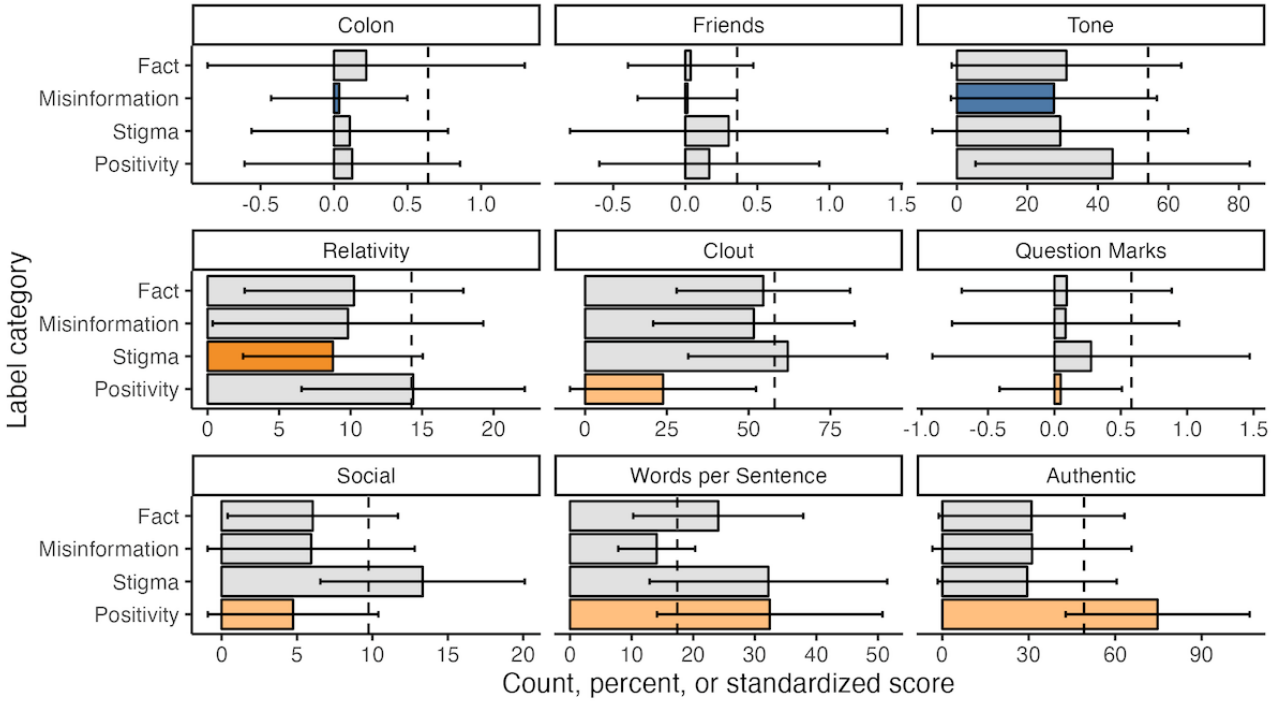
Figure 2. Visualization of each labeled category using t-distributed stochastic neighbor embedding (TSNE) dimensionality reduction. The left plot compares factual content and misinformation, while the right plot compares positivity and stigma. In both instances, the full set of labeled data was sampled randomly without replacement (points per category: n=1000).



External Comparator Analysis With LIWC Grand Means

Of the 93 LIWC features compared against the external grand mean (LIWC μ), 21 (23%) had a large, standardized difference for at least one (but not all) of the labels and 9 (10%) had a large, standardized difference for only 1 category (Figure 3). Compared with the LIWC texts, misinformation sentences had a lower average amount of friend-tagged terms (eg, "buddy" and "neighbor"; LIWC: μ=0.36, =0.01, s=0.35, Cohen *d*=1.30) and a more negative tone on average (LIWC: μ=54.22, =27.5, s=29.2, Cohen *d*=0.92); however, this was not true of factual content. Sentences with stigma had a lower average amount of relativity-related terms (eg, "area," "bend," and "exit") compared with the LIWC texts (LIWC: μ=14.26, =8.76, s=6.28, Cohen *d*=0.87). Sentences tagged as positivity had lower clout on average (LIWC: μ=57.95, =23.8, s=28.4, Cohen *d*=1.20) and number of social-tagged terms (eg, "mate," "talk," "there," and "child"; LIWC: μ=9.74, =4.74, s=5.66, Cohen *d*=0.88) but higher authenticity (LIWC: μ=49.17, =74.7, s=31.9, Cohen *d*=0.80) relative to the LIWC Program texts. There were no terms with an exclusively large, standardized difference for facts, although both facts and positivity had fewer assent-tagged terms on average (eg, "agree," "OK," and "yes") than the LIWC texts (LIWC: μ=0.95, =0.09, s=0.75, Cohen *d*=1.14; =0.18, s=0.84, Cohen *d*=0.92). A comparison of all LIWC features across the 4 categories of interest can be found in Multimedia Appendix 8.

Figure 3. Comparisons between the grand mean of select psycholinguistic features from the 2015 Linguistic Inquiry and Word Count Program (LIWC 2015) and the 4 categories of interest: misinformation, factual content, stigma, and positivity. X-axis values for tone and clout are standardized scores based on the proprietary LIWC 2015 algorithm; x-axis for "Words per Sentence" is word count; x-axis for all other variables are the percentage of total words within the text. Only features that had 1 label category with a large, standardized difference (ie, Cohen d>0.80) are shown, with the substantially different category highlighted within each subplot. Error bars denote 1 SD above and below the mean, while the dashed lines denote the LIWC 2015 grand mean for each category. Note that x-axes are individualized per feature and cannot be compared across subplots. For a full list of numerical comparisons, see Multimedia Appendix 8.



Feature Significance by Category

Many psycholinguistic features were significantly different between factual content and misinformation (Figure 4). For example, on average, there was a lower percentage of words of at least 6 letters in misinformation than fact (β=−1.47, 95% CI −1.85 to −1.10; *P*<.001); however, the log-odds of having no 6-letter words were not significantly different between categories (β=−0.03, 95% CI −0.43 to 0.32; *P*=.87). Misinformation also had a lower compound sentiment score (β=−0.03, 95% CI −0.04 to −0.01; *P*<.001), lower log-odds of containing no negation (eg, "no," "not," and "never"; β=−0.27, 95% CI −0.34 to −0.20; *P*<.001), and more negations overall (β=3.71, 95% CI 3.53-3.90; *P*<.001) compared with facts, suggesting a generally more negative and contrarian sentiment. Although misinformation also had a higher percentage of "net speak" on average (eg, "btw," "lol," and "thx") compared with facts (β=4.34, 95% CI 3.66-5.03; *P*<.001), the presence of net speak within misinformation was not significantly different than the "Other" category (β=1.32, 95% CI −0.43 to 3.07, *P*=.15; see Multimedia Appendix 9 for a full list of comparisons between each label category and "Other"). Interestingly, misinformation also had higher log-odds of no net speak in any given sentence relative to factual content (β=1.69, 95% CI 1.37-2.05; *P*<.001) and other (β=2.17, 95% CI 1.85-2.53; *P*<.001). Taken together, this suggests that net speak is not always in misinformation but, when it is, it is present in large quantities. Although some differences in coefficient directionality emerged within the sensitivity analysis, they are likely attributable to the extremely small sample size of the data used for that analysis. A complete list of comparisons between factual content and misinformation, including findings from the logistic regression portion of the hurdle models, can be found in Multimedia Appendix 10.

Similarly, multiple psycholinguistic features were significantly different between positivity and stigma (Figure 5). There were significantly fewer references to all personal pronouns in sentences containing stigma compared with positivity, especially first-person singular personal pronouns (β=−5.30, 95% CI −5.44 to −5.16; *P*<.001). Stigma also had higher log-odds of containing no first-person singular pronouns relative to positivity (β=2.57, 95% CI 2.51-2.64; *P*=.15). In contrast, third-person plural pronouns were significantly more prevalent in sentences containing stigma (β=1.95, 95% CI 1.74-2.15; *P*<.001), and stigmatizing comments had lower log-odds of containing none of these pronouns relative to positivity pronouns (β=−2.12, 95% CI −2.19 to −2.05; *P*<.001). Negative emotions (β=2.00, 95% CI 1.90-2.10; *P*<.001) were also more prevalent in stigmatizing comments, which had lower log-odds of containing no words with a negative connotation relative to positivity (β=−0.90, 95% CI −0.95 to −0.85; *P*<.001). Similar to misinformation and factual content, some differences in directionality occurred within the sensitivity analysis, although this is likely attributable to variations in sample size. A complete list of comparisons, including log-odds for the first portion of the hurdle model, can be found in Multimedia Appendix 11.

Figure 4. Select psycholinguistic features significantly different between fact and misinformation. Estimates are derived from the semicontinuous part of a 2-part hurdle model. The outcome of interest is the percentage of the sentence containing the psycholinguistic feature, and the exposure is the sentence label of either fact (reference, n=68,276) or misinformation (n=3610). Positive values suggest a higher prevalence of the feature among misinformation compared with fact, while negative values suggest a higher prevalence of the feature among fact compared with misinformation. Only the top 10 features with the largest effect size in each direction are shown, excluding 9 nonsignificant features and 74 significant features.

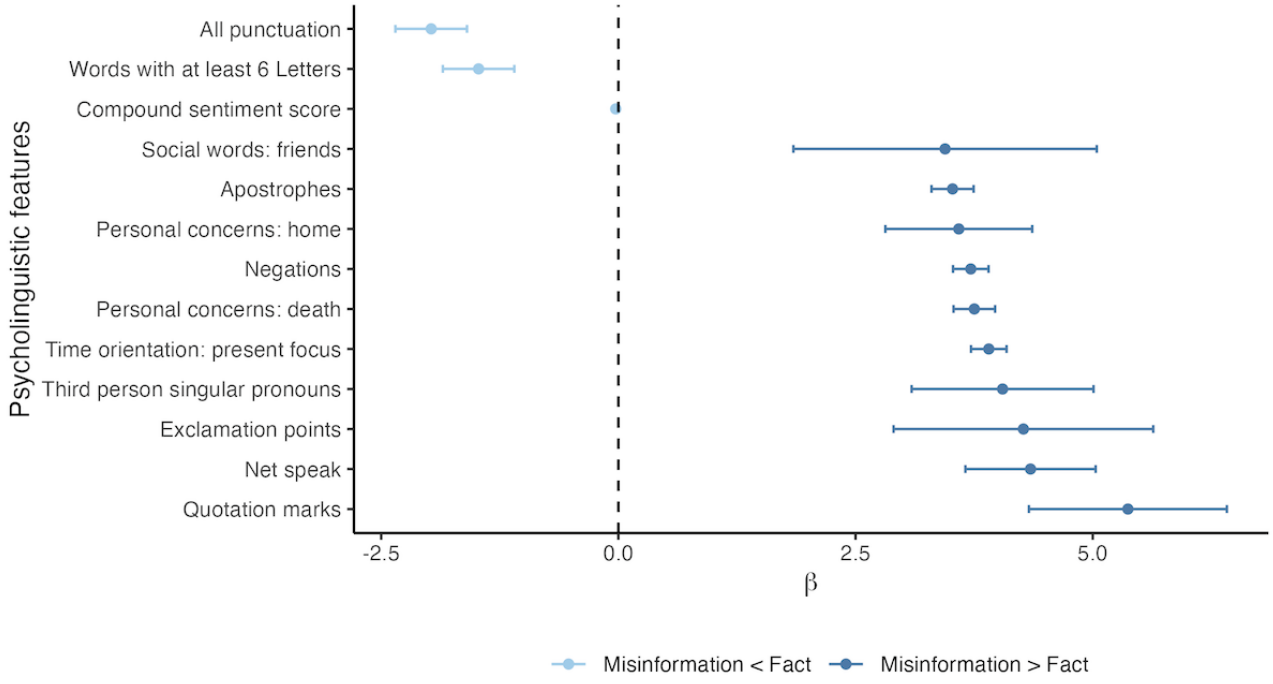
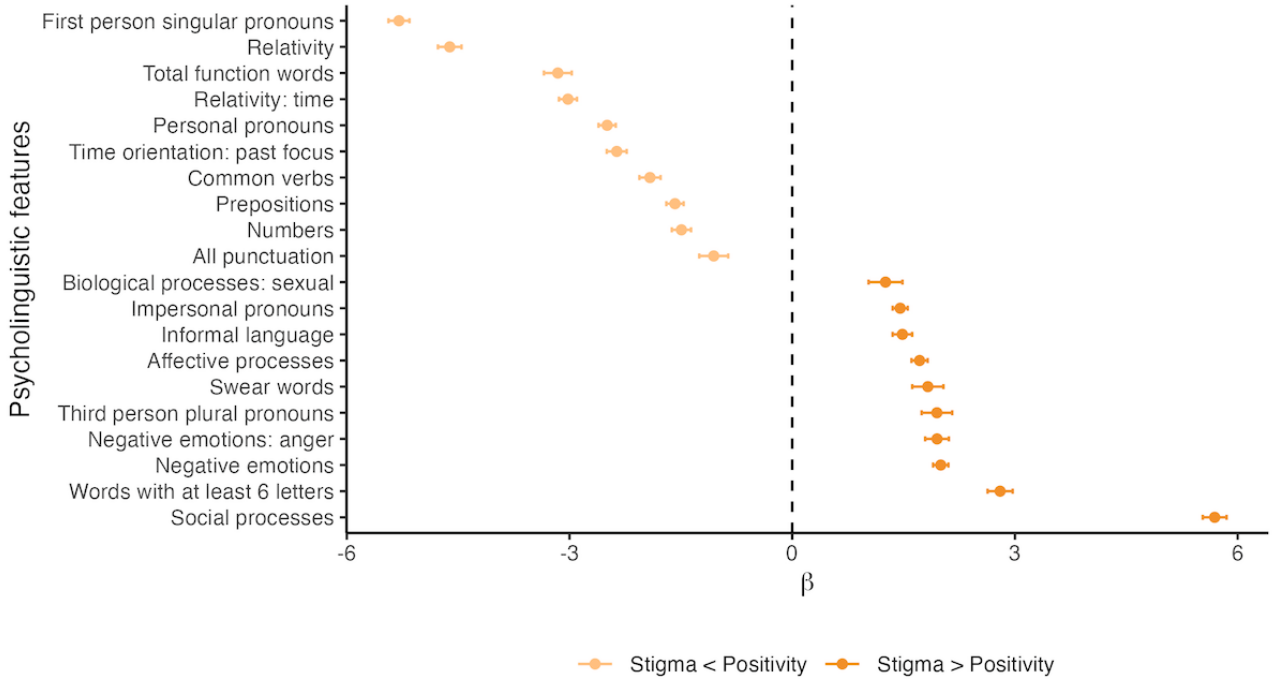


Figure 5. Select psycholinguistic features significantly different between positivity and stigma. Estimates are derived from the semicontinuous part of a 2-part hurdle model. The outcome of interest is the percentage of the sentence containing the psycholinguistic feature, and the exposure is the sentence label of either positivity (reference, n=14,799) or stigma (n=14,366). Positive values suggest a higher prevalence of the feature among stigma compared with positivity, while negative values suggest a higher prevalence of the feature among positivity compared with stigma. Only the top 10 features with the largest effect size in each direction are shown, excluding 20 nonsignificant features and 56 significant features.



Discussion

Principal Findings

This study is the first to comprehensively evaluate obesity-related content on Reddit. Using a multiclass XGBoost model and a suite of NLP features, misinformation was found to be relatively scarce on the platform. However, it had several unique features that distinguished it from both an external corpus and internal facts. There were a relatively similar number of sentences containing stigma and positivity, which could also be distinguished from one another. The novelty of the presented

work is multifaceted and involves both methodological and applied contributions. From a methods standpoint, the pipeline developed here could be readily adapted to understand other public health topics on Reddit with minimum alterations required. From an applied standpoint, the findings provide a baseline for comparison in future work that may look at more specific facets of obesity content (such as nutrition, bariatric surgery, or antiobesity pharmacotherapies). Altogether, the results introduce a candidate set of features that could be explored as indicators for deleterious content, describe a framework for classifying obesity content, and provide important insights into the state of obesity content on the platform that could inform future health communication.

The underlying meanings of the psycholinguistic features enhanced by misinformation and stigma can help characterize how individuals on Reddit perceive obesity. For example, misinformation tended to have more of a present tense than factual content, a finding aligned with other studies on web-based misinformation [44]. Given that tense has been considered a proxy for "psychological distance" (whereby present and future tense suggest closer affinity compared with past tense), this may suggest that individuals experience stronger emotions when describing misinformation compared with factual content [30,45]. In addition, misinformation has a higher number of quotation marks and net speak relative to factual content. Although quotation marks are commonly used to signify an idea originally shared by someone other than the writer (such as the citation of an external reference), they can also be used as scare quotes to highlight something as ironic and distance the writer from the original meaning of the word [46]. Given that quotations could also be used when citing a formal manuscript, it is important to consider other markers beyond punctuation as possible indicators of obesity misinformation. For example, on average, net speak was present in higher quantities in misinformation than in facts. This suggests that individuals who describe untrue content may use more casual language, whereas those who describe something truthful may use more formal language. This could also indicate the presence of bots or trolls, who may use this kind of language to appear more human-like. However, the average amount of quotation marks and net speak within misinformation was not significantly different than the "other" category, which may mean that the labeled misinformation category is just capturing colloquial conversations and not actual misinformation.

In contrast, sentences containing stigma were enriched for social processes, negative emotions (including anger), and third-person plural pronouns (eg, "they"). This aligns with research within the vaccination space, which also found that stigmatized content was more negative and contained more third-person plural pronouns [47]. Similar to that study, this study found that positivity sentences contained a higher prevalence of first-person singular pronouns. Understanding the prevalence of pronouns within a text can help inform where a writer's attention is placed; in this case, the enrichment of third-person pronouns in stigmatizing sentences may suggest that the stigmatizing language is directed at external individuals instead of the commenter [30]. Alternatively, these sentences could be made by users who are describing a situation in which they experienced stigma, as a prior study found that male students (who make up a majority of the Reddit userbase) tended to use more third-person pronouns when describing a scenario in which they were teased [48]. Although these smaller function words are typically removed during traditional NLP analysis, given their abundance in writing, the results presented here suggest that they may be important indicators that could inform how social media users feel about various topics.

The LIWC benchmark analysis demonstrated that each category of interest was also distinct from other heterogeneous texts. This may be due, in part, to the differences in who generated the content for each source. Although Reddit users are primarily men (61.8%) and aged ≤50 (58%) years, LIWC sources are heterogeneous [29]. Thus, this benchmark analysis may highlight unique dimensions of how the Reddit-specific demographic communicate about obesity. For example, there was no LIWC feature that was distinctly different between facts and the LIWC corpus. This may make sense, as facts are likely paraphrased or directly quoted from an external, reputable source. In contrast, stigma and positivity are likely original thoughts, explaining why they have some components that are "enriched" compared with the benchmarked texts. Misinformation may fall somewhere in between, as it could be either one's own misconstrued idea about the topic or a quotation of a source that is misaligned with the scientific consensus. Future work could help inform the type of misinformation that is present on Reddit.

The classification pipeline developed here can be applied in many different ways. For example, tagging posts as misinformation or stigma could aid Reddit content moderators (ie, managers of subreddits that monitor content for policy violations) in identifying which comments may violate their community guidelines. Moderators can leverage historical data where they flagged and removed deleterious content as additional training data, creating a refined model that is tuned to the needs of each subreddit. Furthermore, researchers could use this model as a tool in a larger pipeline that seeks to understand the impact of user-facing flags on misinformation or stigma on behaviors. This may include behaviors of the initial commenter (eg, future comments containing misinformation or stigma) or other users (eg, percent of "upvotes," number of response comments, or sentiment of responding comments). Automating the identification process of candidate comments would allow researchers to allocate more time to other areas of research, such as the development of specific countermessaging for each type of misinformation. Although the current classifier was developed specifically for obesity, it can be readily adapted to other topics of concern, especially if labeled training data already exist. Other adaptations could also include accounting and adjusting for common challenges in automated classification such as misspellings or sarcasm [49].

Comparison With Prior Work

Prior text-based classification analysis on Reddit has been conducted for a myriad of health conditions. The bulk of this work has been focused on the mental health space, including the ability to classify posts into various mental illnesses or identify the risk of suicide [50,51]. Other studies have leveraged NLP tools to classify a user's response to misinformation or to

automate its detection [52,53]. Less work has been done within the stigma space, although some work exists that automates the detection of cyberbullying [54]. Many of these studies rely on similar features that are used in this work, most notably word embeddings [50,51,54]. Although these studies relied on context-independent models, such as Word2Vec or GLOVE, more recent work (especially within the COVID-19 space) has leveraged the bidirectionality of BERT to generate context-dependent embeddings [55,56]. The power of BERT was also evident in this study, as a majority of the most important model features were the dimensions of the BERT embeddings.

Although there are no prior studies that directly quantify obesity content on Reddit, the results presented here align in part with the findings on other social media platforms. A study that specifically looked at nutritional guidelines for selecting Facebook pages of bariatric surgery support groups found that over 50% of posts were either inaccurate or highly ambiguous [57]. This proportion is lower than the amount of inaccurate (ie, misinformation) or ambiguous (ie, other) information found in this study (87.2%), which could be explained by the difference in platform (ie, Reddit vs Facebook), topic (ie, obesity vs nutrition), or purpose (ie, specific advice seeking vs general content). In terms of stigma, a study on Twitter found that tweets on obesity often included jokes, and tweets containing derogatory jokes were retweeted more frequently than tweets with jokes that were not derogatory [11]. Although the presence of jokes was not assessed in this study, this may explain why third-person plural pronouns, informal speech, and negative tone were higher in sentences containing stigma compared with body positivity. This finding was verified in a separate Twitter study that also identified the presence of "unverified health content," which may be considered a type of misinformation [13]. Altogether, although there may be common themes in communication that exist across platforms (such as the use of informal language and select function words and punctuation), there are also likely some platform-specific variations that warrant additional consideration. However, future work is required to precisely quantify the extent of variation in select linguistic and semantic features across platforms.

Limitations

This study excels in its comprehensive approach in analyzing a large corpus of text that could be applied to health domains beyond obesity. Yet, there are several limitations that are important to note. Studies on social media are a crucial tool for understanding health attitudes and behaviors, but each platform has a different user base that may influence the study's generalizability. This is particularly challenging to assess on Reddit, where users post anonymously. External surveys suggest that Reddit is predominately used by adult men, and thus, these results may not be generalizable to other populations or social media platforms. Despite this limitation, Reddit remains a vital platform to study given the growth of its userbase over the past decade (from an average of 46 million monthly active users in 2012 to 430 million monthly active users in 2019) and frequent usage among adults in the United States—approximately 18% report using Reddit "ever," and of those that use Reddit, 43% rely on it for news [1,58,59]. Given the growth of the platform over the analyzed period, coupled with platform-specific changes (such as the ban of r/fatpeoplehate in 2015), it may also be important to explore temporal trends in obesity misinformation and stigma in future work. Second, this study relied on automated machine learning classifiers using their default hyperparameters to categorize each sentence, and as a result, there may have been misclassification of content. It is assumed that this would be randomly distributed and not disproportionately impact one class, but future work could improve the performance of the present model. This would also include refining the process such that fewer sentences received a label of "other," tuning hyperparameters, including metadata in the classification algorithm (such as a post's score or controversiality), leveraging stratified sampling when selecting promising keywords, and assessing the robustness of the model when novel types of misinformation and stigma are presented (eg, misinformation that is enriched for 6-letter words with less net speak).

Third, this study only focused on sentences containing "obese" or "obesity," meaning that a large portion of content was excluded from analysis. These words were chosen given their direct relationship to the research question as compared with more colloquial terminology such as "weight" (which frequently referred to the heaviness of inanimate objects) and "fat" (which frequently referred to the dietary fat found in food). Although this resulted in high precision compared with the use of the full set of search terms, it likely decreased the recall of our study. In addition, given that the selected search terms could be considered clinical in nature, the true amount of misinformation and stigma on the platform is likely higher than what is presented here. Thus, future work should integrate this type of casual terminology into the data, starting with terms that are synonymous with obesity but do not have additional meanings (eg, "chubby"). In addition, this study does not consider any comment-level or user-level clustering, which can be addressed in future work. However, the impact on the current work is estimated to be minimum, as there were approximately 1.13 sentences per comment and a maximum of 1.83 comments per author (assuming all authors, including those with deleted account information, were distinct). Fourth, only top-level comments were analyzed, and future work could extend this to evaluate posts or lower-level comments. Fifth, the hurdle model point estimates in the labeled training data differed from the full model estimates on occasion. This may be attributable to the size of the training data, and future studies should explore this in detail. Finally, this study only focused on obesity, and future work could expand this to include tangential topics, such as weight loss, bariatric surgery, or nutrition.

Conclusions

This study presents the first comprehensive analysis of the state of obesity-related misinformation, facts, stigma, and positivity on Reddit. Although the prevalence of misinformation and stigma appears low, these numbers are likely to be the lower bound of the actual amount on the platform. Given the rapid growth of the Reddit userbase, public health researchers should increasingly consider Reddit as a source of misinformation and stigma, especially because misinformation that begins on this platform could spread to other platforms and into everyday

conversations. Furthermore, the identification of distinct psycholinguistic features that separate misinformation and stigma from facts and positivity can help Reddit moderators to more rapidly identify content that should be flagged and removed. Beyond this manual process, future work should consider how to leverage these features to aid in the automated identification of deleterious content on Reddit and other social media platforms in real time.

Acknowledgments

This publication was supported in part by a National Library of Medicine (NLM)–funded predoctoral fellowship (5T32LM012204-05) and an Alumni Research Award from the Dartmouth College Guarini School of Graduate and Advanced Studies. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NLM, National Institutes of Health, or Dartmouth College.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of keywords used in Reddit search.
[DOCX File , 13 KB - jmir_v24i12e36729_app1.docx]

Multimedia Appendix 2
Confusion matrix for keyword refinement.
[DOCX File , 15 KB - jmir_v24i12e36729_app2.docx]

Multimedia Appendix 3
Comparison of sentiment and Linguistic Inquiry and Word Count output by whether a majority label was able to be reached among the three trained research assistants.
[DOCX File , 18 KB - jmir_v24i12e36729_app3.docx]

Multimedia Appendix 4
Model performance comparisons.
[DOCX File , 15 KB - jmir_v24i12e36729_app4.docx]

Multimedia Appendix 5
Model ablation testing results.
[DOCX File , 15 KB - jmir_v24i12e36729_app5.docx]

Multimedia Appendix 6
Training data size versus model performance.
[DOCX File , 56 KB - jmir_v24i12e36729_app6.docx]

Multimedia Appendix 7
Confusion matrix of final model.
[DOCX File , 16 KB - jmir_v24i12e36729_app7.docx]

Multimedia Appendix 8
Comparisons between Linguistic Inquiry and Word Count psycholinguistic features and label categories.
[DOCX File , 32 KB - jmir_v24i12e36729_app8.docx]

Multimedia Appendix 9
Full hurdle models: other versus all categories.
[DOCX File , 53 KB - jmir_v24i12e36729_app9.docx]

Multimedia Appendix 10
Full hurdle models: fact versus misinformation.
[DOCX File , 42 KB - jmir_v24i12e36729_app10.docx]

Multimedia Appendix 11
Full hurdle models: positivity versus stigma.
[DOCX File , 45 KB - jmir_v24i12e36729_app11.docx]

References

1. Auxier B, Anderson M. Social Media Use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-10-26]
2. Bovet A, Makse HA. Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 2019 Jan 02;10(1):7 [FREE Full text] [doi: 10.1038/s41467-018-07761-2] [Medline: 30602729]
3. Islam MS, Sarkar T, Khan SH, Mostofa Kamal AH, Hasan SM, 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]
4. 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]
5. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017-2018. NCHS Data Brief 2020 Feb(360):1-8 [FREE Full text] [Medline: 32487284]
6. Lydecker JA, Cotter EW, Palmberg AA, Simpson C, Kwitowski M, White K, et al. Does this Tweet make me look fat? A content analysis of weight stigma on Twitter. Eat Weight Disord 2016 Jun;21(2):229-235. [doi: 10.1007/s40519-016-0272-x] [Medline: 27068174]
7. Wanniarachchi VU, Mathrani A, Susnjak T, Scogings C. A systematic literature review: what is the current stance towards weight stigmatization in social media platforms? Int J Human Comput Stud 2020 Mar;135:102371. [doi: 10.1016/j.ijhcs.2019.102371]
8. Lazarus JV, Kakalou C, Palayew A, Karamanidou C, Maramis C, Natsiavas P, et al. A Twitter discourse analysis of negative feelings and stigma related to NAFLD, NASH and obesity. Liver Int 2021 Oct;41(10):2295-2307. [doi: 10.1111/liv.14969] [Medline: 34022107]
9. Wu YK, Berry DC. Impact of weight stigma on physiological and psychological health outcomes for overweight and obese adults: a systematic review. J Adv Nurs 2018 May;74(5):1030-1042. [doi: 10.1111/jan.13511] [Medline: 29171076]
10. Holland G, Tiggemann M. A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image 2016 Jun;17:100-110. [doi: 10.1016/j.bodyim.2016.02.008] [Medline: 26995158]
11. So J, Prestin A, Lee L, Wang Y, Yen J, Chou WS. What do people like to "share" about obesity? A content analysis of frequent retweets about obesity on Twitter. Health Commun 2016;31(2):193-206. [doi: 10.1080/10410236.2014.940675] [Medline: 26086083]
12. Chou WY, Prestin A, Kunath S. Obesity in social media: a mixed methods analysis. Transl Behav Med 2014 Sep;4(3):314-323 [FREE Full text] [doi: 10.1007/s13142-014-0256-1] [Medline: 25264470]
13. Mejova Y. Information sources and needs in the obesity and diabetes Twitter discourse. In: Proceedings of the 2018 International Conference on Digital Health. 2018 Presented at: DH '18; April 23-26, 2018; Lyon, France p. 21-29. [doi: 10.1145/3194658.3194664]
14. Homepage. Reddit. URL: https://www.redditinc.com/ [accessed 2021-11-17]
15. Lee J. The 20 Biggest Subreddits That Are Still Worth Subscribing To. whatNerd. 2021 Jul 12. URL: https://whatnerd.com/biggest-subreddits-worth-subscribing-to/ [accessed 2021-10-26]
16. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Couto Silva AP, Meira W, et al. Factors associated with weight change in online weight management communities: a case study in the LoseIt Reddit community. J Med Internet Res 2017 Jan 16;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]
17. Liu Y, Yin Z. Understanding weight loss via online discussions: content analysis of Reddit posts using topic modeling and word clustering techniques. J Med Internet Res 2020 Jun 08;22(6):e13745 [FREE Full text] [doi: 10.2196/13745] [Medline: 32510460]
18. Chandrasekharan E, Pavalanathan U, Srinivasan A, Glynn A, Eisenstein J, Gilbert E. You can't stay here: the efficacy of Reddit's 2015 ban examined through hate speech. Proc ACM Hum Comput Interact 2017 Nov;1(CSCW):1-22. [doi: 10.1145/3134666]
19. Medical Subject Headings. National Library of Medicine. URL: https://www.nlm.nih.gov/mesh/meshhome.html [accessed 2020-01-17]
20. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift Reddit dataset. Proc Int AAAI Conf Web Soc Media 2020 May 26;14(1):830-839 [FREE Full text] [doi: 10.1609/icwsm.v14i1.7347]
21. Exam Resources. American Board of Obesity Medicine. URL: https://www.abom.org/exam-resources-2/ [accessed 2020-11-05]
22. 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]

23. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [doi: 10.1016/0895-4356(90)90158-l] [Medline: 2348207]
24. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43(6):551-558. [doi: 10.1016/0895-4356(90)90159-m] [Medline: 2189948]
25. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. J Documentation 1972 Jan;28(1):11-21. [doi: 10.1108/eb026526]
26. 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 Jun Presented at: NAACL '19; June 2-7, 2019; Minneapolis, MN, USA p. 4171-4186 URL: https://aclanthology.org/N19-1423/ [doi: 10.18653/v1/N19-1423]
27. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proc Int AAAI Conf Web Soc Media 2014 May 16;8(1):216-225. [doi: 10.1609/icwsm.v8i1.14550]
28. Hutto CJ. cjhutto/vaderSentiment. GitHub. 2022. URL: https://github.com/cjhutto/vaderSentiment [accessed 2022-07-15]
29. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. The University of Texas at Austin. Austin, TX, USA: The University of Texas at Austin; 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2020-09-01]
30. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
31. Tadesse MM, Lin H, Xu B, Yang L. Detection of depression-related posts in Reddit social media forum. IEEE Access 2019 Apr 4;7:44883-44893. [doi: 10.1109/ACCESS.2019.2909180]
32. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
33. Neelon B, O'Malley AJ, Smith VA. Modeling zero-modified count and semicontinuous data in health services research Part 1: background and overview. Stat Med 2016 Nov 30;35(27):5070-5093. [doi: 10.1002/sim.7050] [Medline: 27500945]
34. Neelon B, O'Malley AJ, Smith VA. Modeling zero-modified count and semicontinuous data in health services research part 2: case studies. Stat Med 2016 Nov 30;35(27):5094-5112. [doi: 10.1002/sim.7063] [Medline: 27500973]
35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
36. Benjamini Y, Heller R, Yekutieli D. Selective inference in complex research. Philos Trans A Math Phys Eng Sci 2009 Nov 13;367(1906):4255-4271 [FREE Full text] [doi: 10.1098/rsta.2009.0127] [Medline: 19805444]
37. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res 2017;18(17):1-5.
38. Bird S, Loper E, Klein E. Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit. O'Reilly Media Inc. 2009. URL: https://www.nltk.org/book/ [accessed 2022-05-05]
39. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019 Nov Presented at: EMNLP-IJCNLP '19; November 3-7, 2019; Hong Kong, China p. 3982-3992 URL: https://aclanthology.org/D19-1410 [doi: 10.18653/v1/d19-1410]
40. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12(85):2825-2830.
41. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Aug Presented at: KDD '16; August 13-17, 2016; San Francisco, CA, USA p. 785-794. [doi: 10.1145/2939672.2939785]
42. 2018 Requirements (2018 Common Rule). Office for Human Research Protections (OHRP). 2018. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html [accessed 2020-09-01]
43. Reddit Obesity Misinformation Stigma. GitHub. URL: https://github.com/cpollack736/reddit_obesity_misinformation_stigma [accessed 2022-12-07]
44. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R. Automatic detection of fake news. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018 Presented at: COLING '18; August 20-26, 2018; Santa Fe, NM, USA p. 3391-3401 URL: https://aclanthology.org/C18-1287
45. Pasupathi M. Telling and the remembered self: linguistic differences in memories for previously disclosed and previously undisclosed events. Memory 2007 Apr;15(3):258-270. [doi: 10.1080/09658210701256456] [Medline: 17454663]
46. Gutzmann D, Stei E. How quotation marks what people do with words. J Pragmatics 2011 Aug;43(10):2650-2663. [doi: 10.1016/j.pragma.2011.03.010]
47. Straton N, Jang H, Ng R, Vatrapu R, Mukkamala RR. Computational modeling of stigmatized behaviour in pro-vaccination and anti-vaccination discussions on social media. In: Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine. 2019 Presented at: BIBM '19; November 18-21, 2019; San Diego, CA, USA p. 2673-2681. [doi: 10.1109/bibm47256.2019.8983311]

48. Kowalski RM. “I was only kidding!”: victims’ and perpetrators’ perceptions of teasing. Pers Soc Psychol Bull 2000 Feb;26(2):231-241. [doi: 10.1177/0146167200264009]
49. Nobata C, Tetreault J, Thomas A, Mehdad Y, Chang Y. Abusive language detection in online user content. In: Proceedings of the 25th International Conference on World Wide Web. 2016 Apr Presented at: WWW '16; April 11-15, 2016; Montreal, Canada p. 145-153. [doi: 10.1145/2872427.2883062]
50. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJ, Dobson RJ, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. Sci Rep 2017 Mar 22;7:45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
51. Gaur M, Alambo A, Sain JP, Kursuncu U, Thirunarayan K, Kavuluru R, et al. Knowledge-aware assessment of severity of suicide risk for early intervention. In: Proceedings of the 2019 The World Wide Web Conference. 2019 Presented at: WWW '19; May 13-17, 2019; San Francisco, CA, USA p. 514-525. [doi: 10.1145/3308558.3313698]
52. Dang A, Smit M, Moh’d A, Minghim R, Milios E. Toward understanding how users respond to rumours in social media. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2016 Presented at: ASONAM '16; August 18-21, 2016; San Francisco, CA, USA p. 777-784. [doi: 10.1109/asonam.2016.7752326]
53. Srinivas J, Venkata Subba Reddy K, Sunny Deol GJ, VaraPrasada Rao P. Automatic fake news detector in social media using machine learning and natural language processing approaches. In: Proceedings of the 4th International Conference on Smart Computing and Informatics. 2020 Presented at: SCI '20; June 5-6, 2020; Hyderabad, India p. 295-305. [doi: 10.1007/978-981-16-1502-3_30]
54. Aggarwal A, Maurya K, Chaudhary A. Comparative study for predicting the severity of cyberbullying across multiple social media platforms. In: Proceedings of the 4th International Conference on Intelligent Computing and Control Systems. 2020 Presented at: ICICCS '20; May 13-15, 2020; Madurai, India p. 871-877. [doi: 10.1109/iciccs48265.2020.9121046]
55. Pandey C. redBERT: a topic discovery and deep sentiment classification model on COVID-19 online discussions using BERT NLP model. Int J Open Source Softw Process 2021;12(3):32-47. [doi: 10.4018/ijossp.2021070103]
56. Choudrie J, Patil S, Kotecha K, Matta N, Pappas I. Applying and understanding an advanced, novel deep learning approach: a Covid 19, text based, emotions analysis study. Inf Syst Front 2021;23(6):1431-1465 [FREE Full text] [doi: 10.1007/s10796-021-10152-6] [Medline: 34188606]
57. Koball AM, Jester DJ, Pruitt MA, Cripe RV, Henschied JJ, Domoff S. Content and accuracy of nutrition-related posts in bariatric surgery Facebook support groups. Surg Obes Relat Dis 2018 Dec;14(12):1897-1902. [doi: 10.1016/j.soard.2018.08.017] [Medline: 30274739]
58. Dean B. Reddit User and Growth Stats (Updated Oct 2021). Backlinko. 2021 Oct 12. URL: https://backlinko.com/reddit-users [accessed 2022-05-05]
59. U.S. Reddit usage reasons 2019. Statista. 2019. URL: https://www.statista.com/statistics/262331/digital-communication-platforms-of-us-internet-users/ [accessed 2021-11-17]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
LIWC: Linguistic Inquiry and Word Count program
NLP: natural language processing
RA: research assistant
TF-IDF: term frequency–inverse document frequency
VADER: Valence Aware Dictionary and Sentiment Reasoner
XGBoost: Extreme Gradient Boosting

Edited by A Mavragani; submitted 25.01.22; peer-reviewed by C Shoults, J Chen, D Huang, O Serban; comments to author 12.04.22; revised version received 06.06.22; accepted 07.09.22; published 30.12.22.

Please cite as:
Pollack CC, Emond JA, O'Malley AJ, Byrd A, Green P, Miller KE, Vosoughi S, Gilbert-Diamond D, Onega T
Characterizing the Prevalence of Obesity Misinformation, Factual Content, Stigma, and Positivity on the Social Media Platform Reddit Between 2011 and 2019: Infodemiology Study
J Med Internet Res 2022;24(12):e36729
URL: https://www.jmir.org/2022/12/e36729
doi:10.2196/36729
PMID:36583929

Original Paper

Association Between Acoustic Features and Neuropsychological Test Performance in the Framingham Heart Study: Observational Study

Huitong Ding[1], PhD; Amiya Mandapati[2,3], BA; Cody Karjadi[1,4], MSc; Ting Fang Alvin Ang[1,4,5,6], MPH, MD; Sophia Lu[5], BS; Xiao Miao[7], MD; James Glass[8], PhD; Rhoda Au[1,4,5,6,9], MBA, PhD; Honghuang Lin[10], PhD

[1]Department of Anatomy and Neurobiology, Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, United States
[2]Department of Religious Studies, Brown University, Providence, RI, United States
[3]The Warren Alpert Medical School, Brown University, Providence, RI, United States
[4]The Framingham Heart Study, Chobanian & Avedisian School of Medicine, Boston University, Framingham, MA, United States
[5]Slone Epidemiology Center, Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, United States
[6]Department of Epidemiology, School of Public Health, Boston University, Boston, MA, United States
[7]Innovation Research Institute of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China
[8]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, United States
[9]Department of Neurology, Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, United States
[10]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

Corresponding Author:
Honghuang Lin, PhD
Department of Medicine
University of Massachusetts Chan Medical School
55 N Lake Ave
Worcester, MA, 01655
United States
Phone: 1 7744554881
Email: honghuang.lin@umassmed.edu

Abstract

Background: Human voice has increasingly been recognized as an effective indicator for the detection of cognitive disorders. However, the association of acoustic features with specific cognitive functions and mild cognitive impairment (MCI) has yet to be evaluated in a large community-based population.

Objective: This study aimed to investigate the association between acoustic features and neuropsychological (NP) tests across multiple cognitive domains and evaluate the added predictive power of acoustic composite scores for the classification of MCI.

Methods: This study included participants without dementia from the Framingham Heart Study, a large community-based cohort with longitudinal surveillance for incident dementia. For each participant, 65 low-level acoustic descriptors were derived from voice recordings of NP test administration. The associations between individual acoustic descriptors and 18 NP tests were assessed with linear mixed-effect models adjusted for age, sex, and education. Acoustic composite scores were then built by combining acoustic features significantly associated with NP tests. The added prediction power of acoustic composite scores for prevalent and incident MCI was also evaluated.

Results: The study included 7874 voice recordings from 4950 participants (age: mean 62, SD 14 years; 4336/7874, 55.07% women), of whom 453 were diagnosed with MCI. In all, 8 NP tests were associated with more than 15 acoustic features after adjusting for multiple testing. Additionally, 4 of the acoustic composite scores were significantly associated with prevalent MCI and 7 were associated with incident MCI. The acoustic composite scores can increase the area under the curve of the baseline model for MCI prediction from 0.712 to 0.755.

Conclusions: Multiple acoustic features are significantly associated with NP test performance and MCI, which can potentially be used as digital biomarkers for early cognitive impairment monitoring.

(J Med Internet Res 2022;24(12):e42886) doi:10.2196/42886

KEYWORDS

mild cognitive impairment; digital voice; neuropsychological test; association; prediction

Introduction

Alzheimer disease (AD) is a chronic neurodegenerative disease characterized behaviorally by memory loss, language impairment, motor problems, loss of executive function, and emotional distress, which can progress to severe levels. There are currently no definitive disease-modifying treatment methods [1], but general consensus is that early detection is critical. Interventions through the reduction of modifiable risk factors may serve to delay, attenuate, or even prevent disease onset and progression [2,3]. Mild cognitive impairment (MCI) is a prodromal stage of AD in which cognitive decline does not affect essential functions of daily life [4], but some individuals may have difficulty remembering events and situations, as well as problems with executive function [5]. The detection of MCI is critical to initiate current interventions that may slow down the neurodegenerative process [6] and participate in clinical trials that may lead to effective treatments.

At present, diagnosis relies largely on some combination of clinical examination [7], neuroimaging (eg, magnetic resonance imaging [8] and positron emission tomography [9]), and neuropsychological (NP) testing [10]. Fluid biomarkers are being developed as alternatives to expensive and burdensome imaging through the analysis of cerebrospinal fluid [11] and blood analysis [12]. Although substantial advancements have been made in developing pathological indicators of AD (eg, imaging and fluid biomarkers), surprisingly little has been done to develop better cognitive assessment methods beyond the traditional NP tests. The well-documented heterogeneity of cognition has made the accurate diagnosis of MCI elusive [13,14].

Producing speech is a cognitively complex task [15], and recording speech is relatively easy given the widespread accessibility to recording devices. Research has found that language deficits may occur in the prodromal stages [16] of cognitive impairment, which present years prior to clinical diagnosis [17,18], potentially making it an effective indicator for MCI. Meanwhile, the development of speech feature extraction technology offers the possibility of quantifying voice signal properties from multiple dimensions. It empowers the comprehensive description of specific pathologies by voice features. The lexical, acoustic, and syntactic features extracted from the human voice have been shown to be significantly associated with dementia [19,20]. Using voice-based biomarkers as a screening method presupposes an economic solution for the early diagnosis of MCI. Increasing evidence suggests that the human voice could be used as a powerful resource to derive pathologically appropriate biomarkers for dementia. Multiple acoustic biomarkers have also been related to the future risk of dementia [21].

Applying the findings of earlier research to a general population, however, is difficult due to the small sample sizes and use of cognitive assessment protocols that are not sufficiently comprehensive. Further, voice analyses that include linguistic features are difficult to generalize to other languages. There remains a paucity of research determining the relationship between acoustic features and NP tests that span across multiple cognitive domains. In addition, a comprehensive characterization of acoustic features that are associated with incident MCI is warranted. The objective of this study was to investigate the association of acoustic features and different NP test scores across cognitive domains and how they compare in identifying prevalent and incident MCI in the Framingham Heart Study (FHS) community-based cohort.

Methods

Sample Selection

The original sample included 9253 observations from 5189 participants who completed at least one NP assessment that was voice recorded. A subset of participants had multiple recordings over the course of the study period. Each digital voice recording and the corresponding NP tests were treated as 1 observation. Exclusion criteria included those observations with missing education information (n=492), prevalent dementia (n=313), flagged as potential MCI but have not gone through dementia review (n=551), and those whose voice recording was less than 10 minutes in length (n=23).

Ethics Approval

The Institutional Review Board of the Boston University Medical Campus approved the procedures and protocols of the Framingham Heart Study (FHS is H-32132). All participants provided written informed consent.

NP Assessment

The details of FHS NP test administration have been reported previously [22]. Multiple cognitive domains are measured by 18 different tests [23-27] including verbal memory, verbal fluency, visual memory, attention and concentration, executive function, abstract reasoning, visuoperceptual organization, and language, as is illustrated in Table 1.

Table 1. Cognitive domain and corresponding neuropsychological (NP) tests.

Cognitive domain	NP test
Verbal memory	• Logical Memory—Immediate Recall • Logical Memory—Delayed Recall • Logical Memory—Recognition • Paired Associate Learning—Immediate Recall • Paired Associate Learning—Delayed Recall • Paired Associate Learning—Recognition
Visual memory	• Visual Reproduction—Immediate Recall • Visual Reproduction—Delayed Recall • Visual Reproduction—Recognition
Attention and concentration	• Digit Span—Forward • Trail Making Test A
Executive function	• Digit Span—Backward • Trail Making Test B
Abstract reasoning	• Similarities
Language	• Boston Naming Test—30-item version
Visuoperceptual organization	• Hooper Visual Organization Test
Verbal fluency	• Controlled Oral Word Association Test • Category Naming Test—Animal

Voice Recordings

Since 2005, the FHS has been digitally recording all spoken responses during NP test administration, which encompasses the verbal interactions between the tester and the participant. This study included digital voice recordings obtained from September 2005 to March 2020. OpenSMILE software (version 2.1.3) [28] was used to extract an acoustic feature set [29], which contains 65 low-level descriptors (LLDs) from these recordings. This acoustic feature set covers a broad range of information of the voice recordings including pitch, voice quality, loudness, signal energy, waveform, auditory, fast Fourier transform spectrum, spectral, and cepstral, which has been described in detail in a prior study [30]. The feature set has also been used in many fields, such as speech processing, music information retrieval, and emotion recognition [31]. The description of these features is summarized in Table S1 in Multimedia Appendix 1. More details of these features can be found in the previous publication [30]. There are some audio recordings with 1 channel (mono; n=4738), and the others were recorded with 2 channels (stereo; n=3136). For the recordings with 2 channels, we included the first channel in the analysis. Each recording was divided into segments of 20 milliseconds using a sliding window approach with a shifting size of 10 milliseconds. The LLD features were extracted from these segments. For each recording, we further computed the mean of each LLD feature to capture its high-level statistical features, which were then normalized.

Ascertainment of MCI

The cognitive status of FHS participants included assessments by NP tests. For those identified with possible cognitive impairment, NP tests were administered on average about every 1 to 2 years. When potential cognitive impairment decline was present, a clinical review was conducted by a panel with at least one neurologist and one neuropsychologist. MCI diagnosis was determined by the review panel, which required that the participant exhibit evidence of a decline in cognitive performance in 1 or more cognitive domains, have no records indicating functional decline, and do not meet the criteria for dementia [32]. The Clinical Dementia Rating scale [33] was used to quantify the severity of impairment. In all, 2 outcomes were considered in this study. The prevalent MCI cases were subjects who were diagnosed with MCI before or at the time when the voice was recorded. The incident MCI cases were all subjects who were cognitively intact at baseline but were diagnosed with MCI during the follow-up.

Statistical Analyses

To compare the difference between demographics and standard NP test scores in MCI and normal control groups, Wilcoxon rank sum test was used for continuous variables [34]. The chi-square test was used to compare differences in frequencies for categorical variables [35]. Log transformations were applied for NP tests with skewed distributions to normalize them. Normalized values of NP tests and acoustic features were used in the analysis. Linear mixed-effects models were used to quantify the association between each acoustic feature and NP tests [36].

A set of acoustic composite scores was generated by regressing each NP test against the group of acoustic features that were significantly associated with each NP test. The acoustic composite score is a weighted combination of acoustic features. The weight of each acoustic feature in the composite score was derived by training a linear mixed-effects effect model. For

participant *i*, the acoustic composite score of an NP test is defined as

where *m* is the number of acoustic features significantly associated with the NP test, α_j is the estimate of effect size for the acoustics feature *j* derived from the linear mixed-effects effect model, and V_{ij} is the normalized acoustics feature *j* for participant *i*. The association between normalized acoustic composite scores with corresponding NP tests was assessed by linear mixed-effects models.

The association of normalized acoustic composite scores with prevalent MCI was assessed by logistic regression models. Based on the regression coefficients, the odds ratios (ORs) and 95% CIs were estimated.

To determine the relationship between acoustic composite scores and incident MCI, participants whose age at the voice recording was <60 years (n=2718) and those with prevalent MCI (n=222) were excluded. The first observation of each participant was included in this analysis. The association between acoustic composite scores with incident MCI was quantified by Cox proportional hazards models (censored at the last date of contact or death) [37]. All models were adjusted for age, sex, and education. Bonferroni correction was used to adjust for multiple tests.

We further evaluated the added predictability of the acoustic composite score for incident MCI. Receiver operating characteristic (ROC) analysis was performed to estimate the area under the curve (AUC) using a random forest model. A baseline model was constructed using age, sex, and education as predictors. A second model was constructed using these predictors and additional acoustic composite scores that were found to be significantly related to specific NP tests. The mean AUC of 10-fold cross-validation was computed for each model for comparison. We also performed a secondary analysis by including NP tests and clinical risk factors in the prediction of incident MCI. The statistical analyses were performed using Python software (version 3.9.7; Python Software Foundation).

Results

Our study included 7874 observations from 4950 participants of FHS (age: mean 62, SD 14 years; 4336/7874, 55.07% women; 4279/7874, 54.34% self-reported college-level education or higher). Most participants (2657/4950, 53.68%) had 1 voice recording. Some participants (1775/4950, 35.86%) had 2 recordings, and the remaining participants (518/4950, 10.46%) had 3 or more recordings. Among these observations, 453 of these observations were diagnosed with MCI. The details of sample characteristics are shown in Table 2.

We examined the association of acoustic features with NP tests. As shown in Table 3, eight NP tests (Visual Reproduction—Immediate Recall [VRi], Visual Reproduction—Delayed Recall [VRd], Digit Span—Forward, Digit Span—Backward, Similarities [SIM], Boston Naming Test—30-item version, Controlled Oral Word Association Test [FAS], and Category Naming Test—Animal) were associated with more than 15 acoustic features. The *mfcc_sma* [2] was the most significant acoustic feature with 3 NP tests (Boston Naming Test—30-item version, FAS, and Category Naming Test—Animal) after Bonferroni correction ($P<7.7 \times 10^{-4}$) that represents Mel-frequency cepstral coefficient (MFCC) 2. The details of associations between acoustic features and NP tests are fully depicted in Table S2 in Multimedia Appendix 1. We also summarized the acoustic features that were significantly associated with NP tests across cognitive domains in Table S3 in Multimedia Appendix 1. It shows that visual memory was associated with 49 acoustic features. Each cognitive domain had an average of 28 associated acoustic features. In the sensitivity analysis, besides age, sex, and education, we further included employment as an additional covariate to examine the stability of the association between acoustic features and NP tests. As shown in Table S4 in Multimedia Appendix 1, similar acoustic features were found to be associated with NP tests. In addition, we also examined the correlation between acoustic features and NP tests collected at the same time and a later time. For each NP test conducted at the first exam, we compared its correlation with acoustic features collected at the first exam and the second exam. As shown in Table S5 in Multimedia Appendix 1, only moderate changes were observed between the 2 exams.

Acoustic composite scores were also generated using the significant acoustic features for each NP test. As shown in Table 4, all these scores were significantly associated with their corresponding NP tests.

We then performed association analysis of acoustic composite scores with prevalent MCI. Table 5 shows that 4 acoustic composite scores (*acoustic_LMr*, *acoustic_TrailsB*, *acoustic_FAS*, and *acoustic_CNT_Animal*) were significantly associated with prevalent MCI (OR ranging from 0.69 to 1.23; $P<3.1 \times 10^{-3}$). Lower acoustics composite scores (*acoustic_TrailsB*, *acoustic_FAS*, and *acoustic_CNT_Animal*) were associated with higher OR of MCI after adjusting for age, sex, and education ($P<3.1 \times 10^{-3}$). The most significant acoustic composite score was for FAS Animal test ($P=2.3 \times 10^{-7}$).

We further examined the association of acoustic composite scores with incident MCI by restricting the analysis to 2010 participants who were aged ≥60 years. Among them, 145 participants have incident MCI. As shown in Table 6, the acoustic composite scores for Logical Memory—Immediate Recall (LMi), VRi, VRd, Visual Reproduction—Recognition (VRr), SIM, Trail Making Test B (TrailsB), and Hooper Visual Organization Test (HVOT) tests were significantly associated with incident MCI ($P<3.1 \times 10^{-3}$). Higher acoustic composite scores for VRi, VRd, SIM, and TrailsB tests were associated with higher MCI risk. The other 3 scores were negatively associated with MCI risk with hazard ratio lower than 1 after adjusting for age, sex, and education. We further built 2 Cox regression models for incident MCI to show the contribution of acoustic features. Model 1 includes age, sex, and education as predictors. Model 2 includes age, sex, education, and all significant associated acoustic composite scores with incident MCI. The change in Akaike information criterion [38] with the addition of acoustic composite scores to the model was

calculated. We observed a smaller Akaike information criterion for model 2, suggesting that the model better fit the prediction.

The added predictive power of *acoustic_LMi*, *acoustic_VRi*, *acoustic_VRd*, *acoustic_VRr*, *acoustic_SIM*, *acoustic_TrailsB*, and *acoustic_HVOT* for incident MCI were evaluated by comparing the AUC of different models. Model 1 only included age, sex, and education as the predictors of incident MCI. Besides age, sex, and education, Model 2 included 7 composite scores that were significantly associated with incident MCI as the predictors. Model 3 included age, sex, education, and 18 NP tests as predictors. Figure 1 shows that the AUC of MCI prediction can be improved from 0.712 (model 1) to 0.755 (model 2) by including acoustic composite scores of LMi, VRi, VRd, VRr, SIM, TrailsB, and HVOT tests. As shown in Figure S1 in Multimedia Appendix 1, the model with NP tests reached AUC=0.761, which is comparable to the one including demographic factors and acoustic composite scores (DeLong test *P*=.97). However, both models showed significant improvement over model 1 that included only demographic factors (DeLong test *P*=.03 and *P*=.03 for model 2 and model 3, respectively). These results indicate that the acoustics composite scores have similar predictive power to traditional NP tests. Compared to the burden of conducting NP tests, the prediction model based on acoustic features relied minimally on NP expertise; these results suggest the feasibility of developing real-time cognitive screening tools.

Table 2. Baseline characteristics.

Variable	Total observation (N=7874)	MCI[a] (n=453)	NC[b] (n=7421)	*P* value[c]
Age (years), mean (SD)	62 (14)	81 (8)	61 (14)	<.001
Gender, n (%)				.84
Women	4336 (55.07)	252 (55.63)	4084 (55.03)	
Men	3538 (44.93)	201 (44.37)	3337 (44.97)	
Education, n (%)				<.001
No high school	202 (2.57)	53 (11.70)	149 (2.01)	
High school	1443 (18.33)	118 (26.05)	1295 (17.45)	
Some college	1950 (24.77)	134 (29.58)	1816 (24.47)	
College and higher	4279 (54.34)	148 (32.67)	4161 (56.07)	
NP[d] test score, mean (SD)				
LMi[e]	12.35 (3.62)	8.53 (3.76)	12.58 (3.48)	<.001
LMd[f]	11.36 (3.83)	6.93 (4.11)	11.62 (3.65)	<.001
LMr[g]	9.52 (1.28)	8.59 (1.72)	9.57 (1.23)	<.001
VRi[h]	8.61 (2.91)	4.48 (2.23)	8.85 (2.76)	<.001
VRd[i]	7.91 (3.17)	3.11 (2.30)	8.19 (2.99)	<.001
VRr[j]	3.11 (1.01)	1.89 (1.06)	3.18 (0.96)	<.001
PASi[k]	14.45 (3.58)	10.02 (2.79)	14.71 (3.45)	<.001
PASd[l]	8.56 (1.47)	6.56 (1.60)	8.68 (1.38)	<.001
PASr[m]	9.82 (0.64)	8.83 (1.74)	9.88 (0.45)	<.001
DSf[n]	6.71 (1.31)	6.06 (1.20)	6.75 (1.30)	<.001
DSb[o]	4.92 (1.30)	4.12 (1.01)	4.97 (1.30)	<.001
SIM[p]	16.83 (3.61)	12.63 (4.30)	17.08 (3.40)	<.001
BNT30[q]	27.22 (2.81)	23.66 (4.14)	27.43 (2.56)	<.001
TrailsA[r]	0.42 (0.15)	0.66 (0.21)	0.40 (0.14)	<.001
TrailsB[s]	0.85 (0.34)	1.54 (0.50)	0.82 (0.29)	<.001
HVOT[t]	3.26 (0.15)	3.06 (0.22)	3.27 (0.13)	<.001
FAS[u]	39.85 (12.52)	28.76 (11.68)	40.50 (12.26)	<.001
CNT_Animal[v]	19.48 (5.68)	12.22 (4.37)	19.91 (5.46)	<.001

[a]MCI: mild cognitive impairment.
[b]NC: normal control.
[c]Significant associations were claimed if $P<.05/18\approx.002$.
[d]NP: neuropsychological.
[e]LMi: Logical Memory—Immediate Recall.
[f]LMd: Logical Memory—Delayed Recall.
[g]LMr: Logical Memory—Recognition.
[h]VRi: Visual Reproduction—Immediate Recall.
[i]VRd: Visual Reproduction—Delayed Recall.
[j]VRr: Visual Reproduction—Recognition.
[k]PASi: Paired Associate Learning—Immediate Recall.
[l]PASd: Paired Associate Learning—Delayed Recall.
[m]PASr: Paired Associate Learning—Recognition.

[n]DSf: Digit Span—Forward.
[o]DSb: Digit Span—Backward.
[p]SIM: Similarities.
[q]BNT30: Boston Naming Test—30-item version.
[r]TrailsA: Trail Making Test A.
[s]TrailsB: Trail Making Test B.
[t]HVOT: Hooper Visual Organization Test.
[u]FAS: Controlled Oral Word Association Test.
[v]CNT_Animal: Category Naming Test—Animal.

Table 3. The most significant acoustic feature for each neuropsychological (NP) test.

NP test	Significant acoustic features, n	The most significant acoustic feature	Effect size	SE	*P* value[a]
LMi[b]	7	audSpec_Rfilt_sma [25]	0.0490	0.0095	2.7×10^{-7}
LMd[c]	3	audSpec_Rfilt_sma [25]	0.0402	0.0094	1.9×10^{-5}
LMr[d]	3	audSpec_Rfilt_sma [23]	0.0397	0.0108	2.3×10^{-4}
VRi[e]	49	mfcc_sma [11]	0.1409	0.0082	8.4×10^{-66}
VRd[f]	43	mfcc_sma [11]	0.1137	0.0082	3.7×10^{-44}
VRr[g]	10	pcm_fftMag_spectralRollOff75.0_sma	–0.0358	0.0095	1.7×10^{-4}
PASi[h]	0	N/A[i]	N/A	N/A	N/A
PASd[j]	0	N/A	N/A	N/A	N/A
PASr[k]	7	audSpec_Rfilt_sma [1]	–0.0709	0.0112	2.3×10^{-10}
DSf[l]	44	audSpec_Rfilt_sma [6]	0.0898	0.0107	4.8×10^{-17}
DSb[m]	30	audSpec_Rfilt_sma [5]	0.0624	0.0110	1.2×10^{-8}
SIM[n]	24	pcm_fftMag_spectralRollOff75.0_sma	–0.0530	0.0084	2.4×10^{-10}
BNT30[o]	23	mfcc_sma [2]	0.0433	0.0069	3.2×10^{-10}
TrailsA[p]	15	pcm_fftMag_spectralSkewness_sma	–0.0363	0.0075	1.4×10^{-6}
TrailsB[q]	1	pcm_fftMag_spectralSkewness_sma	–0.0269	0.0074	3.1×10^{-4}
HVOT[r]	5	F0final_sma	–0.0472	0.0093	3.6×10^{-7}
FAS[s]	26	mfcc_sma [2]	0.0534	0.0073	3.6×10^{-13}
CNT_Animal[t]	34	mfcc_sma [2]	0.0715	0.0082	2.6×10^{-18}

[a]Significant associations were claimed if $P<.05/65\approx7.7 \times 10^{-4}$.
[b]LMi: Logical Memory—Immediate Recall.
[c]LMd: Logical Memory—Delayed Recall.
[d]LMr: Logical Memory—Recognition.
[e]VRi: Visual Reproduction—Immediate Recall.
[f]VRd: Visual Reproduction—Delayed Recall.
[g]VRr: Visual Reproduction—Recognition.
[h]PASi: Paired Associate Learning—Immediate Recall.
[i]N/A: not applicable.
[j]PASd: Paired Associate Learning—Delayed Recall.
[k]PASr: Paired Associate Learning—Recognition.
[l]DSf: Digit Span—Forward.
[m]DSb: Digit Span—Backward.
[n]SIM: Similarities.
[o]BNT30: Boston Naming Test—30-item version.
[p]TrailsA: Trail Making Test A.
[q]TrailsB: Trail Making Test B.
[r]HVOT: Hooper Visual Organization Test.
[s]FAS: Controlled Oral Word Association Test.
[t]CNT_Animal: Category Naming Test—Animal.

Table 4. Association between acoustic composite scores and corresponding neuropsychological tests.

Acoustic composite score	Effect size	SE	*P* value[a]
acoustic_LMi[b]	0.0579	0.0094	6.6×10^{-10}
acoustic_LMd[c]	0.0310	0.0095	1.1×10^{-3}
acoustic_LMr[d]	0.0358	0.0105	6.8×10^{-4}
acoustic_VRi[e]	0.1510	0.0086	3.3×10^{-69}
acoustic_VRd[f]	0.1079	0.0086	6.5×10^{-36}
acoustic_VRr[g]	–0.0291	0.0098	3.0×10^{-3}
acoustic_PASr[h]	0.0841	0.0114	1.3×10^{-13}
acoustic_DSf[i]	0.1298	0.0097	1.8×10^{-40}
acoustic_DSb[j]	0.0553	0.0102	6.2×10^{-8}
acoustic_SIM[k]	0.0719	0.0089	5.1×10^{-16}
acoustic_BNT30[l]	0.0458	0.0071	1.4×10^{-10}
acoustic_TrailsA[m]	0.0408	0.0088	3.0×10^{-6}
acoustic_TrailsB[n]	–0.0269	0.0075	3.1×10^{-4}
acoustic_HVOT[o]	0.0284	0.0090	1.7×10^{-3}
acoustic_FAS[p]	0.0827	0.0079	1.4×10^{-25}
acoustic_CNT_Animal[q]	0.0529	0.0098	6.5×10^{-8}

[a]Significant associations were claimed if $P<.05/16\approx3.1 \times 10^{-3}$.
[b]LMi: Logical Memory—Immediate Recall.
[c]LMd: Logical Memory—Delayed Recall.
[d]LMr: Logical Memory—Recognition.
[e]VRi: Visual Reproduction—Immediate Recall.
[f]VRd: Visual Reproduction—Delayed Recall.
[g]VRr: Visual Reproduction—Recognition.
[h]PASr: Paired Associate Learning—Recognition.
[i]DSf: Digit Span—Forward.
[j]DSb: Digit Span—Backward.
[k]SIM: Similarities.
[l]BNT30: Boston Naming Test—30-item version.
[m]TrailsA: Trail Making Test A.
[n]TrailsB: Trail Making Test B.
[o]HVOT: Hooper Visual Organization Test.
[p]FAS: Controlled Oral Word Association Test.
[q]CNT_Animal: Category Naming Test—Animal.

Table 5. Association between acoustic composite scores and prevalent mild cognitive impairment.

Acoustic composite score	Odds ratio (95% CI)	*P* value[a]
acoustic_LMi[b]	1.09 (0.94-1.26)	2.6×10^{-1}
acoustic_LMd[c]	1.14 (0.99-1.31)	7.4×10^{-2}
acoustic_LMr[d]	1.23 (1.08-1.40)	1.6×10^{-3}
acoustic_VRi[e]	1.05 (0.92-1.19)	4.7×10^{-1}
acoustic_VRd[f]	1.07 (0.94-1.21)	3.2×10^{-1}
acoustic_VRr[g]	0.94 (0.80-1.10)	4.6×10^{-1}
acoustic_PASr[h]	0.9 (0.81-0.99)	3.6×10^{-2}
acoustic_DSf[i]	1.17 (1.04-1.32)	1.1×10^{-2}
acoustic_DSb[j]	0.94 (0.83-1.07)	3.5×10^{-1}
acoustic_SIM[k]	0.94 (0.84-1.06)	3.1×10^{-1}
acoustic_BNT30[l]	0.92 (0.82-1.04)	2.0×10^{-1}
acoustic_TrailsA[m]	1.12 (0.98-1.28)	9.6×10^{-2}
acoustic_TrailsB[n]	0.69 (0.59-0.81)	1.0×10^{-5}
acoustic_HVOT[o]	0.91 (0.81-1.03)	1.4×10^{-1}
acoustic_FAS[p]	0.72 (0.64-0.81)	3.9×10^{-8}
acoustic_CNT_Animal[q]	0.70 (0.61-0.80)	2.3×10^{-7}

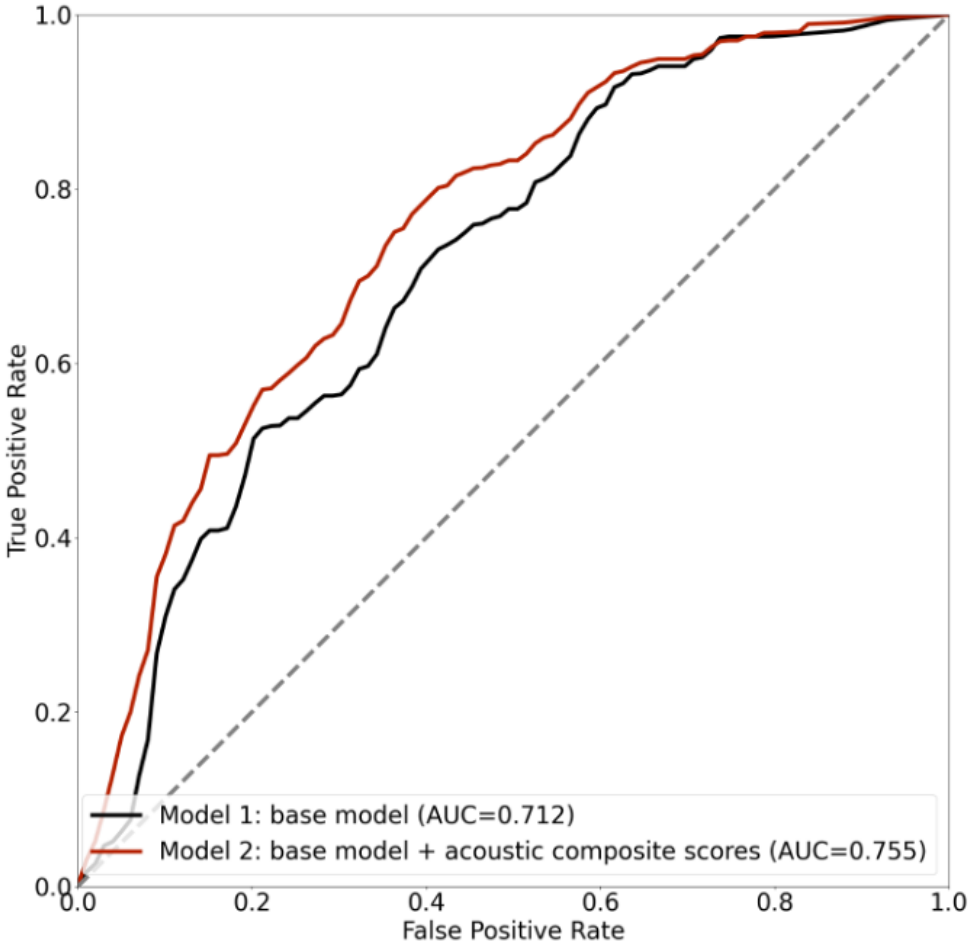
[a]Significant associations were claimed if $P<.05/16\approx3.1 \times 10^{-3}$.
[b]LMi: Logical Memory—Immediate Recall.
[c]LMd: Logical Memory—Delayed Recall.
[d]LMr: Logical Memory—Recognition.
[e]VRi: Visual Reproduction—Immediate Recall.
[f]VRd: Visual Reproduction—Delayed Recall.
[g]VRr: Visual Reproduction—Recognition.
[h]PASr: Paired Associate Learning—Recognition.
[i]DSf: Digit Span—Forward.
[j]DSb: Digit Span—Backward.
[k]SIM: Similarities.
[l]BNT30: Boston Naming Test—30-item version.
[m]TrailsA: Trail Making Test A.
[n]TrailsB: Trail Making Test B.
[o]HVOT: Hooper Visual Organization Test.
[p]FAS: Controlled Oral Word Association Test.
[q]CNT_Animal: Category Naming Test—Animal.

Table 6. Association between acoustic composite scores and incident mild cognitive impairment.

Acoustic composite score	Hazard ratio (95% CI)	*P* value[a]
acoustic_LMi[b]	0.60 (0.47-0.77)	5.1×10^{-5}
acoustic_LMd[c]	0.76 (0.59-0.97)	2.9×10^{-2}
acoustic_LMr[d]	0.74 (0.61-0.91)	3.9×10^{-3}
acoustic_VRi[e]	1.28 (1.10-1.48)	1.1×10^{-3}
acoustic_VRd[f]	1.25 (1.08-1.44)	2.4×10^{-3}
acoustic_VRr[g]	0.44 (0.33-0.59)	6.0×10^{-8}
acoustic_PASr[h]	1.11 (0.95-1.30)	2.0×10^{-1}
acoustic_DSf[i]	1.11 (0.96-1.29)	1.6×10^{-1}
acoustic_DSb[j]	1.09 (0.93-1.27)	2.9×10^{-1}
acoustic_SIM[k]	1.37 (1.16-1.61)	1.7×10^{-4}
acoustic_BNT30[l]	1.23 (1.06-1.43)	6.4×10^{-3}
acoustic_TrailsA[m]	0.75 (0.61-0.93)	7.9×10^{-3}
acoustic_TrailsB[n]	2.03 (1.58-2.60)	2.5×10^{-8}
acoustic_HVOT[o]	0.78 (0.67-0.91)	1.7×10^{-3}
acoustic_FAS[p]	0.87 (0.76-1.01)	6.1×10^{-2}
acoustic_CNT_Animal[q]	0.85 (0.70-1.02)	8.6×10^{-2}

[a]Significant associations were claimed if $P<.05/16\approx3.1 \times 10^{-3}$.
[b]LMi: Logical Memory—Immediate Recall.
[c]LMd: Logical Memory—Delayed Recall.
[d]LMr: Logical Memory—Recognition.
[e]VRi: Visual Reproduction—Immediate Recall.
[f]VRd: Visual Reproduction—Delayed Recall.
[g]VRr: Visual Reproduction—Recognition.
[h]PASr: Paired Associate Learning—Recognition.
[i]DSf: Digit Span—Forward.
[j]DSb: Digit Span—Backward.
[k]SIM: Similarities.
[l]BNT30: Boston Naming Test—30-item version.
[m]TrailsA: Trail Making Test A.
[n]TrailsB: Trail Making Test B.
[o]HVOT: Hooper Visual Organization Test.
[p]FAS: Controlled Oral Word Association Test.
[q]CNT_Animal: Category Naming Test—Animal.

Figure 1. The receiver operating characteristic curves of 2 models to predict incident mild cognitive impairment. AUC: area under the curve.



Discussion

Principal Findings

Relating acoustic features with NP test performance is potentially a novel way for screening at the preclinical stages of AD and other dementias. This paper clarifies the relationship between comprehensive acoustic features and NP test performance on large cohort data. Representations relative spectra–style filtered auditory spectrum (spectral), MFCC (cepstral), and magnitude of spectral features (spectral) are 3 categories of acoustic features that were significantly associated with NP test performance. Representations relative spectra–style filtered auditory spectrum is a filtered representation of an audio signal that is robust to additive and convolutional noise [39]. MFCC is a standardized technique for audio feature extraction [40]. It helps in reducing the frequency information of the input speech signal into coefficients, which represent audio based on the perception of human auditory systems. Prior studies have detected changes of these features in people with neurodegenerative processes [41-43]. The acoustic composite score generated for each NP test was a linear combination of LLD features, which are clinically easily interpretable. As stated in the results above, 4 acoustic composite scores were significantly associated with prevalent MCI, and 7 were also found to be significantly associated with incident MCI. Furthermore, the score corresponding to TrailsB test is significantly associated with both prevalent MCI and incident MCI.

Results could expand current evidence regarding the predictive ability of digital voice on MCI that are critical to monitor early cognitive decline. The added predictive ability of acoustic features was evaluated by constructing random forest models with baseline features and additional acoustic composite scores. The model with baseline features and 7 acoustic composite scores corresponding to LMi, VRi, VRd, VRr, SIM, TrailsB, and HVOT tests could achieve an AUC of 0.755 for incident MCI prediction. Monitoring acoustic features outside of the clinical settings offers a more convenient way to aid in the assessment of cognitive health than traditional methods. Increasing evidence suggests that the human voice can be a predictor of cognitive decline before a clinical diagnosis of AD is made [44]. It has been used to screen for MCI [45], dementia [46], and other neurodegenerative diseases such as Parkinson [47] and Huntington disease [48] because of its ease of administration and clinical assessment capability. Moreover, the easy acquisition of voice in daily life makes it an ideal measure for long-term monitoring of cognitive status. However, there is a lack of research about the relationship between acoustic features and NP tests that reflect multiple cognitive functions. Our study could provide some construct validity for this point. In this study, we recorded voice for NP tests that require verbal responses. Although some NP tests do not require verbal responses, these tests might tap some cognitive domains

similar to those that require verbal responses. We therefore included these tests as well to capture potential application of acoustic characteristics to assess different cognitive domains. Each NP test might require multiple cognitive domains to complete, which might be shared with other NP tests with subtle differences. Given the rich information from human voice, our study suggests that acoustic features might serve as a new data modality to test this nuance.

Notably, the association between acoustic features and a standard epidemiologic NP test procedure was examined based on participants from a community-based cohort with a diverse range of ages and health conditions. The large volume of voice data provides a more robust representation of participants. Each voice recording lasts, on average, around an hour and contains a wealth of information. The longitudinal collection of data provides a great opportunity to assess the cognitive health of participants throughout the entire course of the disease and prospectively reveals a temporal relationship between acoustic features and MCI. It is worth to noting that 4 of the acoustic composite scores (*acoustic_LMr*, *acoustic_TrailsB*, *acoustic_FAS*, and *acoustic_CNT_Animal*) were significantly associated with prevalent MCI, but 7 acoustic composite scores (*acoustic_LMi*, *acoustic_VRi*, *acoustic_VRd*, *acoustic_VRr*, *acoustic_SIM*, *acoustic_TrailsB*, and *acoustic_HVOT*) were associated with incident MCI. It seems that the voice characteristics differentiating prevalent MCI cases from patients who are still cognitively intact are different from the voice characteristics that are predictive of future risk of cognitive impairment. Future research is needed to further investigate the potential mechanisms that underlie these features and help to account for the MCI prevalence and incidence difference. Further, this study found differences in acoustic features between TrailsA and TrailsB, which provides confirmatory evidence that acoustic features are differential for different cognitive domains. TrailsA, as a measure of simple attention compared to the more complex executive functions measured by TrailsB [49], would be expected to have different acoustic features that would be aligned with motor control and perceptual complexity [50] in the latter and not the former. These differential results suggest that acoustic features might provide a way to detect such subtle differences across cognitive domains. The patterns of acoustic features that are accurately representative of the comprehensive range of cognitive domains will be further explored in future studies.

This study also has some limitations. First, the use of NP tests to diagnose MCI may have led to some circularity and an overestimation of the diagnosis performance [32]. Second, despite that diagnoses are arrived at through a careful adjudication process, there may be some misclassification of MCI. Third, although the FHS collected the voice recordings in a well-controlled environment, there might still be some other factors affecting the quality of voice that were not taken into account. Fourth, this study did not consider linguistic features, which has been shown to be effective in predicting cognitive status. Although we recognize that the inclusion of linguistic features might further improve the prediction of incident MCI, we chose to focus on acoustic features because they are much more generalizable to a broader population, including potentially to other languages. Linguistic features are much more likely to be biased by language, culture, and education. Finally, FHS participants were mostly of European ancestry and English speakers; therefore, the applicability of our findings to populations of another race and ethnicity needs to be examined.

Conclusion

We examined the association of acoustic features with specific cognitive functions—prevalent and incident MCI—in a large community-based population. Overall, this study's establishment of a relationship between MCI risk and human voice features provides foundational evidence for an alternative cognitive assessment approach that is cost-effective and easy to administer for detecting cognition-related disorders. Multiple acoustic features were significantly associated with NP test performance and MCI and could be potentially used as a digital biomarker for early cognitive impairment monitoring.

Acknowledgments

We acknowledge the Framingham Heart Study (FHS) participants for their dedication. This study would not be possible without them. We also thank the researchers in the FHS for their efforts over the years in the examination of subjects.

This work was supported by the National Heart, Lung, and Blood Institute (contract N01-HC-25195) and by grants from the National Institute on Aging (AG-008122, AG-16495, AG-062109, AG-049810, AG-068753, AG054156, and U01AG068221) and the National Institute of Neurological Disorders and Stroke (NS017950). It was also supported by Defense Advanced Research Projects Agency (contract FA8750-16-C-0299) and Pfizer, Inc. This work was also supported by the grants from the Alzheimer's Association (AARG-NTF-20-643020) and the American Heart Association (20SFRN35360180). The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Institutes of Health or the US Department of Health and Human Services. The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The derived acoustic features could be requested through a formal research application to the Framingham Heart Study [51].

Conflicts of Interest

RA is a scientific advisor to Signant Health and a consultant to Biogen and the Davos Alzheimer's Collaborative.

Multimedia Appendix 1
Supplemental tables and figure.
[DOCX File , 288 KB - jmir_v24i12e42886_app1.docx]

References

1. Sang Z, Wang K, Dong J, Tang L. Alzheimer's disease: updated multi-targets therapeutics are in clinical and in progress. Eur J Med Chem 2022 Aug 05;238:114464. [doi: 10.1016/j.ejmech.2022.114464] [Medline: 35635955]
2. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 08;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
3. Rosenberg A, Mangialasche F, Ngandu T, Solomon A, Kivipelto M. Multidomain interventions to prevent cognitive impairment, Alzheimer's disease, and dementia: from FINGER to World-Wide FINGERS. J Prev Alzheimers Dis 2020 Oct 10;7(1):29-36 [FREE Full text] [doi: 10.14283/jpad.2019.41] [Medline: 32010923]
4. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. Lancet 2006 Apr 15;367(9518):1262-1270. [doi: 10.1016/S0140-6736(06)68542-5] [Medline: 16631882]
5. Themistocleous C, Eckerström M, Kokkinakis D. Identification of mild cognitive impairment from speech in Swedish using deep sequential neural networks. Front Neurol 2018 Nov 15;9:975 [FREE Full text] [doi: 10.3389/fneur.2018.00975] [Medline: 30498472]
6. Morrison MS, Aparicio HJ, Blennow K, Zetterberg H, Ashton NJ, Karikari TK, et al. Ante-mortem plasma phosphorylated tau (181) predicts Alzheimer's disease neuropathology and regional tau at autopsy. Brain 2022 Oct 21;145(10):3546-3557. [doi: 10.1093/brain/awac175] [Medline: 35554506]
7. Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007 Aug;6(8):734-746. [doi: 10.1016/S1474-4422(07)70178-3] [Medline: 17616482]
8. Chincarini A, Bosco P, Calvini P, Gemme G, Esposito M, Olivieri C, Alzheimer's Disease Neuroimaging Initiative. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. Neuroimage 2011 Sep 15;58(2):469-480. [doi: 10.1016/j.neuroimage.2011.05.083] [Medline: 21718788]
9. Chételat G, Arbizu J, Barthel H, Garibotto V, Law I, Morbelli S, et al. Amyloid-PET and F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias. Lancet Neurol 2020 Nov;19(11):951-962. [doi: 10.1016/S1474-4422(20)30314-8] [Medline: 33098804]
10. Ang TF, An N, Ding H, Devine S, Auerbach SH, Massaro J, et al. Using data science to diagnose and characterize heterogeneity of Alzheimer's disease. Alzheimers Dement (N Y) 2019 Jun 27;5(1):264-271 [FREE Full text] [doi: 10.1016/j.trci.2019.05.002] [Medline: 31304232]
11. Blennow K, Hampel H, Weiner M, Zetterberg H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat Rev Neurol 2010 Mar 16;6(3):131-144. [doi: 10.1038/nrneurol.2010.4] [Medline: 20157306]
12. Henriksen K, O'Bryant SE, Hampel H, Trojanowski JQ, Montine TJ, Jeromin A, Blood-Based Biomarker Interest Group. The future of blood-based biomarkers for Alzheimer's disease. Alzheimers Dement 2014 Jan 15;10(1):115-131 [FREE Full text] [doi: 10.1016/j.jalz.2013.01.013] [Medline: 23850333]
13. Ganguli M, Snitz BE, Saxton JA, Chang CH, Lee C, Vander Bilt J, et al. Outcomes of mild cognitive impairment by definition: a population study. Arch Neurol 2011 Jun 13;68(6):761-767 [FREE Full text] [doi: 10.1001/archneurol.2011.101] [Medline: 21670400]
14. Tabatabaei-Jafari H, Shaw ME, Cherbuin N. Cerebral atrophy in mild cognitive impairment: a systematic review with meta-analysis. Alzheimers Dement (Amst) 2015 Dec 11;1(4):487-504 [FREE Full text] [doi: 10.1016/j.dadm.2015.11.002] [Medline: 27239527]
15. Seraji-Bzorgzad N, Paulson H, Heidebrink J. Neurologic examination in the elderly. Handb Clin Neurol 2019;167:73-88 [FREE Full text] [doi: 10.1016/B978-0-12-804766-8.00005-4] [Medline: 31753158]
16. Cuetos F, Arango-Lasprilla JC, Uribe C, Valencia C, Lopera F. Linguistic changes in verbal expression: a preclinical marker of Alzheimer's disease. J Int Neuropsychol Soc 2007 May;13(3):433-439. [doi: 10.1017/S1355617707070609] [Medline: 17445292]
17. Deramecourt V, Lebert F, Debachy B, Mackowiak-Cordoliani MA, Bombois S, Kerdraon O, et al. Prediction of pathology in primary progressive language and speech disorders. Neurology 2010 Jan 05;74(1):42-49. [doi: 10.1212/WNL.0b013e3181c7198e] [Medline: 19940270]
18. Taler V, Phillips NA. Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. J Clin Exp Neuropsychol 2008 Jul 05;30(5):501-556. [doi: 10.1080/13803390701550128] [Medline: 18569251]

19. Le X, Lancashire I, Hirst G, Jokel R. Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. Literary and Linguistic Computing 2011 May 24;26(4):435-461. [doi: 10.1093/llc/fqr013]
20. Beltrami D, Gagliardi G, Rossini Favretti R, Ghidoni E, Tamburini F, Calzà L. Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline? Front Aging Neurosci 2018 Nov 13;10:369 [FREE Full text] [doi: 10.3389/fnagi.2018.00369] [Medline: 30483116]
21. Lin H, Karjadi C, Ang TFA, Prajakta J, McManus C, Alhanai TW, et al. Identification of digital voice biomarkers for cognitive health. Explor Med 2020 Dec 31;1:406-417 [FREE Full text] [doi: 10.37349/emed.2020.00028] [Medline: 33665648]
22. Au R, Seshadri S, Wolf PA, Elias MF, Elias PK, Sullivan L, et al. New norms for a new generation: cognitive performance in the framingham offspring cohort. Exp Aging Res 2004 Oct;30(4):333-358. [doi: 10.1080/03610730490484380] [Medline: 15371099]
23. Wechsler D, Stone CP. Wechsler Memory Scale (WMS). New York, NY: The Psychological Corporation; 1948.
24. Benton AL, Hamsher KD, Sivan AB. Multilingual Aphasia Examination. Iowa City, IA: University of Iowa; 1976.
25. Kaplan E, Goodglass H, Weintraub S. Boston Naming Test. Philadelphia, PA: Lea & Febi; 1983.
26. Spreen O, Strauss E. A Compendium of Neuropsychological Tests: Administration, Norms and Commentary. New York, NY: NY Oxford University Press; 1991.
27. Hooper HE. Hooper Visual Organization Test Manual. Los Angeles, CA: Western Psychological Services (WPS); 1958.
28. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor. 2010 Oct 25 Presented at: MM '10: the 18th ACM international conference on Multimedia; October 25-29, 2010; Firenze, Italy p. 1459-1462. [doi: 10.1145/1873951.1874246]
29. Schuller B, Steidl S, Batliner A, Hirschberg J, Burgoon JK, Baird A, et al. 2016 Presented at: Interspeech 2016; September 8-12, 2016; San Francisco, CA p. 2001-2005. [doi: 10.21437/interspeech.2016-129]
30. Weninger F, Eyben F, Schuller BW, Mortillaro M, Scherer KR. On the acoustics of emotion in audio: what speech, music, and sound have in common. Front Psychol 2013 May 27;4:292 [FREE Full text] [doi: 10.3389/fpsyg.2013.00292] [Medline: 23750144]
31. Tahon M, Devillers L. Towards a small set of robust acoustic features for emotion recognition: challenges. IEEE/ACM Trans Audio Speech Lang Process 2016 Jan;24(1):16-28. [doi: 10.1109/taslp.2015.2487051]
32. Yuan J, Libon DJ, Karjadi C, Ang AFA, Devine S, Auerbach SH, et al. Association between the digital clock drawing test and neuropsychological test performance: large community-based prospective cohort (Framingham Heart Study). J Med Internet Res 2021 Jun 08;23(6):e27407 [FREE Full text] [doi: 10.2196/27407] [Medline: 34100766]
33. Hughes CP, Berg L, Danziger W, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982 Jun 29;140(6):566-572. [doi: 10.1192/bjp.140.6.566] [Medline: 7104545]
34. Haynes W. Wilcoxon rank sum test. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H, editors. Encyclopedia of Systems Biology. New York, NY: Springer; 2013:2354-2355.
35. McHugh ML. The chi-square test of independence. Biochem Med (Zagreb) 2013;23(2):143-149 [FREE Full text] [doi: 10.11613/bm.2013.018] [Medline: 23894860]
36. Pinheiro JC, Bates DM. Linear mixed-effects models: basic concepts and examples. In: Mixed-Effects Models in S and S-Plus. New York, NY: Springer; 2000:3-56.
37. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. J Am Stat Assoc 1989 Dec;84(408):1074-1078. [doi: 10.1080/01621459.1989.10478874]
38. Burnham KP, Anderson DR. Multimodel Inference. Sociological Methods & Research 2016 Jun 30;33(2):261-304. [doi: 10.1177/0049124104268644]
39. Hermansky H, Morgan N. RASTA processing of speech. IEEE Trans Speech Audio Process 1994;2(4):578-589. [doi: 10.1109/89.326616]
40. Rabiner L, Schafer R. Theory and Applications of Digital Speech Processing. Hoboken, NJ: Prentice Hall Press; 2010.
41. Al-Hameed S, Benaissa M, Christensen H, Mirheidari B, Blackburn D, Reuber M. A new diagnostic approach for the identification of patients with neurodegenerative cognitive complaints. PLoS One 2019 May 24;14(5):e0217388 [FREE Full text] [doi: 10.1371/journal.pone.0217388] [Medline: 31125389]
42. Martínez-Nicolás I, Llorente TE, Martínez-Sánchez F, Meilán JJG. Ten years of research on automatic voice and speech analysis of people with Alzheimer's disease and mild cognitive impairment: a systematic review article. Front Psychol 2021 Mar 23;12:620251 [FREE Full text] [doi: 10.3389/fpsyg.2021.620251] [Medline: 33833713]
43. Li J, Yu J, Ye Z, Wong S, Mak M, Mak B, et al. A comparative study of acoustic and linguistic features classification for Alzheimer's disease detection. 2021 May 13 Presented at: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); June 6-11, 2021; Toronto, ON p. 6423-6427. [doi: 10.1109/icassp39728.2021.9414147]
44. Gonnerman LM, Aronoff JM, Almor A, Kempler D, Andersen ES. From Beetle to bug: progression of error types in naming in Alzheimer's disease. Proceedings of the Annual Meeting of the Cognitive Science Society 2004;26(26):1563 [FREE Full text]

45. Nagumo R, Zhang Y, Ogawa Y, Hosokawa M, Abe K, Ukeda T, et al. Automatic detection of cognitive impairments through acoustic analysis of speech. Curr Alzheimer Res 2020 Mar 20;17(1):60-68 [FREE Full text] [doi: 10.2174/1567205017666200213094513] [Medline: 32053074]
46. Meilán JJG, Martínez-Sánchez F, Carro J, López DE, Millian-Morell L, Arana JM. Speech in Alzheimer's disease: can temporal and acoustic parameters discriminate dementia? Dement Geriatr Cogn Disord 2014 Jan 30;37(5-6):327-334. [doi: 10.1159/000356726] [Medline: 24481220]
47. Mittal V, Sharma RK. Machine learning approach for classification of Parkinson disease using acoustic features. J Reliable Intell Environ 2021 Apr 26;7(3):233-239. [doi: 10.1007/s40860-021-00141-6]
48. Perez M, Jin W, Le D, Carlozzi N, Dayalu P, Roberts A, et al. Classification of Huntington disease using acoustic and lexical features. Interspeech 2018;2018:1898-1902 [FREE Full text] [doi: 10.21437/interspeech.2018-2029] [Medline: 33241056]
49. Crowe SF. The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. J Clin Psychol 1998 Aug;54(5):585-591. [doi: 10.1002/(sici)1097-4679(199808)54:5<585::aid-jclp4>3.0.co;2-k] [Medline: 9696108]
50. Arbuthnott K, Frank J. Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. J Clin Exp Neuropsychol 2000 Aug;22(4):518-528. [doi: 10.1076/1380-3395(200008)22:4;1-0;FT518] [Medline: 10923061]
51. For researchers. Framingham Heart Study. URL: https://www.framinghamheartstudy.org/fhs-for-researchers/

Abbreviations

AD: Alzheimer disease
AUC: area under the curve
FAS: Controlled Oral Word Association Test
FHS: Framingham Heart Study
HVOT: Hooper Visual Organization Test
LLD: low-level descriptor
LMi: Logical Memory—Immediate Recall
MCI: mild cognitive impairment
MFCC: Mel-frequency cepstral coefficient VRi: Visual Reproduction—Immediate Recall
NP: neuropsychological
OR: odds ratio
ROC: receiver operating characteristic
SIM: Similarities
TrailsB: Trail Making Test B
VRd: Visual Reproduction—Delayed Recall
VRr: Visual Reproduction—Recognition

Edited by G Eysenbach, T Leung; submitted 22.09.22; peer-reviewed by I van der Linde, J Chan; comments to author 25.10.22; revised version received 29.11.22; accepted 03.12.22; published 22.12.22.

Please cite as:
Ding H, Mandapati A, Karjadi C, Ang TFA, Lu S, Miao X, Glass J, Au R, Lin H
Association Between Acoustic Features and Neuropsychological Test Performance in the Framingham Heart Study: Observational Study
J Med Internet Res 2022;24(12):e42886
URL: https://www.jmir.org/2022/12/e42886
doi:10.2196/42886
PMID:36548029

Original Paper

Predicting Publication of Clinical Trials Using Structured and Unstructured Data: Model Development and Validation Study

Siyang Wang[1*], MSc; Simon Šuster[1*], PhD; Timothy Baldwin[1,2], PhD; Karin Verspoor[3], PhD

[1]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[2]Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates

[3]School of Computing Technologies, RMIT University, Melbourne, Australia

[*]these authors contributed equally

Corresponding Author:
Simon Šuster, PhD
School of Computing and Information Systems
University of Melbourne
Parkville
Melbourne, 3000
Australia
Phone: 61 40834491
Email: simon.suster@unimelb.edu.au

Abstract

Background: Publication of registered clinical trials is a critical step in the timely dissemination of trial findings. However, a significant proportion of completed clinical trials are never published, motivating the need to analyze the factors behind success or failure to publish. This could inform study design, help regulatory decision-making, and improve resource allocation. It could also enhance our understanding of bias in the publication of trials and publication trends based on the research direction or strength of the findings. Although the publication of clinical trials has been addressed in several descriptive studies at an aggregate level, there is a lack of research on the predictive analysis of a trial's publishability given an individual (planned) clinical trial description.

Objective: We aimed to conduct a study that combined structured and unstructured features relevant to publication status in a single predictive approach. Established natural language processing techniques as well as recent pretrained language models enabled us to incorporate information from the textual descriptions of clinical trials into a machine learning approach. We were particularly interested in whether and which textual features could improve the classification accuracy for publication outcomes.

Methods: In this study, we used metadata from ClinicalTrials.gov (a registry of clinical trials) and MEDLINE (a database of academic journal articles) to build a data set of clinical trials (N=76,950) that contained the description of a registered trial and its publication outcome (27,702/76,950, 36% published and 49,248/76,950, 64% unpublished). This is the largest data set of its kind, which we released as part of this work. The publication outcome in the data set was identified from MEDLINE based on clinical trial identifiers. We carried out a descriptive analysis and predicted the publication outcome using 2 approaches: a neural network with a large domain-specific language model and a random forest classifier using a weighted bag-of-words representation of text.

Results: First, our analysis of the newly created data set corroborates several findings from the existing literature regarding attributes associated with a higher publication rate. Second, a crucial observation from our predictive modeling was that the addition of textual features (eg, eligibility criteria) offers consistent improvements over using only structured data (F_1-score=0.62-0.64 vs F_1-score=0.61 without textual features). Both pretrained language models and more basic word-based representations provide high-utility text representations, with no significant empirical difference between the two.

Conclusions: Different factors affect the publication of a registered clinical trial. Our approach to predictive modeling combines heterogeneous features, both structured and unstructured. We show that methods from natural language processing can provide effective textual features to enable more accurate prediction of publication success, which has not been explored for this task previously.

(J Med Internet Res 2022;24(12):e38859) doi:10.2196/38859

KEYWORDS

clinical trials; study characteristics; machine learning; natural language processing; pretrained language models; publication success

Introduction

Background

Rigorously conducted randomized controlled trials provide the highest level of scientific evidence, enabling medical practitioners to provide better care for patients and ultimately improving public health. Available, findable, and accessible clinical research results are necessary for the successful transfer of findings into evidence-based practice and further research [1]. In recent years, improved clinical trial registration has meant that more trials than ever are now discoverable and searchable according to a variety of metadata. However, registration does not offer detailed information about important aspects of the study execution and results, such as specification of outcomes and pointers to all resulting publications [2]. Scientific publications resulting from completed clinical trials offer a means of disseminating the findings comprehensively, which is essential for supporting subsequent clinical trials, increasing possibilities for research collaboration, and advancing medical practice and research [3]. In addition to research results, detailed information about the study methods provided in publications is also critical to appraising the validity, reliability, and applicability of clinical evidence in clinical practice [4].

Despite the importance of publication, many clinical trials are never published. Estimates of the publication rate of trials vary depending on the medical area and length of the follow-up period. Broadly, publication rates are in the range of 52% to 77% [5-8]. On the basis of a shorter follow-up period of 30 months from clinical trial completion, the rates tend to be lower, at approximately 11% to 46% [3,6,9]. When results are not published, are substantially delayed, or are published selectively based on the direction or strength of the findings, the ability of health care professionals and consumers to make informed decisions based on the full body of current evidence is impeded [10,11]. Such gaps in the evidence base can lead to the use of ineffective or harmful interventions and potentially waste scarce health care resources. In a study by Eyding et al [12] on the treatment of depression, it was found that, when unpublished studies were included in a meta-analysis, the antidepressant reboxetine had more adverse effects but no better efficacy than placebo for treatment of major depression, a different finding from that when only published studies were included. Additional ethical concerns have also been raised by some researchers [7,13], highlighting that, in the case of nonpublication, the trial participants are still exposed to the risks of participation but without the societal benefits resulting from the dissemination of study results.

In this work, we explore the factors affecting publication of the outcomes of individual clinical trials through the tool of predictive modeling of clinical trial–publication outcomes based on a large data set of clinical trials and associated literature. The adoption of this approach provides a mechanism for both predicting the publication outcome of a given trial and identifying the key factors driving those outcomes.

Existing Work and Contributions

Publication Outcome Studies

Many studies have addressed the publication rates of clinical trials and the factors influencing them. However, previous studies used different statistical analysis methods to examine the association between study characteristics and the publication outcome of a clinical trial. The available studies either analyzed a small number of clinical trials (in the order of hundreds) [3,7,14] or included only clinical trials with specific populations (eg, children or patients with cancer [5,15,16]). Conversely, in our work, we focused on approaching the modeling of publication outcomes *through a predictive lens*, although we also provided a descriptive analysis to better characterize the data set that we developed. Our analysis examined factors that may affect the publication outcome without any constraints regarding the population or medical specialty and, therefore, was more general.

A number of studies have focused on analyzing and remedying the quality of linkage between ClinicalTrials.gov and PubMed [17-22]. The presence of incomplete links may hamper efforts to measure publication and outcome reporting biases and identify relevant trials for systematic reviews. As a result of this, semiautomated methods that rank articles using natural language processing (NLP) techniques and allow humans to scan the top-ranked documents are valuable in supporting the effective identification of clinical trial publications [17,18].

Factors Affecting Publication

A variety of factors have been identified as influencing publication outcome, which can be summarized as follows: (1) large clinical trials and those with noncommercial funding are more likely to be published [8,13,23]; (2) industry-funded clinical trials are less likely to appear as publications [7]; (3) the likelihood of publication is associated with the direction and significance of study findings [11,24], although whether to assign this publication bias to rejection by journals or the lack of time and interest by the investigators has been disputed [7]; (4) place of conduct of the research may affect the odds of publication [23]; (5) some fields have higher publication rates, for example, neurology and psychiatry [13] (this may in certain cases be related to the existence of subareas, eg, vascular neurology, with niche journals allowing for easier dissemination [25]); and (6) lack of time and resources by the authors, and even disagreement between coauthors, have been mentioned as potential factors in the literature [26] but are not captured directly in the description of clinical trials and, therefore, are difficult to quantify.

Completion Status and Drug Approval Studies

Although we are not aware of any work that analyzes publishability within a predictive framework, several related

problems have been treated as classification problems [27-29]. One such task is predicting the completion of a clinical trial. Noncompletion can be seen as similar to nonpublication in terms of undesired consequences. A clinical trial that is not completed typically still involves significant financial resources, so it would make sense to ensure that decision makers are aware of the likelihood of termination or nonpublication in the early stages of a clinical trial, potentially allowing for changes in the study design. Admittedly, having such predictive power would mean that the decision makers are shouldered with the additional responsibility of considering the potential for nonpublication and have the ability to interpret the output of such predictive models. Care would also need to be taken on an ongoing basis to mitigate potential biases in the model and its use [30,31].

Another task related to publication outcome prediction is whether a drug intervention studied in a clinical trial will result in the approval of the drug. Machine learning (ML) over structured data has been explored in this context [32-34], relying on features pertaining to drug and trial characteristics as well as those covering commercial figures relating to indication. Lo et al [33] proposed a large data set consisting of approval outcomes of >6000 drug-indication pairs across almost 16,000 phase-2 trials. Although this represents the largest data collection for applying supervised ML to drug approval, our task was more general (concerning clinical trials without needing to identify drug-indication pairs), allowing us to include an even larger number of clinical trials paired with publication outcomes.

In contrast to descriptive studies on publication status, studies on trial completion and drug approval do include textual inputs from trial descriptions in the modeling, which leads to better sensitivity and specificity than using structured features alone [27,35]. These studies generally use relatively simple methods to represent text. Elkin and Zhu [27] included word-embedding features [36,37] in predicting trial completion but only used static word representations rather than more advanced contextualized word representations derived from pretrained language models [38,39]. In drug approval prediction, features constructed over unstructured input data have been studied by Feijoo et al [35], who focused on predicting drug transitions across clinical trial phases. The authors used simple pattern matching to develop an eligibility criteria complexity metric defined in terms of the number of inclusion and exclusion criteria. Although these criteria were shown to be useful (a higher number of criteria has been connected with a higher risk of trial failure), their representation is still rather rudimentary. In our work, we included the eligibility criteria using state-of-the-art NLP techniques that can capture the meaning of the eligibility criteria.

Contributions

We constructed and made available a new data set that provides publication outcomes for trials registered in ClinicalTrials.gov. It is the largest data set of its kind to date.

Predicting the publication status of a clinical trial using numerical, categorical, and textual input features in a single ML model leads to a classification performance of an area under the curve (AUC) of >0.7. We found that textual descriptions of registered trials are an important source of information and are effectively represented using NLP techniques.

We identified a lack of studies investigating publishability within a *predictive* framework. Thus, we confirmed several factors known from *descriptive* studies to influence the publication outcome and identified *new* ones from textual descriptions of clinical trials (eg, eligibility criteria). Our work lays the foundation for a technology that would support trial planning and decision-making by providing, for a given trial, the prominent features that lead to a particular publication outcome. How such technology can best benefit trial developers in increasing the value of their prospective study should be a subject of future research.

Methods

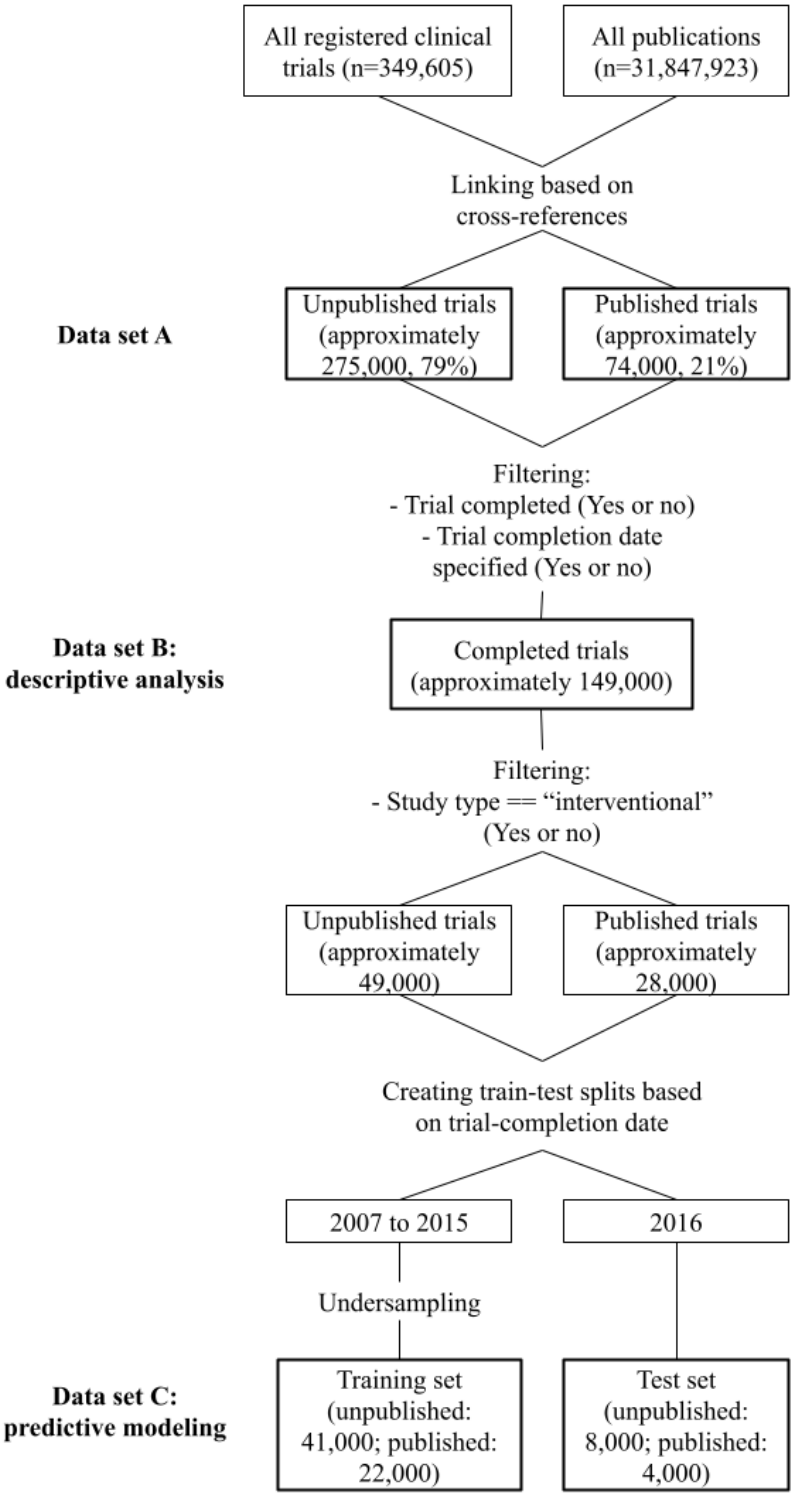
Constructing a Data Set Automatically

We used 2 primary resources in our work: the largest available registry of clinical trials, ClinicalTrials.gov, and MEDLINE, a bibliographic database of academic journal articles. For both data sources, we used the data dumps in XML available as of the start of our study in August 2020 [40,41]. To find out which clinical trials were actually published, we adopted a 2-step procedure and took the union over clinical trial-publication links found at each step. The first step recognized all PubMed article IDs directly listed in the registry of clinical trials. However, as some clinical trials lacked this information, we also looked for clinical trial–related information within the publications themselves (second step). We located that information in MEDLINE inside the databank list, from which we retrieved the clinical trial identifier provided that the databank name equaled “clinicaltrials.gov.” To consider a trial published, we required that there be at least one publication associated with it in MEDLINE. If a trial had more than one associated publication, additional pairs were created for each publication.

The final result was a map between clinical trial IDs and PubMed article ID values (*trial-publication map*). In our data set, the number of clinical trials that had an associated publication was 74,394, and there were approximately 275,000 clinical trials without publication, totaling approximately 349,000 trials (data set A). We illustrate the data creation procedure in Figure 1. We made the mapping openly available to promote further work on this topic.

The complete list of data fields and model features used in our work is shown in Table S1 in Multimedia Appendix 1 [42]. Although most of the features were obtained directly from the trial file, information such as the number of research sites and the number of primary or secondary outcomes was not explicitly stated. Therefore, we added those features as they pertain to clinical trial design and may contain an important signal for the prediction of publication status.

Figure 1. Data set construction.



The data set used in our descriptive analysis and predictive modeling (*data set B*) was based on selecting the instances that satisfied a few additional criteria. Specifically, we filtered out data instances that did not satisfy the following two conditions: (1) the study had both started and been *completed*, with known start and end dates and without "anticipated" status (as the information about a clinical trial may be updated several times after registration, such as updating the enrollment field, which indicates the planned number of participants, the information remains stable after completion, thus increasing the representativeness); and (2) the *completion date* of the study was later than 2006 (to remove older studies whose information was less complete) but earlier than 3 years before our data collection (to allow time for publication, similarly to Jones et al [7] and Ross et al [3]).

Performing these steps reduced the size of the data considerably. The resulting data set was used to obtain the descriptive statistics.

In addition, we constrained the type of study to be *interventional* to obtain the data set used in predictive modeling (*data set C*). We decided to exclude observational studies as they are less common and are characterized by several features that are different from those of interventional studies.

To emulate the real-world scenario of predicting publishability of future trials, we partitioned the data such that the completion dates of all trials in the test set postdated those in the training data set. This also made the task more challenging as we could expect previously unseen interventions in the test set. Finally, we removed all features from each trial record that would not have been known at the time of registration of the trial, such as the trial duration and results. Although including them would simplify the prediction, it would also make the task less realistic. By comparison, we note that, in the related ML task of the drug approval prediction work by Lo et al [33], the authors assumed that the same information about clinical trials *is* accessible. As these features are found to be strong predictors of drug approval, the predictive performance is likely to suffer in the more realistic scenario of this information not being available.

As the number of unpublished clinical trials in data set C was much larger than that of published clinical trials, we randomly undersampled the unpublished trials for our publication prediction experiments. We performed the undersampling by stratifying per completion year, keeping roughly equal percentages of positive and negative labels in each year. Note that we performed this step for the training set only, preserving the real-world label bias in the test set, again to make the task as faithful to reality as possible.

Manually Constructed Test Set

The aforementioned data construction approach provided a large-scale data set that allowed us to analyze and predict the publication status at scale using ML models. However, some links between clinical trials and publications may be incomplete, as we mentioned in the *Existing Work and Contributions* section. Therefore, we gathered data from 3 previously published studies [3,18,20] that included manual publication status annotations (see Table 1 for the statistics). Although the scale of these annotations was smaller than in our automatically constructed data set, because of human effort, it was less likely that the publication of a clinical trial would go unnoticed. We used this data set as an additional test set and also made it publicly available with the permission of the original authors [43].

Table 1. Data from previously published studies. A total of 5 studies were included in more than one original work but received the same annotation. Owing to this, the size of the resulting test set was less than the sum of the sizes of the individual data sets.

	Size	Proportion of positive labels (“published”) out of all
Ross et al [3]	630	0.54
Zarin et al [20]	148	0.23
Dunn et al [18]	199	0.45
Combined	972	0.48

Modeling Approach

To study factors associated with publication status and learn to predict whether a clinical trial is likely going to be published, we created 3 types of features for our models: numerical, categorical (both can be seen as structured inputs), and textual features. The textual features encode a wealth of information that augments the structured information and have the potential to improve predictive modeling, but they are also potentially much noisier. An example of textual fields that can be indicative of publication status are the inclusion and exclusion criteria. A possible link between eligibility criteria, sample size, significant effect, and publication status has been pointed out by Elkin and Zhu [27]. NLP techniques allowed us to extract and represent this information in a predictive model as well as highlight which textual features are important.

As a simple baseline, we used a k-nearest neighbor classifier that only used numerical and categorical features (with no text-based features). At test time, the classifier predicts the predominant label among k training instances that are closest to the test instance in terms of Euclidean distance. Through a random search over various values of k, we settled on k=460.

We trained and evaluated 2 different models that incorporated textual features: a random forest (RF) classifier and a neural network (NN).

For RF, a standard approach to include textual inputs is to convert them into numeric word vectors, extracting both unigrams and bigrams. These terms are weighted using term frequency-inverse document frequency (Schütze et al [44]), whereby the frequency of a term in a document is divided by the proportion of documents that that term appears in within the data set to down-weight common terms. We thresholded the vocabulary by selecting the 20,000 most frequent terms. We used the one-hot encoding method to represent categorical features and included numeric features without additional adaptation. We report other RF details in Multimedia Appendix 2.

In the NN, the categorical features are embedded using a weight matrix that is randomly initialized and updated during training. The textual inputs (examples are included in Table 2) are embedded using pretrained language models that output context-dependent token activations [39], as explained in more detail next.

Table 2. Examples of selected textual features from clinical trial metadata.

Feature name and identifier	Textual excerpt
Brief title	
NCT01309919	Bleeding Patterns and Complications After Postpartum IUD Placement: a Pilot Study
NCT00230971	Study Comparing Tigecycline Versus Ceftriaxone Sodium Plus Metronidazole in Complicated Intra-abdominal Infection (cIAI)
NCT01364948	Effect of Coconut Oil Application in Reducing Water Loss From Skin of Premature Babies in First Week of Life (TEWL) (TopOilTewl)
Brief summary	
NCT01309919	The purpose of the study is to determine the feasibility of placing the levonorgestrel-releasing intrauterine system (LNG - IUS, Mirena®) post-delivery. The investigators will gain information about complications at the time of placement; the investigators will also examine the expulsion rate, side effects, bleeding patterns and subject satisfaction at various time periods after insertion.
NCT00230971	This is a study of the safety and efficacy of tigecycline to ceftriaxone sodium plus metronidazole in hospitalised subjects with cIAI. Subjects will be followed for efficacy through the test-of-cure assessment. Safety evaluations will occur through the treatment and post-treatment periods and continue through resolution or stability of the adverse event(s).
NCT01364948	The skin of newborn infants is immature and ineffective as a barrier. Preterm skin exhibits even more vulnerability to the environment due to poor self regulatory heat mechanisms, paucity of fatty tissue and its thinness. Most preterm babies lose up to 13\% of their weight as water loss from their skin during the first week of life. Many strategies have been utilised by neonatologists to decrease this water loss. Oil application on the skin can act as a non permeable barrier and can help in reducing water loss from the skin. Edible coconut oil, often used for traditional massage of babies by Indian communities, is culturally acceptable and Hence the investigators decided to undertake this study to objectively assess the reduction in water loss from skin after oil application
Inclusion criteria	
NCT01309919	Age 18 years or older, speak either English or Spanish, desire to use an IUD as their postpartum contraception (IUD arm), do NOT desire an IUD as their contraception (Diary Only arm), plan to deliver at Baystate Medical Center
NCT00230971	Clinical diagnosis of complicated intra-abdominal infection that requires surgery within 24 hours. Fever plus other symptoms such as nausea, vomiting, abdominal pain.\\
NCT01364948	All preterm babies born at the study center with birth weight 1500gms were eligible for inclusion in the study.
Participant condition	
NCT01309919	Postpartum period
NCT00230971	Appendicitis, cholecystitis, diverticulitis, intra-abdominal abscess, intra-abdominal infection, and peritonitis
NCT01364948	Trans Epidermal Water Loss (TEWL)
Keywords	
NCT01309919	Intrauterine device, Mirena, levonorgestrel intrauterine system, postpartum contraception
NCT00230971	Intra-abdominal infections, abscess
NCT01364948	Preterm, VLBW, coconut oil application, transepidermal water loss, weight gain

We evaluated the RF and NN classifiers that used textual features compared with those without, in which only structured features were used.

We opted for 2 different encoders: Bidirectional Encoder Representations from Transformers (BERT) [39], pretrained on general-domain English corpora, and BERT for scientific texts (SciBERT) [38], pretrained on the biomedical domain. We used the same idea as Adhikari et al [45], who took the hidden layer output at the sentence-level classification level as the representation of the document. In addition, we used the hidden outputs of the 3 last layers [46] as inputs to the top dense layers of our classifier. To refine the model's representational capacity, we included 2 additional sources of information: positional and segmental. For the first one, a trainable positional embedding [47], which is unique to each token, is added to the token vector to endow the model with a sense of word order. For the second one, a trainable segment embedding helps the encoder discriminate between the multiple, independent textual fields (Table S1 in Multimedia Appendix 1) that are passed to the model as one long string of text. We found the interchangeable segment scheme illustrated in Figure S1 in Multimedia Appendix 1 to work best. Another variation represents each text field with a different segment embedding but works less well, although the difference is small. In addition, an alternative scheme for positional embeddings in which the embedding index is restarted with each text field yields similar results. We took inspiration for that from Herzig et al [48], who used positional embeddings in the context of table parsing to enhance input structuring.

A limitation of the original BERT architecture is that it can only accept sequences of up to 512 tokens. Therefore, we needed to truncate the textual inputs exceeding this limit. We started by selecting the first n=512/*T* tokens of each field (*T* being the total number of textual fields to encode). As some textual fields can be shorter, we progressively raised *n* across all fields until we reached the maximum number of tokens. Finally, the parameters of the encoder were fine-tuned jointly with the remaining NN parameters on our publication outcome prediction data set, minimizing the cross-entropy loss during training.

In addition to adopting the standard BERT model in the NN, we looked at 2 adaptations of the training regime: a special case when the encoder parameters are left unchanged during training (named "frozen" in the table of results) and a model that receives cased text as input ("cased"; ie, text that has not been previously lowercased), the latter being the most common practice. Finally, for RF, we tested an adaptation that, instead of the term frequency-inverse document frequency encoder, uses language model representations previously induced in the text. These representations were kept fixed throughout the training and testing phases.

Evaluation Details

We evaluated the predictive performance using the F_1-score measure (F_1 = 2 × [P × R / (P + R)]), which is the harmonic mean of precision (P = TP / [TP + FP]; the proportion of trials predicted as published out of all predictions, where TP are true positives and FP are false positives) and recall (R = TP / [TP + FN]; the proportion of trials predicted as published out of all published trials, where FN are false negatives). We also reported the area under the receiver operating characteristic curve (itself indicative of the trade-off between recall and false-positive rate at various thresholds over the predicted probabilities), which was useful in summarizing the classifier's ability to distinguish between classes via a single figure of merit.

Results

Descriptive Analysis

Overview

To obtain a clear idea of the *publication rate* in our data set, we plotted the number of published and unpublished studies per year, as shown in Multimedia Appendix 3. We observed that the number of registered trials was monotonically increasing (with >20,000 trials registered in 2016), but the number of published trials increased less strongly. For trials with an earlier completion year, the publication rate was approximately 45%, whereas, for later trials, it decreased by approximately 10%. For comparison, existing studies on publication rates reported highly variable publication percentages, up to 77% in Huiskens et al [6] and as low as 11% in Chen et al [9] depending on the medical area and length of follow-up considered.

Furthermore, we examined the *time needed to publish*. Analyzing only the published studies, we found a median time to publish of 27 months. We show the distribution of publication times in Figure 2. For a smaller number of trials, it can take much longer to publish, as seen by the long tail on the right of the plot. The previous studies generally reported shorter times of approximately 19 to 23 months [3,9,16].

An additional way of analyzing publication time is to plot the probability that a study will go unpublished for an interval longer than some time *t*. We borrowed here a tool from survival analysis, the Kaplan-Meier plot. By analogy, the survival time in our case represents the time that a clinical trial remains unpublished, and the relevant event is the publication. Some individuals (clinical trials) may be lost to follow-up (right censoring), which is also considered by the method. We see in Figure 3 that, when given a very short period (eg, a few months after completion), the chance is still high that the trial will not be published. When given more time, the probability of nonpublication drops, although it remains fairly high even for very long intervals (at 80 months, it is still >70%).

Figure 2. The distribution of publication times in months.

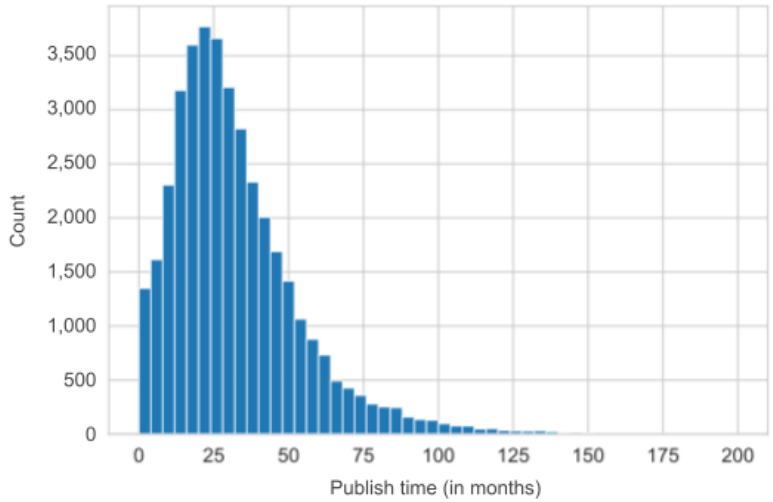
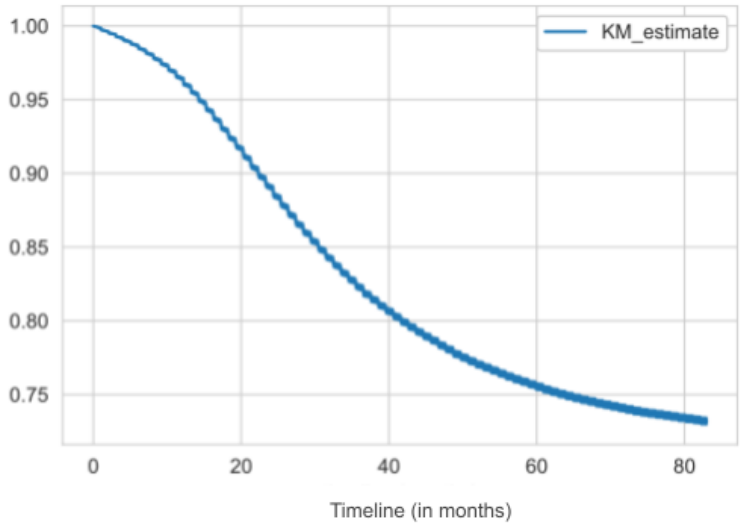


Figure 3. A Kaplan-Meier (KM) plot representing the probability (y-axis) that a trial will go unpublished for longer than the number of months shown on the x-axis.



Association Between Publication Outcome and Categorical Features

To analyze the relationship between a feature and the publication outcome, we applied the chi-square test (in line with the related literature [8,9,14,16,23,49,50]) but, because of its sensitivity to the sample size [51,52], we also carried out the Cramér *V* association test for discrete variables. In this analysis, we followed the related work and focused on categorical features only. In the *Predictive Performance* section, we analyze the importance of all feature types in predictive performance. The results for all categorical features are shown in Table 3. The features with the highest values of *V* include the overall status (eg, a value such as "Suspended" may be indicative of future publication), whether the results were reported, enrollment type (anticipated vs actual), and the phase of the trial (when calculating the odds ratio over different phases of the trial, we found that trials in phase 3 were 2 times more likely to be published than trials in other phases). By contrast, some features such as the type of observational study (retrospective, prospective, or cross-sectional) and the class of funding agency (US National Institutes of Health, other US Federal agencies, industry, or other) can hardly be associated with publication status. The latter example is particularly surprising as most previous works have reported that the source of funding is a strong indicator of publication status [8,23,50], with the exception of Gandhi et al [14].

Table 3. Strength of association between categorical features extracted directly from structured metadata associated with clinical trials and publication status. For the definition of each feature, see Table S1 in Multimedia Appendix 1.

Feature name	Chi-square *P* value	Cramér V
overall_status	.001	0.26
were_results_reported	.001	0.157
enrollment_type	.001	0.153
Phase	.001	0.126
plan_to_share_ipd	.001	0.095
intervention_type_behavioral	.001	0.06
has_dmc	.001	0.056
intervention_model	.001	0.053
intervention_type_diagnostic_test	.001	0.047
has_single_facility	.001	0.044
intervention_type_device	.001	0.039
Country	.001	0.035
study_type	.001	0.034
Allocation	.001	0.026
primary_purpose	.001	0.025
is_fda_regulated_device	.001	0.023
Masking	.001	0.022
intervention_type_dietary_supplement	.001	0.021
intervention_type_biological	.001	0.019
Gender	.001	0.018
intervention_type_combination_product	.001	0.017
intervention_type_other	.001	0.016
intervention_type_radiation	.001	0.013
sampling_method	.001	0.013
intervention_type_drug	.001	0.012
intervention_type_procedure	.001	0.012
observational_model	.002	0.012
is_us_export	.13	0.011
responsible_party_type	.001	0.011
intervention_type_genetic	.001	0.01
healthy_volunteers	.001	0.009
is_fda_regulated_drug	.001	0.009
observational_prospective	.14	0.006
agency_class	.32	0.002

Predictive Performance

Overview

The main results of our predictive models for data set C are shown in Table 4. Interestingly, the k-nearest neighbor baseline already set a high bar for the use of structured inputs. We see that the best performance on the test set was achieved with the models that used textual information. The 2 evaluation metrics show slightly different trends (ie, when looking at F_1-score, the neural models using BERT-based representations performed better than the RF classifier using the bag-of-words representation); however, according to AUC, the RF classifier outperformed different variants of the neural model. Judging by the improvement obtained when including the textual features in both models, the NN model makes more effective use of these features. We found that the difference between the NN model using only structured features and the NN model using

SciBERT-encoded text features was statistically significant at $P<.001$ (statistic value: 778.4), measured with the McNemar test for binary classification tasks [53]. Although it had a considerably lower performance compared with the RF classifier when including only the structured features, the performance difference between the 2 models vanished when including the textual features. For the neural model, choosing a BERT model with a better domain fit (ie, SciBERT) appears to boost F_1-score, but the differences are too small to make a judgment in the case of AUC. We include the precision-recall curves in Figures 4 and 5, calculated using the predictions of the model that tested best in terms of F_1-score (ie, NN with structured and SciBERT textual features).

Table 4. Results for publication prediction[a].

Method	Input	Validation		Test	
		F_1-score	AUC[b]	F_1-score	AUC
K-nearest neighbor	Structured	0.592	N/A[c]	0.611	N/A
RF[d]	Structured	0.64	0.701	0.614	0.704
RF	Structured+text (TF-IDF[e])	0.656	0.721	0.623	0.719
RF	Structured+text (SciBERT[f])	0.65	0.709	0.63	0.711
NN[g]	Structured	0.611	0.672	0.607	0.612
NN	Structured+text (frozen SciBERT)	0.642	0.689	0.63	0.696
NN	Structured+text (SciBERT)	0.648	0.708	0.641	0.7
NN	Structured+text (cased SciBERT)	0.641	0.697	0.637	0.701
NN	Structured+text (BERT[h])	0.64	0.699	0.633	0.7

[a]All models use categorical and numerical features ("structured"). When textual features are added, this is marked with "+ text." As the k-nearest neighbor classifier does not output probabilities, we cannot calculate the area under the curve.

[b]AUC: area under the curve.

[c]N/A: not applicable.

[d]RF: random forest.

[e]TF-IDF: term frequency-inverse document frequency.

[f]SciBERT: Bidirectional Encoder Representations from Transformers model for scientific texts.

[g]NN: neural network.

[h]BERT: Bidirectional Encoder Representations from Transformers.

Figure 4. Precision-recall curve for the positive class (publication) using the neural network model with structured and textual features from a Bidirectional Encoder Representations from Transformers model for scientific texts. AP: average precision.

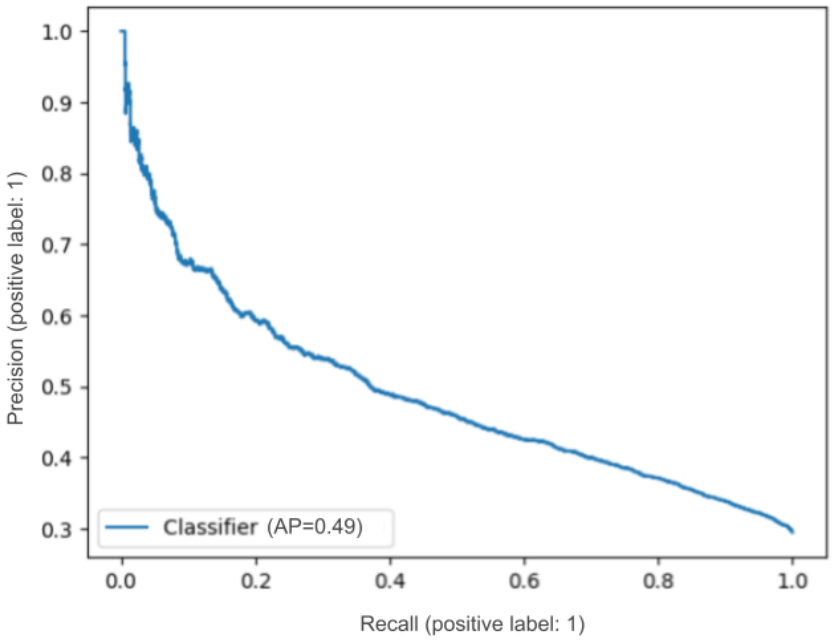
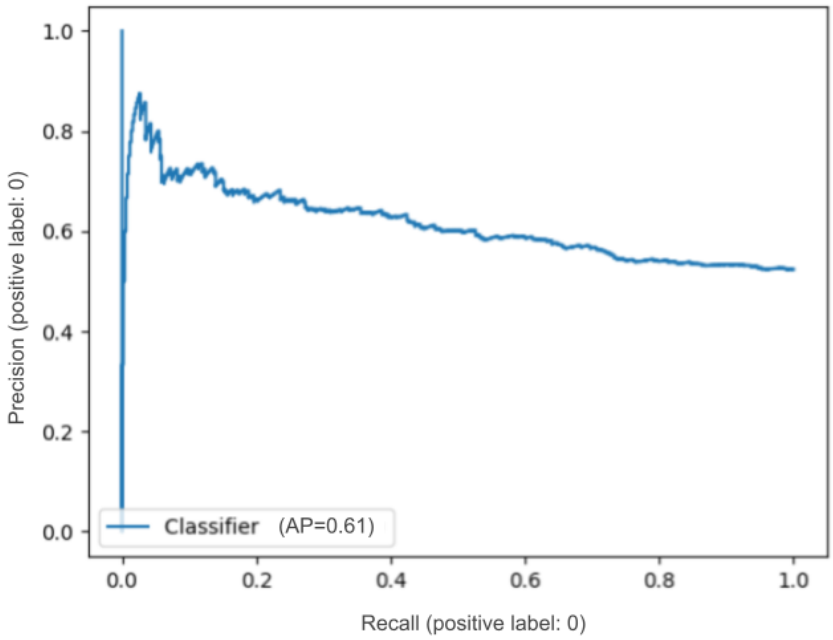


Figure 5. Precision-recall curve for the negative class (nonpublication) using the neural network model with structured and textual features from a Bidirectional Encoder Representations from Transformers model for scientific texts. AP: average precision.



Factors Affecting Publication

To determine which features play a key role in prediction, we used a feature permutation technique to obtain the features ranked by their respective drop in performance. We performed this analysis using RF only because of faster inference times. The classifier is trained once; then, at test time, a corrupted representation of a feature is obtained by shuffling its possible feature values in the test set. After that, the model is applied to the test set, and the drop in accuracy is calculated compared with the performance on the noncorrupted data set. We only corrupted one feature at a time and repeated the process for all features. The entire process was performed 5 times using different random seeds for shuffling, after which the reported scores were averaged.

The results, organized according to feature type, are shown in Table 5. The most significant numerical feature is the number of enrolled participants, with a possible explanation being that it may affect the reliability of the results (thus ultimately increasing the odds of publication). Similarly, a larger number of facilities has been linked to higher publication rates [8]. The number of outcomes indicates the size and complexity of the study, which may in turn also affect publishability. For textual inputs, the narrative describing the trial (the detailed description and brief summary) as well as the eligibility criteria are the strongest features. We observed that some textual features contained overlapping information. For example, the brief title could be subsumed into the official title. The same word often occurred in different inputs, and this redundancy can be a strong indicator for predicting publication status. For example, when we measured the importance of the words in RF using the impurity criterion of our RF implementation [9], we found that the presence of *randomized* (occurring in both the official title and detailed description) was a strong discriminator between published and unpublished studies.

In the case of categorical inputs, we found similar features to be important, as mentioned in the *Descriptive Analysis* section, including the country of the main institution (“country”) and whether the study had a data monitoring committee (“has dmc”). However, some features that were found to be important in our descriptive analysis and in the prior work were less important in the predictive approach (eg, the phase of investigation [“phase”], the allocation of participants to trial arms [“allocation”], and the method used to assign an intervention to participants [“intervention model”]).

Table 5. The drop in accuracy after permuting the values of a feature as measured with random forest using term frequency-inverse document frequency representation of text. The values for each feature type are ranked in decreasing order, so the most important features are mentioned first.

Feature type and feature	Drop in accuracy
Numerical	
number_of_facilities	0.007364
outcome_counts_secondary	0.004911
outcome_counts_others	0.004068
outcome_counts_primary	0.003702
number_study_directors	0.003518
number_study_chairs	0.003359
minimum_age	0.003235
number_principal_investigators	0.003157
maximum_age	0.002719
number_of_arms	0.000985
Textual	
detailed_description	0.010193
brief_summary	0.008551
criteria_Exclusion	0.008313
criteria_Inclusion	0.004971
official_title	0.003428
brief_title	0.001433
Source	0.001342
responsible_party_keywords	0.001064
participant_condition	0.00064
Categorical	
has_single_facility	0.004591
intervention_type_Behavioral	0.004211
primary_purpose	0.003914
Country	0.003804
intervention_type_Biological	0.003643
is_fda_regulated_device	0.003376
is_us_export	0.003333
intervention_type_Diagnostic_Test	0.003322
intervention_type_Combination_Product	0.003322
intervention_type_Genetic	0.003322
is_fda_regulated_drug	0.003321
intervention_type_Procedure	0.003205
has_dmc	0.003185
intervention_type_Other	0.003144
intervention_type_Radiation	0.003144
intervention_type_Device	0.003078
Gender	0.003012
responsible_party_type	0.002925
intervention_type_Dietary_Supplement	0.002873

Feature type and feature	Drop in accuracy
plan_to_share_ipd	0.002819
healthy_volunteers	0.002607
intervention_type_Drug	0.00227
agency_class	0.001854
Phase	0.001426
Allocation	0.001347
intervention_model	0.00131

Performance on the Manually Verified Test Set

As an additional experiment, we took the model that achieved the highest F_1-score on the automatically constructed data set (NN with structured+text [SciBERT] input features) and applied it to the test set built from the manually verified publication links introduced in the *Manually Constructed Test Set* section. We measured an F_1-score of 55.9 and area under the receiver operating characteristic curve of 58.6. To better understand this drop in performance with respect to automatically obtained test sets, we calculated a confusion matrix, which revealed that the model too eagerly predicted "publication" (ie, it was more likely to commit a type-1 error [a false positive, 272/972, 28% of the time] than a type-2 error [a false negative, 146/972, 15% of the time]). As the test data consisted of 3 subsets, there might be important individual variations in the performance that we need to consider. Indeed, splitting the results according to each subset (Table 6), we noticed that the subset from Zarin et al [20] showed lower performance than the subsets from Ross et al [3] and Dunn et al [18], both with similar performance. Our explanation is that these subsets contain varying proportions of positive labels, which, if different from those seen during training, will negatively affect the test performance. Specifically, the Zarin et al [20] subset has only 23% (34/148) of positive labels compared with approximately 50% (410/824, 49.8%) in the remaining subsets. Understandably, the model that was trained on roughly equal portions of positive and negative instances overpredicted the positive class on the Zarin et al [20] subset, and almost all modeling mistakes in this case were due to false positives (78/87, 90% compared with 9/87, 10% of false negatives). We found that this negative effect vanished when the model was retrained with a similar ratio of positive to negative instances. We used the nonbalanced version of our training data set (data set C in Figure 1).

Table 6. Data statistics and performance on the subsets of the manually verified test set.

	Ross et al [3]	Zarin et al [20]	Zarin et al [20] with nonbalanced training set	Dunn et al [18]
Percentage positive[a]	54	23	23	45
F_1-score	58.4	43.4	58.2	55.0
AUROC[b]	62.3	52.6	53.5	60.4

[a]*Percentage positive* represents the percentage of instances bearing the positive label (*published*) out of all instances.

[b]AUROC: area under the receiver operating characteristic curve.

Discussion

Limitations

Although our work established at scale the various attributes associated with a higher publication rate and the positive impact of including textual descriptions of clinical trials in a predictive framework, a few additional considerations are necessary.

The qualitative performance of an ML model is sensitive to the quality of the underlying data that are used for training and testing, and predicting publication success is no different. When constructing our data set, we noticed that incorrect information existed in the trial registration entries (eg, the estimated completion year may be set to 2099). In addition, the current status of the study (eg, ongoing, completed, or terminated) may not be always up to date, and this is similar for other registered information. Incompleteness and incorrect information in ClinicalTrials.gov have been examined in the literature [7,54-56], but the precise extent of this is unknown and difficult to estimate, and it would require substantial manual effort to reveal it. We see noise as an integral part of learning from large data collections, similar to the related work (*Existing Work and Contributions* section) that uses structured resources such as ClinicalTrials.gov [27-29,32-34] and to the work on learning under distant supervision [57-59]. As our classifiers used a very large number of training instances and each instance is represented using multiple features, the effect of occasional noise is deemed small.

Another potential source of noise in our automatically constructed data set could stem from the linkage between clinical trials and their publications, which is established automatically and, hence, prone to incorrect or missed links. The data set was also limited to studies that were publicly available and indexed in public resources. Although conference abstracts and other gray literature resources may provide additional context on trial outcomes, they are not typically considered to be formal publications and require ad hoc strategies for collection that are beyond the scope of our study. Overall, the results presented

reflect the most realistic scenario possible based on accessible resources.

Finally, a more general limitation in the modeling of publication outcomes is that it is difficult to capture and quantify the influence of factors that are not available in trial registries but would otherwise be useful, particularly for understanding nonpublication, for example, whether investigators did not have enough time to publish and instead focused on other tasks, whether there were changing interests or disagreements between coauthors, whether researchers believed that a journal was unlikely to accept their work, and whether financial problems or other contractual issues prevented publication [15,60-62]. Although such information is obtainable from study authors in principle, it would be extremely difficult to carry out such information acquisition at scale, and it is not currently available in public resources.

Impact

In this study, we sought to simulate a real-world situation in which a prospective estimate is desired regarding the publication outcome of a clinical trial. To this end, we carried out a set of experiments on the newly created data set that linked clinical trial records from the period of 2007 to 2016 with their publications, if they existed, with a follow-up period of 4 years. The resulting data set represents the largest such collection available to date. We have shown how a combination of heterogeneous features—including text features derived from the clinical trial registry record—can lead to a classification performance of >0.7 AUC; this means that, if one randomly selects a case that is positive (ie, a trial that will eventually lead to publication), there is at least a 70% chance that the case is also classified as such. This technology has strong potential to be used in trial design. It can provide a prospective estimate of publishability in the early stages of a clinical trial when the properties of the study design and environment are already known, more broadly giving an indication of the viability of the trial. The tool could reveal to trial developers the different areas suggestive of lowered publication chances (and, by extension, of a reduced value of their study) before wasting resources unnecessarily. In future work, we will explore the incorporation of this model into a system that can effortlessly and in a human-friendly way provide, for a given trial, the prominent features that lead to a particular outcome, as well as indicate the reliability of the classifier's decision, to support trial planning and decision-making.

Acknowledgments

This research was funded by the Australian Research Council through an Industrial Transformation Training Centre Grant (IC170100030) in collaboration with IBM. The authors sincerely thank Deborah Zarin, Joseph Ross, and Shifeng Liu with Adam G Dunn for making their annotated data available.

Data Availability

We have made the data set publicly available [63].

Authors' Contributions

SW collected the data, conceived and designed the analysis, and performed the analysis. SŠ conceived and designed the analysis, contributed to conceptualization, and wrote the paper. TB conceived and designed the analysis, contributed to conceptualization, wrote the paper, and supervised. KV contributed to conceptualization and wrote the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Text-representation scheme.
[PNG File , 98 KB - jmir_v24i12e38859_app1.png]

Multimedia Appendix 2
Experimental details.
[DOCX File , 13 KB - jmir_v24i12e38859_app2.docx]

Multimedia Appendix 3
The distribution of published and unpublished trials per year of completion.
[PNG File , 33 KB - jmir_v24i12e38859_app3.png]

References

1. Wilkinson MD, Dumontier M, Jan Aalbersberg I, Appleton G, Axton M, Baak A, Jan Velterop, et al. Addendum: the FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2019 Mar 19;6(1):6 [FREE Full text] [doi: 10.1038/s41597-019-0009-6] [Medline: 30890711]
2. Dickersin K, Rennie D. The evolution of trial registries and their use to assess the clinical trial enterprise. JAMA 2012 May 02;307(17):1861-1864. [doi: 10.1001/jama.2012.4230] [Medline: 22550202]
3. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. BMJ 2012 Jan 03;344(jan03 1):d7292 [FREE Full text] [doi: 10.1136/bmj.d7292] [Medline: 22214755]
4. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Hoboken, New Jersey, United States: Wiley; 2019.
5. Chapman PB, Liu NJ, Zhou Q, Iasonos A, Hanley S, Bosl GJ, et al. Time to publication of oncology trials and why some trials are never published. PLoS One 2017 Sep 21;12(9):e0184025 [FREE Full text] [doi: 10.1371/journal.pone.0184025] [Medline: 28934243]
6. Huiskens J, Kool BR, Bakker J, Bruns ER, de Jonge SW, Olthof PB, et al. From registration to publication: a study on Dutch academic randomized controlled trials. Res Synth Methods 2020 Mar 28;11(2):218-226 [FREE Full text] [doi: 10.1002/jrsm.1379] [Medline: 31614063]
7. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. BMJ 2013 Oct 29;347(oct28 9):f6104 [FREE Full text] [doi: 10.1136/bmj.f6104] [Medline: 24169943]
8. Blmle A, Huwiler K, Witschi M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. Swiss Med Wkly 2008 Apr 05:197. [doi: 10.4414/smw.2008.12027]
9. Chen R, Desai NR, Ross JS, Zhang W, Chau KH, Wayda B, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. BMJ 2016 Feb 17;352:i637 [FREE Full text] [doi: 10.1136/bmj.i637] [Medline: 26888209]
10. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014 Jan;383(9912):166-175. [doi: 10.1016/s0140-6736(13)62227-8]
11. Schmucker C, Schell LK, Portalupi S, Oeller P, Cabrera L, Bassler D, OPEN consortium. Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. PLoS One 2014 Dec 23;9(12):e114023 [FREE Full text] [doi: 10.1371/journal.pone.0114023] [Medline: 25536072]
12. Eyding D, Lelgemann M, Grouven U, Härter M, Kromp M, Kaiser T, et al. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. BMJ 2010 Oct 12;341(oct12 1):c4737 [FREE Full text] [doi: 10.1136/bmj.c4737] [Medline: 20940209]
13. Blümle A, Wollmann K, Bischoff K, Kapp P, Lohner S, Nury E, et al. Investigator initiated trials versus industry sponsored trials - translation of randomized controlled trials into clinical practice (IMPACT). BMC Med Res Methodol 2021 Aug 31;21(1):182 [FREE Full text] [doi: 10.1186/s12874-021-01359-x] [Medline: 34465296]
14. Gandhi R, Jan M, Smith HN, Mahomed NN, Bhandari M. Comparison of published orthopaedic trauma trials following registration in Clinicaltrials.gov. BMC Musculoskelet Disord 2011 Dec 07;12(1):278 [FREE Full text] [doi: 10.1186/1471-2474-12-278] [Medline: 22151841]
15. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of investigational drugs in late-stage clinical development and publication of trial results. JAMA Intern Med 2016 Dec 01;176(12):1826-1833. [doi: 10.1001/jamainternmed.2016.6008] [Medline: 27723879]
16. Shamliyan T, Kane RL. Clinical research involving children: registration, completeness, and publication. Pediatrics 2012 May;129(5):e1291-e1300. [doi: 10.1542/peds.2010-2847] [Medline: 22529271]
17. Liu S, Bourgeois FT, Dunn AG. Identifying unreported links between ClinicalTrials.gov trial registrations and their published results. Res Synth Methods 2022 May 23;13(3):342-352. [doi: 10.1002/jrsm.1545] [Medline: 34970844]
18. Dunn AG, Coiera E, Bourgeois FT. Unreported links between trial registrations and published articles were identified using document similarity measures in a cross-sectional analysis of ClinicalTrials.gov. J Clin Epidemiol 2018 Mar;95:94-101. [doi: 10.1016/j.jclinepi.2017.12.007] [Medline: 29277557]
19. Bashir R, Bourgeois FT, Dunn AG. A systematic review of the processes used to link clinical trial registrations to their published results. Syst Rev 2017 Jul 03;6(1):123 [FREE Full text] [doi: 10.1186/s13643-017-0518-3] [Medline: 28669351]
20. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.Gov results database — update and key issues. N Engl J Med 2011 Mar 03;364(9):852-860. [doi: 10.1056/nejmsa1012065]
21. Zarin DA, Fain KM, Dobbins HD, Tse T, Williams RJ. 10-year update on study results submitted to ClinicalTrials.Gov. N Engl J Med 2019 Nov 14;381(20):1966-1974. [doi: 10.1056/nejmsr1907644]
22. Huser V, Cimino JJ. Linking ClinicalTrials.gov and PubMed to track results of interventional human clinical trials. PLoS One 2013 Jul 9;8(7):e68409 [FREE Full text] [doi: 10.1371/journal.pone.0068409] [Medline: 23874614]

23. Al-Durra M, Nolan RP, Seto E, Cafazzo JA, Eysenbach G. Nonpublication rates and characteristics of registered randomized clinical trials in digital health: cross-sectional analysis. J Med Internet Res 2018 Dec 18;20(12):e11924 [FREE Full text] [doi: 10.2196/11924] [Medline: 30485832]
24. Canestaro WJ, Hendrix N, Bansal A, Sullivan SD, Devine EB, Carlson JJ. Favorable and publicly funded studies are more likely to be published: a systematic review and meta-analysis. J Clin Epidemiol 2017 Dec;92:58-68. [doi: 10.1016/j.jclinepi.2017.08.004] [Medline: 28842289]
25. Sreekrishnan A, Mampre D, Ormseth C, Miyares L, Leasure A, Ross JS, et al. Publication and dissemination of results in clinical trials of neurology. JAMA Neurol 2018 Jul 01;75(7):890-891 [FREE Full text] [doi: 10.1001/jamaneurol.2018.0674] [Medline: 29710083]
26. Scherer RW, Ugarte-Gil C, Schmucker C, Meerpohl JJ. Authors report lack of time as main reason for unpublished research presented at biomedical conferences: a systematic review. J Clin Epidemiol 2015 Jul;68(7):803-810 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.01.027] [Medline: 25797837]
27. Elkin ME, Zhu X. Predictive modeling of clinical trial terminations using feature engineering and embedding learning. Sci Rep 2021 Feb 10;11(1):3446 [FREE Full text] [doi: 10.1038/s41598-021-82840-x] [Medline: 33568706]
28. Follett L, Geletta S, Laugerman M. Quantifying risk associated with clinical trial termination: a text mining approach. Inform Process Manage 2019 May;56(3):516-525. [doi: 10.1016/j.ipm.2018.11.009]
29. Geletta S, Follett L, Laugerman M. Latent Dirichlet allocation in predicting clinical trial terminations. BMC Med Inform Decis Mak 2019 Nov 27;19(1):242 [FREE Full text] [doi: 10.1186/s12911-019-0973-y] [Medline: 31775737]
30. Yitong L, Timothy B, Trevor C. Towards robust and privacy-preserving text representations. arXiv 2018.
31. Jieyu Z, Tianlu W, Mark Y, Vicente O, Kai-Wei C. Men also like shopping: reducing gender bias amplification using corpus-level constraints. arXiv 2017.
32. DiMasi J, Hermann J, Twyman K, Kondru R, Stergiopoulos S, Getz K, et al. A tool for predicting regulatory approval after phase II testing of new oncology compounds. Clin Pharmacol Ther 2015 Nov 24;98(5):506-513. [doi: 10.1002/cpt.194] [Medline: 26239772]
33. Lo AW, Siah KW, Wong CH, Lo AW, Siah KW, Wong CH. Machine learning with statistical imputation for predicting drug approval. Harvard Data Sci Rev 2019 Jun 23. [doi: 10.1162/99608f92.5c5f0525]
34. Siah KW, Kelley N, Ballerstedt S, Holzhauer B, Lyu T, Mettler D, et al. Predicting drug approvals: the Novartis data science and artificial intelligence challenge. SSRN J 2021 Mar 2. [doi: 10.2139/ssrn.3796530]
35. Feijoo F, Palopoli M, Bernstein J, Siddiqui S, Albright TE. Key indicators of phase transition for clinical trials through machine learning. Drug Discov Today 2020 Feb;25(2):414-421. [doi: 10.1016/j.drudis.2019.12.014] [Medline: 31926317]
36. Le Q, Mikolov T. Distributed representations of sentences and documents. 2014 Presented at: International Conference on Machine Learning; 2014; Beijing, China p. II-1188-II-1196.
37. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: International Conference on Neural Information Processing Systems; 2013; Lake Tahoe.
38. Beltagy I, Lo K, Cohan A. SciBERT: a pretrained language model for scientific text. 2019 Presented at: Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); Nov, 2019; Hong Kong, China. [doi: 10.18653/v1/d19-1371]
39. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2019; Minneapolis.
40. All public XML. ClinicalTrials.gov. URL: https://clinicaltrials.gov/AllPublicXML.zip [accessed 2022-11-24]
41. Index of /pubmed/baseline. NCBI. URL: https://ftp.ncbi.nlm.nih.gov/pubmed/baseline [accessed 2020-08-15]
42. Pytorch lightning. GitHub. URL: https://github.com/PyTorchLightning/pytorch-lightning [accessed 2022-01-01]
43. Suster S. Manually verified publication prediction dataset. Zenodo. 2022 Sep 27. URL: https://zenodo.org/record/7115724#.Y5x1j-xBxpQ [accessed 2022-12-16]
44. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge, Massachusetts, United States: Cambridge University Press; 2008.
45. Adhikari A, Ram A, Tang R, Lin J. DocBERT: BERT for document classification. arXiv 2019.
46. Sun C, Qiu X, Xu Y, Huang X. How to fine-tune BERT for text classification? In: Chinese Computational Linguistics. Cham: Springer; 2019.
47. Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN. Convolutional sequence to sequence learning. 2017 Presented at: International Conference on Machine Learning; 2017; Sydney, Australia.
48. Herzig J, Nowak PK, Müller T, Piccinno F, Eisenschlos E. TaPas: weakly supervised table parsing via pre-training. 2020 Presented at: Annual Meeting of the Association for Computational Linguistics; Jul, 2020; Online. [doi: 10.18653/v1/2020.acl-main.398]
49. Shamliyan TA, Kane RL. Availability of results from clinical research: failing policy efforts. J Epidemiol Glob Health 2014 Mar;4(1):1-12 [FREE Full text] [doi: 10.1016/j.jegh.2013.08.002] [Medline: 24534330]

50. Trinquart L, Dunn AG, Bourgeois FT. Registration of published randomized trials: a systematic review and meta-analysis. BMC Med 2018 Oct 16;16(1):173 [FREE Full text] [doi: 10.1186/s12916-018-1168-6] [Medline: 30322399]
51. Khalilzadeh J, Tasci AD. Large sample size, significance level, and the effect size: solutions to perils of using big data for academic research. Tourism Manage 2017 Oct;62:89-96. [doi: 10.1016/j.tourman.2017.03.026]
52. Lin M, Lucas HC, Shmueli G. Research commentary—too big to fail: large samples and the p-value problem. Inform Syst Res 2013 Dec;24(4):906-917. [doi: 10.1287/isre.2013.0480]
53. Manually verified publication prediction dataset. Zenodo. 2022 Sep 27. URL: https://doi.org/10.5281/zenodo.7115724 [accessed 2022-12-06]
54. Miron L, Gonçalves RS, Musen MA. Obstacles to the reuse of study metadata in ClinicalTrials.gov. Sci Data 2020 Dec 18;7(1):443 [FREE Full text] [doi: 10.1038/s41597-020-00780-z] [Medline: 33339830]
55. Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider. BMJ 2018 May 25;361:k1452 [FREE Full text] [doi: 10.1136/bmj.k1452] [Medline: 29802130]
56. Chaturvedi N, Mehrotra B, Kumari S, Gupta S, Subramanya HS, Saberwal G. Some data quality issues at ClinicalTrials.gov. Trials 2019 Jun 24;20(1):378 [FREE Full text] [doi: 10.1186/s13063-019-3408-2] [Medline: 31234923]
57. Wallace BC, Kuiper J, Sharma A, Zhu MB, Marshall IJ. Extracting PICO sentences from Clinical Trial Reports using Supervised Distant Supervision. J Mach Learn Res 2016;17:132 [FREE Full text] [Medline: 27746703]
58. Mintz M, Bills S, Snow R, Jurafsky D. Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2. 2009 Presented at: Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2; Aug 2 - 7, 2009; Suntec Singapore. [doi: 10.3115/1690219.1690287]
59. Boudin F, Nie JY, Dawes M. Positional language models for clinical information retrieval. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2010 Presented at: Conference on Empirical Methods in Natural Language Processing; Oct 9-11, 2010; Cambridge, MA.
60. Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. PLoS Med 2013 Dec 3;10(12):e1001566; discussion e1001566 [FREE Full text] [doi: 10.1371/journal.pmed.1001566] [Medline: 24311990]
61. Tice D, Carroll K, Bhatt K, Belknap S, Mai D, Gipson H, et al. Characteristics and causes for Non-Accrued Clinical Research (NACR) at an academic medical institution. J Clin Med Res 2013 Jun;5(3):185-193 [FREE Full text] [doi: 10.4021/jocmr1320w] [Medline: 23671544]
62. Weber EJ, Callaham ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. JAMA 1998 Jul 15;280(3):257-259. [doi: 10.1001/jama.280.3.257] [Medline: 9676674]
63. Wang S, Suster S, Baldwin T, Verspoor K. Predicting Publication of Clinical Trials Using Structured and Unstructured Data. Zenodo. 2022 Feb 28. URL: https://zenodo.org/record/6302910#.Y5x2CuxBxpQ [accessed 2022-12-16]

Abbreviations

AUC: area under the curve
BERT: Bidirectional Encoder Representations from Transformers
ML: machine learning
NLP: natural language processing
NN: neural network
RF: random forest
SciBERT: Bidirectional Encoder Representations from Transformers model for scientific texts

Edited by G Eysenbach; submitted 19.04.22; peer-reviewed by A Dunn, M Syed, R Zhang; comments to author 13.07.22; revised version received 14.10.22; accepted 16.11.22; published 23.12.22.

Please cite as:
Wang S, Šuster S, Baldwin T, Verspoor K
Predicting Publication of Clinical Trials Using Structured and Unstructured Data: Model Development and Validation Study
J Med Internet Res 2022;24(12):e38859
URL: https://www.jmir.org/2022/12/e38859
doi:10.2196/38859
PMID:36563029

Original Paper

Factors Associated With the Acceptance of an eHealth App for Electronic Health Record Sharing System: Population-Based Study

Junjie Huang[1,2], PhD; Wing Sze Pang[1], BSocSci; Yuet Yan Wong[1], MPH; Fung Yu Mak[1], BSocSci; Florence S W Chan[3], MPH; Clement S K Cheung[3], MD; Wing Nam Wong[3], MD; Ngai Tseung Cheung[3], MD; Martin C S Wong[1,2], MD

[1]The Jockey Club School of Public Health and Primary Care, Faculty of Medicine, Chinese University of Hong Kong, Sha Tin, Hong Kong

[2]Centre for Health Education and Health Promotion, Faculty of Medicine, Chinese University of Hong Kong, Sha Tin, Hong Kong

[3]Information Technology and Health Informatics Division, Hospital Authority, Kowloon, Hong Kong

Corresponding Author:
Martin C S Wong, MD
The Jockey Club School of Public Health and Primary Care
Faculty of Medicine
Chinese University of Hong Kong
Room 407, 4/F, School of Public Health Building
Prince of Wales Hospital
Sha Tin, 999077
Hong Kong
Phone: 852 2252 8782
Email: wong_martin@cuhk.edu.hk

Abstract

Background: In the second stage of the Electronic Health Record Sharing System (eHRSS) development, a mobile app (eHealth app) was launched to further enhance collaborative care among the public sector, the private sector, the community, and the caregivers.

Objective: This study aims to investigate the factors associated with the downloading and utilization of the app, as well as the awareness, perception, and future improvement of the app.

Methods: We collected 2110 surveys; respondents were stratified into 3 groups according to their status of enrollment in the eHRSS. The primary outcome measure was the downloading and acceptance of the eHealth app. We collected the data on social economics factors, variables of the Technology Acceptance Model and Theory of Planned Behavior. Any factors identified as significant in the univariate analysis ($P<.20$) will be included in a subsequent multivariable regression analysis model. All P values ≤.05 will be considered statistically significant in multiple logistic regression analysis. The structural equation modeling was performed to identify interactions among the variables.

Results: The respondents had an overall high satisfaction rate and a positive attitude toward continuing to adopt and recommend the app. However, the satisfaction rate among respondents who have downloaded but not adopted the app was relatively lower, and few of them perceived that the downloading and acceptance processes are difficult. A high proportion of current users expressed a positive attitude about continuing to adopt and recommend the app to friends, colleagues, and family members. The behavioral intention strongly predicted the acceptance of the eHealth app (β=.89; $P<.001$). Attitude (β=.30; $P<.001$) and perceived norm; β=.37; $P<.001$) played important roles in determining behavioral intention, which could predict the downloading and acceptance of the eHealth app (β=.14; $P<.001$).

Conclusions: Despite the high satisfaction rate among the respondents, privacy concerns and perceived difficulties in adopting the app were the major challenges of promoting eHealth. Further promotion could be made through doctors and publicity. For future improvement, comprehensive health records and tailored health information should be included.

(J Med Internet Res 2022;24(12):e40370) doi:10.2196/40370

KEYWORDS

digital health; eHealth; electronic health record; system; mobile app; app; public; private; community; caregiver; awareness; perception; improvement; utility; technology; model; health information

Introduction

In Hong Kong, a substantial proportion of hospital services is provided by the public sector (90% of all in-patient bed-days) and up to 70% of the outpatient services are offered by the private sector [1]. In view of the dual-track health care system, the Electronic Health Record Sharing System (eHRSS) was developed by the Hospital Authority (HA) to facilitate the information flow between the public and private sectors. It was launched in March 2016 [2] as an electronic platform to provide accurate and quick retrieval of clinical details, such as patient demographics, clinical information, and prescription profiles. The benefits of eHRSS are facilitation of patient communication, improvement of patient care continuity, accuracy of drug prescription, and enablement of holistic management [3]. Stage 2 development of the eHRSS started in July 2017, which further expanded the benefits to the relevant stakeholders and users. These include the broadening of the scope of sharable data by the system; provision of patients' choice over data sharing scope; and their access to some of the data in the eHRSS [4]. As of May 2022, over 5.5 million citizens, 50,000 health care professionals, all the 13 private hospitals, and over 2400 health care professionals working in the private sectors have enrolled in the eHRSS [5].

In stage 2 development, a mobile app, an "eHealth app," was launched in January 2021 [6]. It facilitates users to access their integrated health records and manage own health. Our team has previously evaluated the perceptions of and factors associated with the acceptance of the eHRSS in 2018 among 2000 patients in Hong Kong [7]. More than 70% (707/1000, 70.70%) of the patients were satisfied with the overall performance of the eHRSS. The expansion of sharable scope in the eHRSS (32/124, 25.8%) and allowing patients to access their medical records (30/124, 24.2%) were considered as the features to be developed in the future development of the eHRSS by the enrollees. This is one of the survey findings that provides support for the second stage of the eHRSS, where the users may access their health records and other health information via the utilization of an eHealth app. This mobile app further enhances collaborative care among the public sector, private sector, community, and caregivers. Importantly, citizens could be empowered in self-health management and disease prevention by recording health data within the app. It further empowers citizens' self-care ability by involving family members and other stakeholders to understand their current health status.

Across the world, similar mobile health apps were developed for people to upload and view health records, manage personal health care activities, share clinical information with doctors, and improve public health. Apps such as "Capzule PHR," "Health and Family," and "Health Notes" allow patients to view and get access to their medical information and record their data at any time and any place through the internet or by offline access [8]. The government of various countries is promoting electronic medical health records. For example, "MIDATA" is the UK government program with the goal of providing consumers a better control over their data [9]. The Mi Health App was developed accordingly to record health data and support long-term health management [10]. In 2019, the Korean government initiated the "MyData" program, which aims to give citizens increased access to personal data through mobile phones. In the medical field, it enables the public to manage their medical record [11]. The My HealthWay app was launched in 2021 by the Korean Ministry of Health and Welfare to integrate scattered medical data [12].

To further promote quality and efficiency, as well as to recommend the future development of the mobile (eHealth) app, perceptions and views from users are required to inform a more system-friendly design. The objectives of this project are to evaluate the factors associated with the downloading and utilization of the eHealth app; to examine the awareness, use, and acceptability of the mobile eHealth app; to explore whether eHealth app use may be associated with the joining of the eHRSS; the reasons for nonuse among those who joined the eHRSS; the extent to which the app improves user experience and influences health service utilization; and to recommend a potential room for improvement of the eHealth apps.

Methods

Sampling Frame and Recruitment

A self-administered questionnaire was adopted in this study. Prospective study participants were based on a list of patients provided by the HA. A simple random sampling methodology was mainly used. Over 5.5 million existing eHRSS users were included in the population, and computer-generated numbers were listed correspondingly for participant recruitment. An invitational SMS was first sent by the HA to existing eHRSS users. This served to alert the participants that they would receive a subsequent survey invitation by Chinese University of Hong Kong via SMS [13]. Then research teams at the Chinese University of Hong Kong sent messages to those people who had received an invitation from the HA through a bulk SMS sending platform (MD SMS by Media Digital Technologies Corporation Limited). Supplemented by a convenience sampling methodology, the online survey link was shared on the website of the HA, eHealth Facebook and Instagram page, eHealth app, eHealth website so that both health care recipients and non–health care recipients could access the questionnaires. The overall response rate was 66.71% (3026/4536).

Survey Instruments

Survey items focused on the awareness, use, and acceptability of the eHealth app; the association between the use of the eHealth app and the joining of the eHRSS; the reasons why some users did not use the eHealth app after joining the eHRSS; the extent to which the eHealth app improved user experience, modified health service access, and health management; and recommendations for possible improvement of the eHealth app.

The surveys were designed by an academic physician with relevant experience in projects related to the eHRSS, and extensive expertise in both clinical and public health research studies. The questionnaire draft was face-validated by a panel of epidemiologists, biostatisticians, and professionals in the field of health care policy, public health, and primary care. It was subsequently pilot tested for feasibility and item comprehensiveness among 20 people. The completion rate was 65% (13/20), and the average response time was 7 minutes and 40 seconds (Multimedia Appendix 1).

The surveys were available in both Chinese and English versions. All surveys were anonymous, and written consent was provided by the participants at the start of the questionnaire. The study participants were informed that all information presented would be in the form of aggregated data that could not identify any individuals.

Ethics Approval

This study was approved by the Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong (approval number SBRE-21-0184).

Statistical Analysis

All surveys were checked for their completeness and the presence of participant consent. All data entry and analysis were conducted using SPSS version 26.0 (IBM, Inc.). As part of quality control, at least 20% (422/2110) of all surveys were randomly checked for the validity, quality, and accuracy. All items in the survey were analyzed as stratified according to the status of enrollment. The primary outcome measure was the downloading and acceptance of the eHealth app. We tested for the presence of statistical association by identifying potential associated factors using univariate and multivariate regression analyses. We included age, gender, educational level, job status, monthly household income, the types of mobile phone operating systems currently in use, the eHRSS enrollment status, perceived enablers of acceptance, and perceived barriers of the eHealth app use. Any factors identified as significant in univariate analysis (P<.20) will be included in a subsequent multivariable regression analysis model. All Pvalues ≤.05 will be considered statistically significant in the multiple logistic regression analysis. Assuming the proportion of the primary outcomes was 50%, which would provide the largest sample size, a total of 2110 surveys would result in precision of approximately 2.2%. In addition, we performed structural equation modeling to identify interactions among the variables.

Health Behavioral Models

To investigate the factors that could predict downloading and acceptance of the eHealth app, we used 2 internationally recognized models that have been widely adopted to examine the use of new technologies. These were the Technology Acceptance Model (TAM), which was first developed by Fred D Davis, Richard P Bagozzi, and Paul R Warshaw [14]. It is an adaptation of the Theory of Reasoned Action (TRA) to the discipline of information systems. The TAM hypothesizes that perceived usefulness and perceived ease of use could influence an individual's intention to use an information system [15]. The meditator of actual acceptance of the system is the intention to use. The model also considered perceived ease of use as a direct determinant of perceived usefulness. The TAM has been simplified by omitting the construct pertinent to attitude, as used in the TRA model. In the survey, perceived usefulness has been assessed using a series of questions related to the convenience and the benefit of using the app. To measure the ease of use, respondents have been asked about their experience in the downloading and acceptance processes, whether the app is easy to download, easy to find function, and contains the health information they want. For perceived barriers, respondents were asked about factors preventing them from downloading or adopting the app, such as doctors do not recommend or participate and concerns about personal privacy (Multimedia Appendix 2).

Furthermore, we employed the Theory of Planned Behavior (TPB), a commonly used psychological theory that links people's beliefs and behaviors [16]. The underpinning theory identified 3 core predictors, namely, attitude (A1-4), subjective norms (SN1-3), and perceived behavioral control (BI1-2) as modifiers of intention. Items from these 3 constructs, for example, suggestions from people who influence users' behavior, were recoded into the questionnaire as measurement (Multimedia Appendix 3) [17,18]. A 5-point Likert scale (strongly disagree, disagree, neutral, agree, and strongly agree) and a 2-point Likert scale (yes and no) were used in the survey design. Survey questions related to the acceptance and use of the app were designed according to the components of the TAM and TPB models. In our survey, attitude was the measurement of enabling factors, and the subjective norm referred to how the respondents viewed the idea of other people's perceptions about the app, including the recommendation from doctors, friends, and family members. The specific questions related to attitude and subjective norm are "Do you agree with the following reasons that can increase your motivation to continuously use/start to use the eHealth app" and "Do you agree with the following reasons that hinder your motivation to continuously use/start to use the eHealth App" (Multimedia Appendix 4). The theory hypothesized that behavioral intention is the most antecedent influencer of behavior. In the current structural equation modeling, we included the following additional variables into the TAM: age, gender, educational level, occupation, types of mobile phone operating systems, and enrollment status of the eHRSS. All P values ≤.05 were regarded as statistically significant.

Results

Participant Characteristics

A total of 2110 completed surveys were collected (Table 1). Overall, there were more male than female respondents (1184/2110, 56.11%, vs 926/2110, 43.89%). Among the study participants, 46.16% (974/2110) were aged between 41 and 60, while 39.72% (838/2110) were aged above 60. Over one-half of the respondents attained secondary educational level (1118/2107, 53.06%). Nearly half of the respondents had full-time or part-time jobs (999/2024, 49.36%). For income level, the highest proportion of monthly household income was HK $10,000-19,999 (1HK $=US $0.12; 458/2110, 26.44%).

Table 1. Participant characteristics (N=2110).

Characteristics	Values, n (%)
Age (years)	
16-30	136 (6.45)
31-40	162 (7.68)
41-50	343 (16.26)
51-60	631 (29.91)
61-70	636 (30.14)
>70	202 (9.57)
Gender	
Male	1184 (56.11)
Female	926 (43.89)
Educational level (n=2107)	
Primary or below	150 (7.12)
Secondary	1118 (53.06)
Tertiary or above	839 (39.82)
Other	3 (not counted)[a]
Job status (n=2024)	
Employed (Full-time/part-time)	999 (49.36)
Unemployed	100 (4.94)
Retired	695 (34.34)
Housewives	138 (6.82)
Students	53 (2.62)
Others	39 (1.93)
Refuse to answer	86 (not counted)[a]
Monthly household income (HK $; n=1732)[b]	
<10,000	373 (21.54)
10,000-19,999	458 (26.44)
20,000-29,999	335 (19.34)
30,000-39,999	154 (8.89)
40,000-59,999	180 (10.39)
≥60,000	232 (13.39)
Refuse to answer	378 (not counted)[a]
Phone currently in use	
Apple iOS	700 (33.18)
Android	1110 (52.61)
Huawei	174 (8.25)
Others	126 (5.97)

[a]As these options are out of the original categories, the answers were “not counted” and thus not used in the analysis.

[b]1HK $=US $0.12.

Participants were classified into several groups according to downloading and acceptance of the eHealth app (Multimedia Appendix 5). A total of 1242 respondents have enrolled in the eHRSS, downloaded, and adopted the eHealth app (group 1). There were 275 participants who have enrolled in the eHRSS, downloaded the eHealth app, but did not adopt the app (group 2). The third group included 203 respondents that have enrolled in the eHRSS, but have neither downloaded nor adopted the

app (group 3). In the following paragraphs, the findings were stratified according to these 3 groups of respondents.

The COVID-19 vaccination program (649/2110, 30.76%), medical doctors (647/2110, 30.66%), publicity (posters, pamphlets, television, outdoor advertisements; 533/2110, 25.26%), and friends or family members (388/2110, 18.39%) were the 4 major sources of information about the eHealth app among respondents (Multimedia Appendix 6). We did not observe a distinct difference in the distribution of sources among the 3 groups.

Perceived Enablers and Barriers of the App

In group 1, the majority of participants agreed that the app can show their accurate vaccination records (1118/1242, 90.02%) and other health records (1081/1242, 87.04%). They also expressed that the app provides useful administrative functions, including giving consent to health care providers for sharing their data (1044/1242, 84.06%), easier management of eHealth accounts (1005/1242, 80.92%), and empowerment of their family members and own health (940/1242, 75.68%). A similar result was also noted in the other 2 groups (Tables 2 and 3).

Among the study participants in group 1 (Tables 4 and 5), the major barrier was that their physicians had not joined the eHealth app (505/1028, 49.12%) and that their doctors did not mention, recommend, or think it is necessary to use the eHealth app (417/1078, 38.68%). Respondents in groups 2 and 3 perceived that the downloading procedure is complicated (172/382, 45%) and were concerned about their personal information and privacy (243/461, 52.7%), respectively.

Table 2. Perceived enablers of downloading the eHealth app.

Enablers of downloading	Downloaded and used eHealth app (n=1242)	Downloaded but not used eHealth app (n=399)	Not having downloaded and used eHealth app (n=469)
	Strongly agree or agree, n (%)	Strongly agree or agree, n (%)	Strongly agree or agree, n (%)
It is convenient to get information about different government-subsidized medical programs	920 (74.07)	293 (73.43)	332 (70.79)
I can view my accurate health records	1081 (87.04)	309 (77.44)	380 (81.02)
I can manage my eHealth account easily (eg, update the communication means)	1005 (80.92)	281 (70.43)	359 (76.55)
I can give sharing consents to health care providers easily so that they can view my health records	1044 (84.06)	307 (76.94)	378 (80.60)
I can find the health care providers and doctors that are participating in different health programs with ease	899 (72.38)	269 (67.42)	368 (78.46)
I can check the remaining balance and record of the Elderly Health Care Voucher Scheme	904 (72.79)	270 (67.67)	371 (79.10)
I can show the vaccination record/QR code	1118 (90.02)	321 (80.45)	383 (81.66)
It helps to manage my health and my families' health	940 (75.68)	274 (68.67)	367 (78.25)
My friend recommended me to use the "eHealth" app	691 (55.64)	202 (50.63)	244 (52.03)
My family recommended me to use the "eHealth" app	777 (62.56)	225 (56.39)	282 (60.13)
My doctor recommended me to use the "eHealth" app	797 (64.17)	240 (60.15)	312 (66.52)
Government's advertisement of the "eHealth" app	730 (58.78)	216 (54.14)	271 (57.78)
I can get souvenirs	466 (37.52)	148 (37.09)	201 (42.86)

Table 3. Perceived enablers of acceptance of the eHealth app.

Enablers of acceptance	Downloaded and used eHealth app (n=1242)			Downloaded but not used eHealth app (n=399)			Not having downloaded and used eHealth app (n=469)		
	n	Mean (SD)	95% CI	n	Mean (SD)	95% CI	n	Mean (SD)	95% CI
It is convenient to get information about different government-subsidized medical programs	920	3.84 (0.78)	3.80-3.89	293	3.75 (0.85)	3.66-3.83	332	3.76 (0.75)	3.69-3.82
I can view my accurate health records	1081	4.15 (0.79)	4.11-4.20	309	3.87 (0.87)	3.78-3.96	380	3.96 (0.71)	3.89-4.02
I can manage my eHealth account easily (eg, update the communication means)	1005	3.99 (0.73)	3.95-4.03	281	3.70 (0.86)	3.62-3.79	359	3.86 (0.71)	3.79-3.92
I can give sharing consents to health care providers easily so that they can view my health records	1044	4.07 (0.73)	4.03-4.11	307	3.84 (0.85)	3.75-3.92	378	3.92 (0.71)	3.86-3.99
I can find the health care providers and doctors that are participating different health programs with ease	899	3.86 (0.75)	3.82-3.90	269	3.68 (0.80)	3.61-3.76	368	3.89 (0.69)	3.82-3.95
I can check the remaining balance and record of the Elderly Health Care Voucher Scheme	904	3.90 (0.81)	3.86-3.95	270	3.71 (0.86)	3.63-3.80	371	3.88 (0.72)	3.82-3.95
I can show the vaccination record/QR code	1118	4.22 (0.72)	4.18-4.26	321	3.95 (0.86)	3.86-4.03	383	4.00 (0.73)	3.93-4.06
It helps to manage my health and my families' health	940	3.93 (0.79)	3.89-3.98	274	3.73 (0.89)	3.64-3.82	367	3.89 (0.72)	3.83-3.96
My friend recommended me to use the "eHealth" app	691	3.55 (0.91)	3.50-3.60	202	3.41 (0.94)	3.32-3.50	244	3.44 (0.86)	3.37-3.52
My family recommended me to use the "eHealth" app	777	3.68 (0.89)	3.63-3.73	225	3.5 (0.95)	3.40-3.59	282	3.56 (0.87)	3.48-3.64
My doctor recommended me to use the "eHealth" app	797	3.7 (0.88)	3.65-3.75	240	3.58 (0.88)	3.49-3.67	312	3.71 (0.77)	3.64-3.78
Government's advertisement of the "eHealth" app	730	3.61 (0.88)	3.56-3.66	216	3.49 (0.92)	3.40-3.58	271	3.52 (0.86)	3.44-3.59
m. I can get souvenirs	466	3.21 (1.08)	3.15-3.27	148	3.15 (1.05)	3.05-3.25	201	3.24 (0.99)	3.15-3.33

Table 4. Perceived barriers to downloading of the eHealth app.

Barrier	Downloaded and used the eHealth app (n=1028-1222)	Downloaded but not used the eHealth app (n=301-391)	Not having downloaded and used the eHealth app (n=365-461)
	Strongly agree or agree, n (%)	Strongly agree or agree, n (%)	Strongly agree or agree, n (%)
One's physician has not joined	505/1028 (49.12)	133/310 (42.90)	151/365 (41.37)
Only see 1 doctor who is familiar with my health records	392/1092 (35.90)	144/347 (41.50)	181/425 (42.59)
No sickness	295/1157 (25.50)	97/358 (27.09)	156/441 (35.37)
Concerned about personal information and privacy	408/1222 (33.39)	168/388 (43.30)	243/461 (52.71)
My doctor did not mention about/recommend/think it is necessary to use the "eHealth" app	417/1078 (38.68)	136/333 (40.84)	183/403 (45.41)
I do not know how to use a smartphone/mobile app	203/1167 (17.40)	94/372 (25.27)	119/441 (26.98)
Not willing for others to read one's own health records	372/1216 (30.59)	161/391 (41.18)	209/455 (45.93)
Uncertain about the benefits of the eHealth app	266/1198 (22.20)	134/374 (35.83)	172/437 (39.36)
Complicated downloading procedures	321/1216 (26.40)	172/382 (45.03)	173/423 (40.90)

Table 5. Perceived barriers to acceptance of the eHealth app.

Barrier	Downloaded and used the eHealth app (n=1242)			Downloaded but not used the eHealth app (n=399)			Not having downloaded and used the eHealth app (n=469)		
	n	Mean (SD)	95% CI	n	Mean (SD)	95% CI	n	Mean (SD)	95% CI
One's physician has not joined	505	3.30 (1.08)	3.23-3.37	133	3.30 (0.92)	3.18-3.41	151	3.16 (0.95)	3.05-3.26
Only see 1 doctor who is familiar with my health records	392	3.03 (1.02)	2.96-3.09	144	3.09 (0.93)	2.97-3.2	181	3.12 (0.92)	3.02-3.23
No sickness	295	2.73 (1.02)	2.66-2.80	97	2.83 (0.95)	2.71-2.95	156	2.97 (0.97)	2.86-3.08
Concerned about personal information and privacy	408	2.94 (1.14)	2.86-3.02	168	3.09 (1.05)	2.96-3.22	243	3.46 (1.06)	3.34-3.58
My doctor did not mention about/recommend/think it is necessary to use the "eHealth" app	417	3.11 (0.98)	3.05-3.18	136	3.22 (0.84)	3.12-3.33	183	3.23 (0.86)	3.14-3.33
I do not know how to use a smartphone/mobile app	203	2.43 (1.11)	2.36-2.51	94	2.70 (1.07)	2.57-2.83	119	2.78 (1.00)	2.67-2.89
Not willing for others to read one's own health records	372	2.89 (1.09)	2.82-2.96	161	3.08 (0.99)	2.95-3.2	209	3.28 (1.00)	3.17-3.39
Uncertain about the benefits of the eHealth app	266	2.70 (1.04)	2.63-2.77	134	3.06 (0.95)	2.95-3.18	172	3.15 (0.93)	3.05-3.26
Complicated downloading procedures	321	2.79 (1.05)	2.73-2.86	172	3.20 (1.01)	3.07-3.32	173	3.23 (0.88)	3.13-3.33

Perception of Processes of Acceptance of the App

The proportion of participants in group 1 who were positive about the downloading and acceptance processes was in general higher than those in group 2. Most respondents in group 1 were satisfied with the downloading processes (908/1242, 73.11%; Multimedia Appendix 7). However, the proportion of group 2 participants expressing satisfaction about the downloading process was lower (239/399, 59.90%). Regarding the acceptance process, respondents in group 1 were satisfied with the app's user experience and interface. They agreed that the fonts and size of the words were easy to read (947/1242, 76.25%), that the icon and tables were easy to understand (880/1242, 70.85%), and that the app was easy to use overall (869/1242, 69.97%). Among respondents in group 2, 60.6% (242/399) agreed that the fonts and size of the words were easy to read, and nearly half of them agreed that the icons and tables were easy to understand (190/399, 47.6%).

Applicability and Perception of the App

In terms of applicability, vaccine records (1108/1242, 89.21%), appointment records (1055/1242, 84.94%), medication records (1015/1242, 81.72%), allergy records (924/1242, 74.40%), and health management (786/1242, 63.29%) were the top 5 useful functions among the users (Multimedia Appendix 8). These proportions were higher in group 1 than in group 2.

Turning to the perception of the app (Multimedia Appendix 9), a high percentage of group 1 respondents (ie, app users) were satisfied with the app overall (975/1242, 78.50%), agreed that it enhanced the experience of health services (962/1242, 77.46%), enhanced concerns about health information (926/1242, 74.56%), and enhanced management of health on their own (889/1242, 71.58%). Over 50% (211/399, 52.9%) agreed that the app improved the health of family members. Group 2 respondents (ie, nonusers) also reported a positive perception of the app, although the proportion agreeing with these items was lower.

Expectations on the Future Development of the App

A high proportion of group 1 respondents, current users, expressed a positive attitude about continuing to adopt (1105/1242, 88.97%) and recommend the app to friends, colleagues, and family members (1024/1242, 82.45%; Multimedia Appendix 10). The proportion agreeing to continuously use and recommend among the nonusers in groups 2 and 3 was also high. Over 70% and 60% of the respondents in groups 2 (283/399, 70.9%, and 290/399, 72.7%) and 3 (320/469, 68.2%, and 304/469, 64.8%), respectively, expressed positive attitude toward future acceptance and recommendation of the app, respectively. Among all respondents, they expected to access more health records via the app, for example, the laboratory results (1707/2110, 80.90%) and the radiographic images (1484/2110, 70.33%), and to have customized health information, for example, age-specific health care recommendations (1378/2110, 65.31%) and tailored health tips (1121/2110, 53.13%). In group 1, the inclusion of the laboratory result was the most frequently cited item (1094/1242, 88.08%), followed by radiographic images (980/1242, 78.90%) and age-specific health care recommendations (843/1242, 67.87%). The results were similar compared with responses in groups 2 and 3.

Factors Associated With Downloading and Acceptance

Respondents were more likely to download the app when they had joined the eHRSS (adjusted odds ratio [aOR] 9.2, 95% CI 6.35-13.32; P<.001); had attained secondary educational level (aOR 1.63, 95% CI 1.08- 2.46; P=.02); reported being able to view their accurate health records (aOR 1.41, 95% CI 1.02-1.95; P<.035); reported being able to show the vaccination records or QR codes (aOR 1.43, 95% CI 1.03-1.98; P=.031); and reported one's physician had not joined the eHRSS (aOR 1.45,

95% CI 1.18-1.77; *P*<.001; Tables 6 and 7). Housewives (aOR 0.44, 95% CI 0.23-0.84; *P*=.013) and participants who were concerned about personal information and privacy (aOR 0.74, 95% CI 0.60-0.90; *P*=.003) were significantly less likely to download the eHealth app.

The independent factors associated with the acceptance of the eHealth app were similar to those associated with downloading, except that male participants (aOR 1.85, 95% CI 1.36-2.52; *P*<.001) were more likely to adopt, whereas individuals with primary educational level or below (aOR 0.49, 95% CI 0.25-0.94; *P*=.03) and study participants who were uncertain about the benefits of the eHealth app (aOR 0.80, 95% CI 0.66-0.96; *P*=.02) or perceived the downloading procedures as complicated (aOR 0.81, 95% CI 0.68-0.96; *P*=.01) were less likely to adopt (Tables 6 and 7).

Table 6. Factors associated with downloading and acceptance of the eHealth app.

Factor	Users, n (n=1159)	Downloading			Acceptance		
		Values, n (%)	aOR[a] (95% CI)	*P* value	Values, n (%)	aOR (95% CI)	*P* value
Age (years)				.63			.53
16-40	150	105 (70)	1 (reference)		82 (54.7)	1 (reference)	
41-60	571	440 (77.1)	1.22 (0.70-2.11)	.48	347 (60.8)	1.31 (0.81-2.13)	.27
>60	438	361 (82.4)	1.40 (0.71-2.78)	.33	280 (63.9)	1.35 (0.75-2.43)	.32
Gender							
Male	680	553 (81.3)	1.19 (0.83-1.73)	.35	458 (67.4)	1.85 (1.36-2.52)	<.001
Female	479	353 (73.7)	1 (reference)		251 (52.4)	1 (reference)	
Educational level				.03			.04
Primary or below	73	50 (68.5)	0.91 (0.44-1.91)	.81	32 (43.8)	0.49 (0.25-0.94)	.03
Secondary	617	491 (79.6)	1.63 (1.08-2.46)	.02	373 (60.5)	1.05 (0.75-1.48)	.76
Tertiary or above	469	365 (77.8)	1 (reference)		304 (64.8)	1 (reference)	
Job status				.01			.48
Full-time/part-time	642	504 (78.5)	1 (reference)		404 (62.9)	1 (reference)	
Unemployed	49	36 (73.5)	1.38 (0.59-3.21)	.46	26 (53.1)	1.21 (0.57-2.56)	.62
Retired	352	297 (84.4)	1.12 (0.67-1.88)	.67	230 (65.3)	0.89 (0.58-1.35)	.58
Housewives	74	45 (60.8)	0.44 (0.23-0.84)	.01	29 (39.2)	0.63 (0.34-1.15)	.13
Students	22	13 (59.1)	0.49 (0.15-1.57)	.23	11 (50)	0.82 (0.27-2.52)	.73
Others	20	11 (55)	0.27 (0.09-0.82)	.02	9 (45)	0.47 (0.16-1.38)	.17
Monthly household income (HK $)[b]				.27			.82
<10,000	228	170 (74.6)	1 (reference)		118 (51.8)	1 (reference)	
10,000-19,999	300	227 (75.7)	0.91 (0.55-1.51)	.72	177 (59)	1.20 (0.78-1.85)	.41
20,000-29,999	225	174 (77.3)	0.74 (0.43-1.29)	.29	141 (62.7)	1.07 (0.67-1.70)	.79
≥30,000	406	335 (82.5)	1.22 (0.71-2.08)	.47	273 (67.2)	1.18 (0.75-1.86)	.46
Phone currently in use				.05			.19
Apple iOS	392	295 (75.3)	1 (reference)		239 (61)	1 (reference)	
Android	615	501 (81.5)	1.22 (0.82-1.82)	.34	391 (63.6)	0.93 (0.67-1.31)	.69
Huawei	93	73 (78.5)	1.24 (0.63-2.46)	.53	54 (58.1)	0.83 (0.47-1.46)	.51
Others	59	37 (62.7)	0.46 (0.22-0.97)	.04	25 (42.4)	0.47 (0.24-0.94)	.03
Joining of eHRSS[c]							
Yes	924	807 (87.3)	9.20 (6.35-13.32)	<.001	665 (72)	9.77 (6.64-14.38)	<.001
No	235	99 (42.1)	1 (reference)		44 (18.7)	1 (reference)	

[a]aOR: adjusted odds ratio.

[b]1HK $=US $0.12.

[c]eHRSS: electronic Health Record Sharing System.

Table 7. Factors associated with perceived enablers and barriers of the eHealth app.

Factors	aOR[a] (95% CI)	*P* value	aOR[a] (95% CI)	*P* value
Perceived enablers (score: 1 [strongly disagree] to 5 [strongly agree])				
It is convenient to get information about different government-subsidized medical programs	0.94 (0.69-1.28)	.70	0.95 (0.74-1.23)	.71
I can view my accurate health records	1.41 (1.02-1.95)	.04	1.40 (1.08-1.81)	.01
I can manage my eHealth account easily (eg, update the communication means)	0.82 (0.55-1.22)	.32	1.26 (0.90-1.75)	.18
I can give sharing consents to health care providers easily so that they can view my health records	1.14 (0.80-1.63)	.47	1.12 (0.84-1.50)	.44
I can find the health care providers and doctors who participated in different health programs with ease	0.49 (0.33-0.73)	.001	0.62 (0.45-0.85)	.003
I can check the remaining balance and record of the Elderly Health Care Voucher Scheme	0.99 (0.70-1.40)	.95	1.03 (0.78-1.37)	.82
I can show the vaccination record/QR code	1.43 (1.03-1.98)	.03	1.33 (1.02-1.75)	.03
It helps to manage my health and my families' health	0.73 (0.51-1.06)	.09	0.76 (0.56-1.01)	.06
My friend recommended me to use the "eHealth" app	1.28 (0.88-1.86)	.20	0.98 (0.72-1.35)	.92
My family recommended me to use the "eHealth" app	1.10 (0.75-1.62)	.63	1.14 (0.82-1.59)	.42
My doctor recommended me to use the "eHealth" app	0.83 (0.60-1.13)	.23	0.85 (0.65-1.11)	.23
Government's advertisement of the "eHealth" app	1.00 (0.76-1.32)	.97	1.01 (0.80-1.27)	.96
I can get souvenirs	1.14 (0.93-1.39)	.22	1.13 (0.95-1.34)	.17
Perceived barriers (score 1 [strongly disagree] to 5 [strongly agree], discard those answering "N/A")				
One's physician has not joined	1.45 (1.18-1.77)	<.001	1.18 (1.01-1.39)	.04
Only see 1 doctor who is familiar with my health records	1.01 (0.82-1.26)	.90	1.08 (0.90-1.29)	.42
No sickness	0.94 (0.76-1.16)	.58	0.97 (0.81-1.16)	.75
Concerned about personal information and privacy	0.74 (0.60-0.90)	.003	0.89 (0.75-1.05)	.16
My doctor did not mention about/recommend/think it is necessary to use the "eHealth" app	1.20 (0.95-1.51)	.12	1.05 (0.87-1.27)	.58
I do not know how to use a smartphone/mobile app	0.97 (0.81-1.17)	.77	1.04 (0.89-1.22)	.62
Not willing for others to read one's own health records	1.05 (0.84-1.31)	.66	1.04 (0.87-1.24)	.68
Uncertain about the benefits of the eHealth app	0.81 (0.64-1.01)	.06	0.80 (0.66-0.96)	.02
Complicated downloading procedures	0.88 (0.71-1.08)	.23	0.81 (0.68-0.96)	.02

[a]aOR: adjusted odds ratio.

Findings From the Health Behavioral Models

In the TAM, perceived usefulness (β=.52; *P*<.001) and behavioral intention (β=.19; *P*<.001) were determined by perceived ease of use. The behavioral intention strongly predicted the acceptance of the eHealth app (β=.89; *P*<.001). Age (β=.07; *P*<.001) and whether the participant is a student or not (β=–0.09; *P*<.001) predicted the perceived usefulness. However, perceived usefulness did not significantly predict behavioral intention (β=.03; *P*=.32; Figure 1).

Turning to the TPB, attitude (β=.30; *P*<.001) and subjective norm (β=.37; *P*<.001) played important roles in determining behavioral intention, which could predict the downloading and acceptance of the eHealth app (β=.14; *P*<.001). The downloading and acceptance of the eHealth app could also be predicted by perceived behavior control (β=.14; *P*<.001). For the 3 core predictors, attitude was predicted by the subjective norm (β=.36; *P*<.001) and perceived behavior control (β=.23; *P*<.001). Subjective norm was predicted by attitude (β=.36; *P*<.001) and perceived behavior control (β=.11; *P*<.001). Perceived behavior control was predicted by attitude (β=.23; *P*<.001) and subjective norm (β=.27; *P*<.001; Figure 2).

Figure 1. Factors predictive of downloading and acceptance of the eHealth app by the Technology Acceptance Model. *P<.05 (2-tailed).

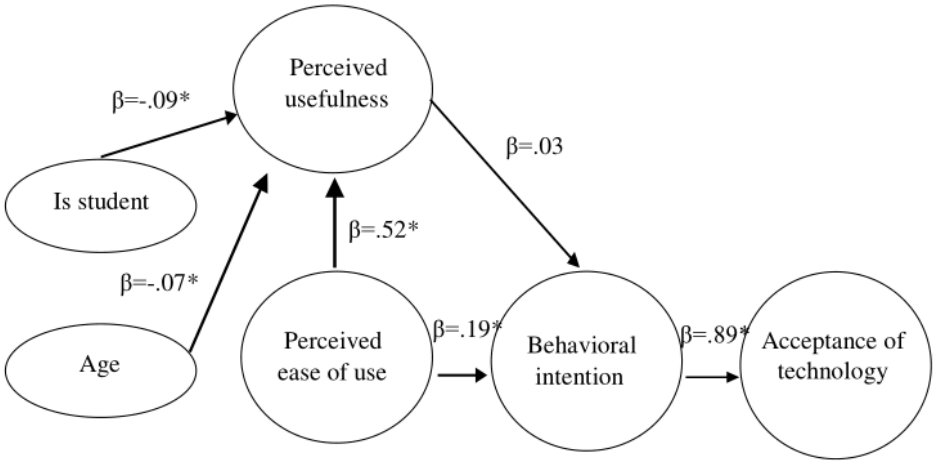
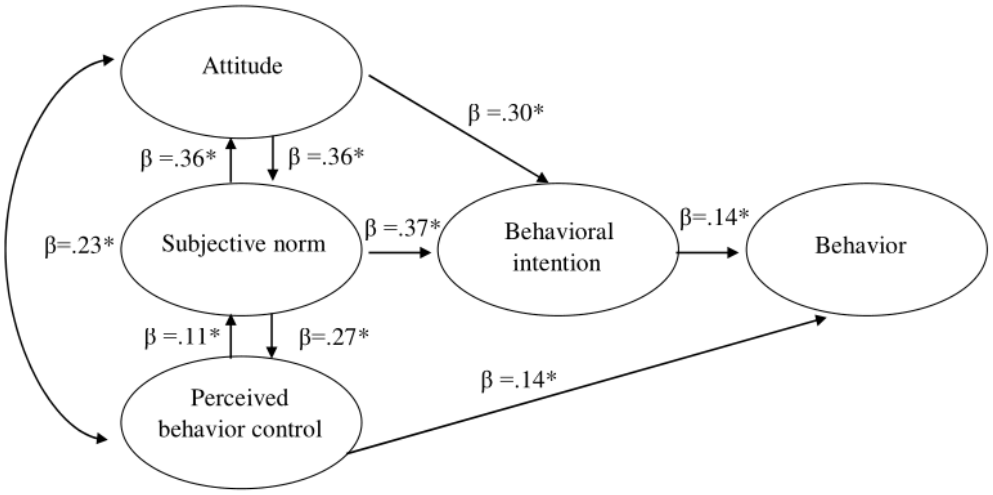


Figure 2. Factors predictive of downloading and acceptance of the eHealth app by the Theory of Planned Behavior. *P<.05 (2-tailed).



Discussion

Principal Findings

Overall, the satisfaction rate among the respondents was high. The satisfaction rate among group 2 respondents was relatively lower, and few of them perceived the downloading process as complicated. The willingness to continue to use and recommend the app was strong among all respondents. The 3 major enablers of adopting the app were the viewing of health records, especially the vaccination record; managing their eHealth accounts and sharing consent; and managing their family members' and their own health. However, respondents of the 3 groups had different perceived barriers. These include one's physician had not joined the eHRSS or had not recommended the eHealth app to them, a complicated downloading process, and privacy concerns. Most of the respondents expected to access more health records in the app, such as laboratory results and radiographic images, and have more personalized health information and health tips based on their age groups and health condition.

Limitations

This study has a few limitations. First, the survey was cross sectional, and so only the correlation could be measured instead of the causal relationship with the possibility of reverse causality. To corroborate the enablers and barriers, prospective longitudinal studies are required. In addition, face validity rather than construct validity was applied in the design of the questionnaire. The consistency reliability of the survey measurements has not yet been evaluated. Besides, some variables that could affect the downloading and acceptance of eHealth app may not be discussed in this study. Hence, there was a possibility of residual confounding. Finally, the study focused on acceptance of the app and examined individual factors affecting its use, which was based on a more individual level by using the TAM and TPB models in study design and analysis. Referring to Shachak et al's [19] study on the complexity of the health information technology implementation, a more sociotechnical-level study that examines the complex and overall cyber-social system in which users, cultures, networks, technologies, and processes are involved should be conducted in the future.

Comparison With Prior Work

eHealth app provides accurate and quick retrieval of clinical details for the citizens, as well as a platform for citizens to record self-monitoring health data. Thus, the app also facilitated the work of health care professionals with the integration and sharing of health records [5,7]. A medical app that contained medication, vaccine, and appointment records was convenient

for the users of health care services. This helps to contribute to a user-friendly system that enhanced more patients' use of the app. Among the eHealth app users in different studies, ease of use, user-friendliness, and availability of resources were the success factors facilitating the use of the app [20]. The eHealth app seems to empower the users to participate in health services, access health information, and manage their family members' and their own health, which has contributed to the overall satisfaction (975/1242, 78.50%) with the eHealth app.

Similar to our previous studies in 2020, many participants learned about the eHRSS from others, including medical doctors, posters, television, and outdoor advertisement [7,21]. However, the occurrence of the coronavirus pandemic has raised public awareness of eHealth technology [22]. Our results showed that the COVID-19 vaccination program has become the major source for people to learn about the app. The practical use of the eHealth app, including COVID-19 vaccination record and vaccine pass, has encouraged a large group of citizens to download and adopt the eHealth app. Based on the systematic analysis of 8 studies from the United States, China, and Switzerland, patient engagement has been enhanced by eHealth technologies, as these supported contact tracing and improved access to surveillance data [23]. A group of Canadian scholars found that the use of an eHealth app could be enhanced and made available widely in a pandemic context when eHealth technologies are integrated with public health policy and programs, which in turn could facilitate the flow of information and communication [24]. These helped to explain why the downloading and acceptance of the eHealth app, as a medical informatics technology, had a large increase during the pandemic.

The participation of doctors was decisive to encourage the citizens to download and use the eHealth app. Our previous study in 2020 found that the actual use of the eHRSS among patients was also significantly associated with the enrollment among physicians [7]. Giving sharing consent to health care providers was one of the major enablers for people to download and use the app. However, if their doctor did not join eHealth, it is of no use for them to give sharing consent to the doctor. This may lower the perceived usefulness of the app and discourage people to adopt. In our result, the TPB implied that subjective norm, doctor's recommendation, could largely determine the participants' willingness to download and adopt the app. The downloading and acceptance processes have been found satisfactory in the responses, especially among the respondents in group 1. However, the respondents hesitated to adopt the app because of perceived complicated downloading procedures. The eHealth app had users with a wide range of demographic characteristics and different levels of technical proficiency. Besides, the elderly and less educated citizens might have difficulties in adopting mobile apps. It was also found that the respondents in group 2 reported a lower satisfaction rate with the app. Based on the TAM, perceived usefulness and perceived ease of use are the key factors in the process of adopting new technology. A cross-sectional study by Canadian medical practitioners found that perceived ease of use was the strongest facilitator for electronic health record use, whereas usefulness and ease of use were the main factors influencing system acceptance among nonusers [25]. A systematic review also stated that lower perceived ease of use may lead to resistance to further acceptance and require additional effort and time [26]. In our study, respondents who faced difficulties in the downloading and acceptance processes had reduced willingness to download and use the app.

Privacy was an important perceived barrier to the acceptability of the eHealth app. The respondents in group 3 were worried about their personal information and privacy. As supported by international studies, privacy was a common concern raised by the public about eHealth technologies [27], especially when patients' lifestyles and activities were collected by multiple mobile health apps [28]. By contrast, our result showed that a significantly lower percentage of the users expressed concern about privacy, and that they had a generally high satisfaction rate with the app. Those who have already used the app valued their experience and benefits outweighed the privacy issue. This result was also suggested by a previous study on perceived benefits and concerns toward health information exchange [29]. Data security was also found to be a major barrier for non-enrollees not registering for the eHRSS in our 2020 study [7].

Implication

More assistance and support should be provided regarding the perceived difficulties in using mobile apps. To enhance the acceptance rate among people who have not adopted or downloaded the app, the utility and benefit of the app should be emphasized among the public. We suggest further promoting the app through doctors by sharing the benefits of health management in using the app with the citizens. For future development, more sharable scope of the health record, such as laboratory results and radiographic images, and customized health information, including age-specific health care recommendations and tailored health tips, should be included.

Regarding the perception of difficulties in using mobile apps, the user interface and user experience should be further enhanced. The acceptance of the eHealth app requires a certain level of technology literacy and a fair understanding of digital technology [30]. To have a full experience of eHealth, users are required to develop fundamental skills in health, information, science, media, computer, and the internet [31]. The publicity channels could be used to educate and provide some quick tips to the citizens. Users should also be encouraged to manage their family members, who are less familiar with the mobile apps, via the eHealth app.

Regarding the privacy issue, the security and privacy measures applied to the eHealth app should be reinforced. Further, it is an effective way to ensure widespread participation in the eHealth app by emphasizing the utility and benefits of the app [29,32]. The strategy is to present positively framed messages to the participants [33]. The usefulness and convenience of the eHealth app should be emphasized as they were strong predictive factors of acceptance of the eHealth app. A high percentage of respondents agreed that using the app could enhance their experience of health services, their concerns about health information, their management of health, and improve the health of family members.

In our findings, doctors had an important role in determining people's acceptance of the app. Doctors could recommend citizens managing the eHealth account and sharing function, which were the top-rated perceived enablers. The app could also improve the workflow of the doctors by allowing them to access patients' health records that have been shared in the eHRSS. Doctor was an important source for citizens to acknowledge the eHealth app. Therefore, it was also important to introduce the eHealth app to doctors and health care providers, encourage them to manage patients' health, and facilitate comanagement by patients and their family with the assistance of the eHealth app.

For future improvement, personalized and age-specific health care recommendations should be provided to facilitate a more patient-centered eHealth app [34]. Health information, health care recommendation, and support could be individualized to the patients. Tailored health information was processed and selected by human or computer algorithms from a database developed for the citizens. The self-monitoring health data recorded in the app by the citizens are also one of the sources for the database. With more self-input health data in the app (eg, BMI, health vital, and medication list), the data collected could be used to provide tailor-made health tips. Tailored health messages or recommendations could thus be individualized according to the patients' needs that were able to command greater attention and were easier to be understood [35]. Health information could be specific to the age and chronic diseases or other personal backgrounds of the citizen, which could improve the design of the app.

Conclusions

Overall, the respondents had a high satisfaction rate and a positive attitude toward continuing to adopt and recommend the app. The eHealth app seemed to empower citizens and their family members by enhancing their health information, self-management strategies, and experience with health services. However, privacy concerns and perceived difficulties in adopting were the major challenges of promoting eHealth. More comprehensive health records and tailored health information were recommended to be included for future improvement.

Acknowledgments

The study was funded by Hospital Authority of Hong Kong Special Administrative Region (Ref.: (20) in HA 7052118).

Data Availability

The data used for the analyses are available upon reasonable request from the corresponding author.

Authors' Contributions

JH and MCSW participated in the conception of the research ideas, study design, interpretation of the findings, and provided intellectual input to the translational aspects of the study. JH, WSP, YYW, and FYM contributed to the implementation of the study, statistical analysis, and writing of the first draft of the manuscript. MCSW, FSWC, CSKC, WNW, and NTC made critical revisions on the manuscripts and provided expert opinions on implications of the study findings.

Conflicts of Interest

FSWC, CSKC, WMW, and NTC are from the funding authority.

Multimedia Appendix 1
Factors predictive of downloading and adoption of the eHealth app by the Technology Acceptance Model (TAM).
[DOCX File , 13 KB - jmir_v24i12e40370_app1.docx]

Multimedia Appendix 2
Factors predictive of downloading and adoption of the eHealth app by the Technology Acceptance Model (TAM).
[DOCX File , 23 KB - jmir_v24i12e40370_app2.docx]

Multimedia Appendix 3
Factors predictive of downloading and acceptance of the eHealth app by the Theory of Planned Behavior (TPB).
[DOCX File , 22 KB - jmir_v24i12e40370_app3.docx]

Multimedia Appendix 4
Survey administered for the different respondents in this study.
[DOCX File , 7119 KB - jmir_v24i12e40370_app4.docx]

Multimedia Appendix 5
Distribution of study participants according to downloading and acceptance of the eHealth app.

[DOCX File , 14 KB - jmir_v24i12e40370_app5.docx]

Multimedia Appendix 6
Sources where the study participants learnt about the eHealth app.
[DOCX File , 15 KB - jmir_v24i12e40370_app6.docx]

Multimedia Appendix 7
Perception on processes of current and future acceptance of the eHealth app.
[DOCX File , 17 KB - jmir_v24i12e40370_app7.docx]

Multimedia Appendix 8
Applicability of the eHealth app.
[DOCX File , 17 KB - jmir_v24i12e40370_app8.docx]

Multimedia Appendix 9
Perception of the eHealth app.
[DOCX File , 15 KB - jmir_v24i12e40370_app9.docx]

Multimedia Appendix 10
Expectations on future development of the eHealth app.
[DOCX File , 17 KB - jmir_v24i12e40370_app10.docx]

References

1. Food and Health Bureau. Your Health, Your Life: Healthcare Reform Consultation Document. Food and Health Bureau. Hong Kong: Food and Health Bureau; 2008. URL: https://www.healthbureau.gov.hk/beStrong/files/consultation/exsummary_eng.pdf [accessed 2022-11-18]
2. eHRSS: The New World of Healthcare. Hong Kong Special Administrative Region Government. 2016 Mar 9. URL: https://www.ehealth.gov.hk/en/whats-new/ehealth-news/ehealth_news_09/the_new_world_of_healthcare.html [accessed 2022-11-18]
3. Wong ACK, Shum LHY, Chan FKK, Cheung CSK, Tong ELH, Cheung NT. Easy access to Electronic Health Record Sharing System (eHRSS) via Clinical Management System (CMS) Extension On-ramp system by Hong Kong healthcare providers in private sectors. 2016 Presented at: The Asia Pacific Electronic Health Records Conference; July 6 and 7, 2016; Science Park, Hong Kong.
4. Electronic Health Record Sharing System wins Asia Pacific Information and Communications Technology Alliance Award (with photo). In: The Government of the Hong Kong Special Administrative Region. Hong Kong: The Government of the Hong Kong Special Administrative Region; Dec 11, 2017.
5. Electronic Health Record Sharing System records millionth patient registrant at third anniversary. In: The Government of the Hong Kong Special Administrative Region. Hong Kong: The Government of the Hong Kong Special Administrative Region; Mar 5, 2019.
6. Hospital Authority. eHealth App. Hospital Authority. 2021. URL: https://www.ehealth.gov.hk/en/you-and-your-family/mobile-app.html [accessed 2022-11-18]
7. Wong MC, Huang J, Chan PS, Lok V, Leung C, Wang J, et al. The Perceptions of and Factors Associated With the Adoption of the Electronic Health Record Sharing System Among Patients and Physicians: Cross-Sectional Survey. JMIR Med Inform 2020 May 21;8(5):e17452 [FREE Full text] [doi: 10.2196/17452] [Medline: 32436855]
8. Pires IM, Marques G, Garcia NM, Flórez-Revuelta F, Ponciano V, Oniani S. A Research on the Classification and Applicability of the Mobile Health Applications. J Pers Med 2020 Feb 27;10(1):11 [FREE Full text] [doi: 10.3390/jpm10010011] [Medline: 32120849]
9. Choi W, Chun J, Lee S, Chang S, Kim D, Choi IY. Development of a MyData Platform Based on the Personal Health Record Data Sharing System in Korea. Applied Sciences 2021 Sep 03;11(17):8208. [doi: 10.3390/app11178208]
10. Department of Business Innovation and Skills, Midata Innovation Lab. The Midata Innovation Opportunity. Midata Innovation Lab. 2013. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/262271/bis-13-1314-the_midata-innovation-opportunity-v2.pdf [accessed 2022-09-23]
11. Choi Y, Kim J, Kwon IH, Kim T, Kim SM, Cha W, et al. Development of a Mobile Personal Health Record Application Designed for Emergency Care in Korea; Integrated Information from Multicenter Electronic Medical Records. Applied Sciences 2020 Sep 25;10(19):6711. [doi: 10.3390/app10196711]
12. Jung J. South Korea's My HealthWay: A "digital highway" of personal health records, but to where? thebmjopinion. 2021 Sep 21. URL: https://blogs.bmj.com/bmj/2021/09/24/south-koreas-my-healthway-a-digital-highway-of-personal-health-records-but-to-where/ [accessed 2022-09-23]

13. eHealth. Hong Kong Special Administrative Region Government. Hong Kong: Hong Kong Special Administrative Region Government; 2022. URL: https://www.portal.ehealth.gov.hk/app/index.html# [accessed 2022-11-18]
14. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
15. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
16. Ajzen I. From Intentions to Actions: A Theory of Planned Behavior. In: Kuhl J, Beckmann J, editors. Action Control: From Cognition to Behavior. Heidelberg, Germany: Springer; 1985:11-39.
17. Woon IM, Kankanhalli A. Investigation of IS professionals' intention to practise secure development of applications. International Journal of Human-Computer Studies 2007 Jan;65(1):29-41. [doi: 10.1016/j.ijhcs.2006.08.003]
18. Taylor S, Todd PA. Understanding Information Technology Usage: A Test of Competing Models. Information Systems Research 1995 Jun;6(2):144-176. [doi: 10.1287/isre.6.2.144]
19. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: Future directions for HIT implementation research. J Biomed Inform 2019 Dec;100:103315 [FREE Full text] [doi: 10.1016/j.jbi.2019.103315] [Medline: 31629923]
20. 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]
21. Wang J, Huang J, Cheung CSK, Wong WN, Cheung NT, Wong MC. Adoption of an Electronic Patient Record Sharing Pilot Project: Cross-Sectional Survey. J Med Internet Res 2020 Apr 06;22(4):e13761 [FREE Full text] [doi: 10.2196/13761] [Medline: 32250279]
22. Nugteren IC, Duineveld LAM, Wieldraaijer T, van Weert HCPM, Verdonck-de Leeuw IM, van Uden-Kraan CF, et al. Need for general practitioner involvement and eHealth in colon cancer survivorship care: patients' perspectives. Fam Pract 2017 Aug 01;34(4):473-478. [doi: 10.1093/fampra/cmw139] [Medline: 28122847]
23. Tebeje TH, Klein J. Applications of e-Health to Support Person-Centered Health Care at the Time of COVID-19 Pandemic. Telemed J E Health 2021 Feb 01;27(2):150-158. [doi: 10.1089/tmj.2020.0201] [Medline: 32746750]
24. Osmanlliu E, Rafie E, Bédard S, Paquette J, Gore G, Pomey M. Considerations for the Design and Implementation of COVID-19 Contact Tracing Apps: Scoping Review. JMIR Mhealth Uhealth 2021 Jun 09;9(6):e27102 [FREE Full text] [doi: 10.2196/27102] [Medline: 34038376]
25. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating Patient Empowerment in Association With eHealth Technology: Scoping Review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
26. 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]
27. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth Technologies: Interests, Preferences, and Concerns of Older Adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
28. Beheshti-Atashgah M, Aref MR, Barari M, Bayat M. Security and privacy-preserving in e-health: A new framework for patient. Internet of Things 2020 Dec;12:100290. [doi: 10.1016/j.iot.2020.100290]
29. Dimitropoulos L, Patel V, Scheffler S, Posnack S. Public attitudes toward health information exchange: perceived benefits and concerns. The American journal of managed care 2011;17(12 Spec No):111-116 [FREE Full text] [Medline: 22216769]
30. van der Kleij RM, Kasteleyn MJ, Meijer E, Bonten TN, Houwink EJ, Teichert M, et al. SERIES: eHealth in primary care. Part 1: Concepts, conditions and challenges. Eur J Gen Pract 2019 Oct 10;25(4):179-189 [FREE Full text] [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
31. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
32. Angst, Agarwal. Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. MIS Quarterly 2009;33(2):339. [doi: 10.2307/20650295]
33. Electronic Health Record Sharing System. What's eHealth: Security and privacy. Government of Hong Kong. 2020 Dec 31. URL: https://www.ehealth.gov.hk/en/whats-ehealth/security-and-privacy.html [accessed 2022-11-18]
34. Alpay L, van der Boog P, Dumaij A. An empowerment-based approach to developing innovative e-health tools for self-management. Health Informatics J 2011 Dec 21;17(4):247-255 [FREE Full text] [doi: 10.1177/1460458211420089] [Medline: 22193825]
35. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019 Feb 05;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]

Abbreviations

eHRSS: Electronic Health Record Sharing System
HA: Hospital Authority

TAM: Technology Acceptance Model
TPB: Theory of Planned Behavior
TRA: Theory of Reasoned Action

Edited by G Eysenbach; submitted 17.06.22; peer-reviewed by A Garcia, M Alyahya, A Shachak; comments to author 20.09.22; revised version received 05.10.22; accepted 15.11.22; published 12.12.22.

Please cite as:
Huang J, Pang WS, Wong YY, Mak FY, Chan FSW, Cheung CSK, Wong WN, Cheung NT, Wong MCS
Factors Associated With the Acceptance of an eHealth App for Electronic Health Record Sharing System: Population-Based Study
J Med Internet Res 2022;24(12):e40370
URL: https://www.jmir.org/2022/12/e40370
doi:10.2196/40370
PMID:36382349

Original Paper

The Agreement Between Virtual Patient and Unannounced Standardized Patient Assessments in Evaluating Primary Health Care Quality: Multicenter, Cross-sectional Pilot Study in 7 Provinces of China

Minrui Zeng[1], BM; Yiyuan Cai[2], PhD; Jin Cao[3], MPH; Qianyu He[1], MSN; Xiaohui Wang[3], PhD; Yun Lu[4], MM; Huijuan Liang[5], PhD; Dong Xu[6,7], PhD; Jing Liao[1], PhD

[1]Department of Medical Statistics & Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Department of Epidemiology and Medical Statistics, School of Public Health, Guizhou Medical University, Guiyang, China

[3]School of Public Health, Zhejiang University School of Medicine, Hangzhou, China

[4]Department of Preventive Medicine & Maternal and Child Health, School of Public Health, Guizhou Medical University, Guiyang, China

[5]Research Institute for Health Policy of Inner Mongolia, Inner Mongolia Medical University, Hohhot, China

[6]Center for World Health Organization Studies, Department of Health Management, School of Health Management of Southern Medical University, Guangzhou, China

[7]ACACIA Lab for Implementation Research, Southern Medical University Institute for Global Health, Dermatology Hospital of Southern Medical University, Guangzhou, China

Corresponding Author:
Jing Liao, PhD
Department of Medical Statistics & Epidemiology
School of Public Health
Sun Yat-sen University
North Campus
No 74, 2nd Yat-Sen Road, Yuexiu District
Guangzhou, 510080
China
Phone: 86 13560339049
Email: liaojing5@sysu.edu.cn

Abstract

Background: The unannounced standardized patient (USP) is the gold standard for primary health care (PHC) quality assessment but has many restrictions associated with high human and resource costs. Virtual patient (VP) is a valid, low-cost software option for simulating clinical scenarios and is widely used in medical education. It is unclear whether VP can be used to assess the quality of PHC.

Objective: This study aimed to examine the agreement between VP and USP assessments of PHC quality and to identify factors influencing the VP-USP agreement.

Methods: Eleven matched VP and USP case designs were developed based on clinical guidelines and were implemented in a convenience sample of urban PHC facilities in the capital cities of the 7 study provinces. A total of 720 USP visits were conducted, during which on-duty PHC providers who met the inclusion criteria were randomly selected by the USPs. The same providers underwent a VP assessment using the same case condition at least a week later. The VP-USP agreement was measured by the concordance correlation coefficient (CCC) for continuity scores and the weighted κ for diagnoses. Multiple linear regression was used to identify factors influencing the VP-USP agreement.

Results: Only 146 VP scores were matched with the corresponding USP scores. The CCC for medical history was 0.37 (95% CI 0.24-0.49); for physical examination, 0.27 (95% CI 0.12-0.42); for laboratory and imaging tests, –0.03 (95% CI –0.20 to 0.14); and for treatment, 0.22 (95% CI 0.07-0.37). The weighted κ for diagnosis was 0.32 (95% CI 0.13-0.52). The multiple linear regression model indicated that the VP tests were significantly influenced by the different case conditions and the city where the test took place.

Conclusions: There was low agreement between VPs and USPs in PHC quality assessment. This may reflect the "know-do" gap. VP test results were also influenced by different case conditions, interactive design, and usability. Modifications to VPs and the reasons for the low VP-USP agreement require further study.

(J Med Internet Res 2022;24(12):e40082) doi:10.2196/40082

KEYWORDS

virtual patient; unannounced standardized patient; primary health care; primary care; quality assessment; quality improvement; scenario; simulation; simulate; medical education; cross-sectional; digital health; eHealth

Introduction

Improving primary health care (PHC) services is one approach to increasing universal health coverage [1]. PHC provides comprehensive essential health care to the community by supporting access to health monitoring, diagnosis, and treatment [2] in an efficient and cost-effective manner [3,4]. PHC service quality is an important factor affecting population health outcomes and should be strengthened as part of health care system reforms [5,6] and in the face of the drastic challenges of the COVID-19 pandemic.

The unannounced standardized patient (USP) is regarded as the gold standard to assess the quality of PHC services [7-10]. USP is a rigorously trained actors portraying patients with certain diseases who anonymously visit PHC services; they can provide a standardized and timely evaluation of health care providers' performance that prevents the Hawthorne effect, that is, changes in practice associated with being observed [11]. However, USP is limited to clinical conditions that have no obvious signs and that do not require invasive examinations [12]; they are also difficult to deploy in low- and middle-income countries due to their heavy reliance on personnel and resources. The virtual patient (VP), an improvement on computerized clinical vignettes [13], has been proposed as a potential low-cost alternative to USP. VP is a software tool; they simulate real clinical scenarios and have been widely used in medical education [14] due to their low requirements for equipment, high interactivity, safety, and capacity for repeatable actions [15].

It is unknown whether assessments of quality based on VP agree with those based on USP. Prior studies mainly applied VP in medical education [16-18] as a tool to train students' clinical thinking, skills in medical history collection and diagnosis [19], and attitudes toward patients [20]. Only a few studies have directly compared VP and standardized patients; these studies have found that skills training was less effective with VPs than with standardized patients as the educational tool [21]. No study so far has used VP for PHC quality assessment. Although VP can examine users' medical knowledge (ie, their competency) in a similar way as vignettes, the results may not accurately reflect the actual performance of users in real clinical practice [22-24], and the Hawthorne effect cannot be avoided. There is some evidence that VP user interfaces and usability may influence VP-based assessment outcomes [14,25]. The extent to which a VP may serve as a quality assessment tool needs further research [26,27].

The current study belongs to a family of studies of PHC quality assessments in China based on the multicenter, nation-wide ACACIA (Health Care Quality Cohort in China) study [28]. This was a pilot study that specifically aimed to (1) examine the agreement of VPs and USPs in assessing the quality of PHC services and (2) identify factors influencing VP-USP agreement.

Methods

Study Design and Procedure

This multicenter, cross-sectional pilot study is part of the ACACIA family of studies. The ACACIA protocol has been published previously [28,29]. Briefly, ACACIA aims to develop and validate USPs and paired VPs to assess clinical quality, cost, and patient experiences in PHC across China. The study sample's representativeness was ensured by its multistage, clustered sample design [30], stratified by the average life expectancy in each province, geographic variations, and feasibility [31]. Altogether, 7 provinces were selected, and their capital cities and prefecture-level municipalities were used as a stratum; 5 townships or urban subdistricts were selected in each city based on probability proportional to size sampling. PHC facilities were then examined in each location. For this study of the agreement of VPs and USPs, a convenience sample was selected of urban PHC facilities in the capital cities of the 7 study provinces, with USP visits to these centers conducted between 2019 and 2021. All PHC providers in these centers, including licensed practicing clinicians and unlicensed clinicians under supervision of licensed physicians, served as our study population.

Altogether, 720 USP visits were conducted. On-duty PHC providers who met our criteria were randomly selected for USP visits. The PHC providers who were visited by the USPs received a VP assessment of the same cases at least a week later to prevent the practice effect [32]. The agreement between these 2 tests was analyzed with the concordance correlation coefficient (CCC) and the weighted κ.

Ethical Approval

Ethical approval was obtained from the Ethics Committee of Sun Yat-sen University (2017-007), and all PHC providers participating in the VP tests provided informed consent.

USP and VP Case Selection and Design

The USPs and VPs shared identical case designs to ensure consistency and simplify the development process. The selection and development process for these case designs was reported previously [33]. Case designs were selected based on whether the disease in question (1) had a high frequency of PHC clinical encounters, (2) had a significant disease burden, (3) was present in the main areas of PHC in China, and (4) was feasible for use

in a USP test (ie, it was without obvious physiological signs and had a low risk of needing invasive tests). Twelve case designs were selected and rigorously developed: angina, asthma, diarrhea, cold, gastritis, hypertension, lower back pain, migraine, postpartum depression, stress urinary incontinence, tuberculosis, and type 2 diabetes. The validity of the case designs was verified, and they were found to have scale-level content validity indices over 0.90, role-playing fidelity over 90%, and checklist completion accuracy of 88% [34]. Most case designs had 5 modules: medical history, a physical examination, laboratory and imaging tests, diagnosis, and treatment. There were exceptions in 4 case designs (hypertension, lower back pain, migraine, and postpartum depression) that did not require laboratory or imaging tests. Due to the COVID-19 pandemic, tuberculosis cases were excluded to protect the USPs from unnecessary physical examinations, potential harm, and conflict [34]. Thus, only 11 case designs were used in this pilot study. Details of the case design development, modification, and validity testing are provided in Multimedia Appendix 1.

USP Training and Implementation

The USP actors all received at least one week of competency-based online-offline training and were assessed by specialists who were not members of the research team [34]. Before the site visits, the USP actors were further examined to ensure they could accurately portray the case designs according to the standardized training manual [30]. On the day of the visit, each USP was companied by a facilitator, who pretended to be a relative of the USP. The visits were audio recorded with a hidden recording device; these recordings were also used to monitor the performance of the USP actors and ensure checklist quality. If audio recordings were not available, field reports on what the providers said and did during the visits were upload to the online database of REDCap (Vanderbilt University) immediately after the visit to reduce recall bias. An example of a REDCap entry is provided in Multimedia Appendix 2.

VP Platform and Implementation

The VP was hosted on an online platform that could be accessed via a mobile phone or computer. The 5 VP modules used 3 different interface designs. For the medical history and diagnosis modules, the PHC providers were required to search for keywords with at least 2 characters to trigger relevant inquiries for selection. The physical examination module displayed all possible options. In the laboratory and imaging test module and the treatment module, some general options (eg, ordering blood tests or electrocardiograms) could be chosen directly, while specific options were made available after searching for keywords. All actions were recorded and uploaded online automatically.

For the field testing, PHC providers who agreed to participate in the VP tests received the VP install package for their mobile phone or personal computer alongside a user demonstration video. For each PHC provider, the cases for the VP test were the same as those for the USP test. The VPs were masked to avoid bias due to providers noticing the tested cases. The VP tests included a training VP case, which allowed the PHC providers to become familiarized with the operation of the system. There were no time limits for any of the VP tests to avoid underestimated results caused by a lack of proficiency. To facilitate the use of the VPs, some tests were organized on-site, which may have led to test results that differed from those completed by the PHC providers independently. Thus, the location and manner of the tests, as well as the number of VP tests assigned to each PHC provider and the age and sex of the providers, were recorded for analysis.

Outcome Measures

The F1 score, recall, and precision were used to measure the continuity of physical examinations, laboratory and imaging tests, and treatment [35]. However, precision and F1 score could not be calculated for medical history due to missing records for unnecessary consultations during USP visits. We used a method adapted from previous studies [36,37] in which recall represents the proportion of PHC providers who completed the checklist based on clinical guidelines, while precision was used to quantitatively assess unnecessary actions in clinical practice. The F1 score considered both recall and precision to be of equal importance and combined them. As shown in the following equation, the F1 score reflected both recall and precision:

In the equation and in Table 1, recall represents the proportion of necessary actions that were performed in the tests and precision represents the proportion of performed actions that were necessary.

The results of the diagnoses were classified as ordinal variables in line with clinical guidelines and were rated as completely correct, partly correct, or incorrect.

Table 1. Explanation of the relationship between test results and case design for virtual patients and unannounced standardized patients. Recall is the number of performed necessary actions divided by the number of necessary actions, while precision is the number of performed necessary actions divided by the number of performed actions.

	Performed actions	Unperformed actions
Necessary actions	Number of performed necessary actions	Number of missing necessary actions
Unnecessary actions	Number of performed unnecessary actions	N/A[a]

[a]N/A: not applicable.

Statistical Analysis

Characteristics of the PHC providers and VP test information are shown as the mean (SD) for continuous variables and percentages for categorical variables. CCC, which reflects the criterion validity of the VP tests [28], was used to analyze the agreement between precision, recall, and F1 score for VP and

USP tests. CCC values <0.90, 0.90 to 0.95, 0.95 to 0.99, and >0.99 were considered to represent poor, moderate, substantial, and almost perfect agreement, respectively [38]. The weighted κ (square weighted) was used to analyze diagnostic agreement [28]; weighted κ values <0.20, 0.20 to 0.40, 0.40 to 0.60, 0.60 to 0.80, and >0.80 were considered to represent poor, moderate, substantial, good, and almost perfect agreement, respectively [39,40].

Multiple linear regression was used to identify factors influencing VP-USP agreement. Using the VP tests as the dependent variable and USP tests as the independent variable, several multiple linear regression models were established, and the models were stepwise adjusted according to cases, characteristics of the PHC providers (ie, age, sex, and city), and test conditions (ie, test deployment and number of tests). Significant covariates in these models were controlled jointly in a fully adjusted model. Partial regression coefficients of the USP tests are reported. Statistical analysis was carried out using the R (version 4.0.5; R Foundation for Statistical Computing) packages stats (version 4.0.5), Desc Tools (version 0.99.43), and psych (version 2.1.9).

Results

Characteristics of PHC Providers and VP Test Information

The recruitment process is shown in Figure 1. Of 268 PHC providers who were visited by USPs, 80 agreed to conduct the VP sessions, yielding 236 valid VP scores. However, only 146 VP scores could be matched with the original USP scores.

The characteristics of the PHC providers included in the analysis were as follows: over 80% (67/80) were between 30 and 50 years old and most were male (48/80, 60%). About 40% (35/80) of the PHC providers worked in Guangzhou. Test deployment type for the VP tests was mainly field-testing (59/80, 74%) and more than half (42/80, 53%) of the PHC providers were tested by a single case. The average VP test time was 13.49 (SD 9.33) minutes. The most frequently tested case design was low back pain, with 25 tests. The least frequently tested cases were asthma, gastritis, migraine, and postpartum depression, with 7 tests each. Details are shown in Tables 2 and 3.

Figure 1. Flow chart of the recruitment process. USP: unannounced standardized patient; VP: virtual patient.

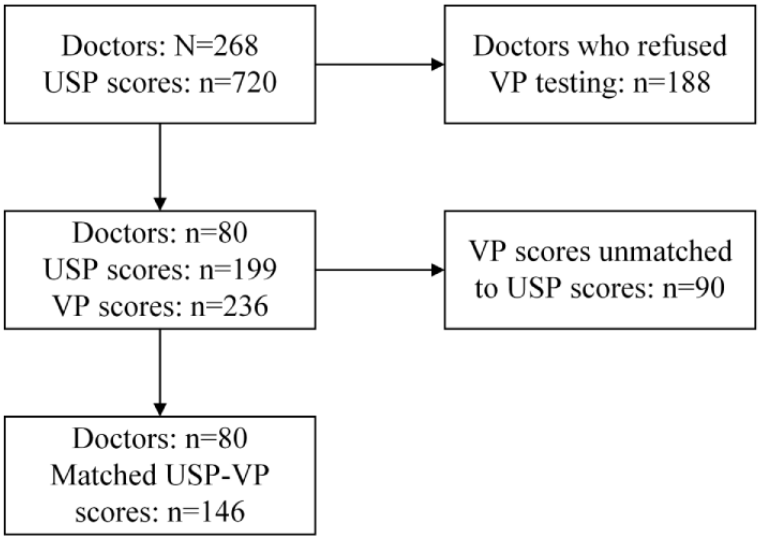


Table 2. Characteristics of primary health care providers (N=80).

Categories	Values, n (%)
Age (years)	
<30	5 (6)
30-50	67 (83)
≥50	8 (10)
Sex	
Male	48 (60)
Female	32 (40)
Location (city)	
Changsha	12 (15)
Xi'an	7 (9)
Guangzhou	35 (44)
Lanzhou	8 (10)
Hohhot	11 (14)
Guiyang	5 (6)
Chengdu	2 (3)
Case designs	
Angina	9 (6)
Asthma	7 (5)
Gastritis	7 (5)
Cold	16 (11)
Type 2 diabetes	18 (12)
Diarrhea	13 (9)
Hypertension	13 (9)
Low back pain	25 (17)
Migraine	7 (5)
Postpartum depression	7 (5)
Stress urinary incontinence	24 (16)

Table 3. Virtual patient test situations (N=80).

Categories	Values, n (%)
Test deployment	
Field-testing	59 (74)
Self-testing	21 (26)
Number of tests	
1	42 (53)
2	22 (28)
≥3	16 (20)

Agreement Between VP and USP Tests

Test outcomes and CCCs for the medical history, physical examination, laboratory and imaging tests, and treatment modules are listed in Table 4. The USP test results showed high precision (over 0.47), but the VP test results showed varying degrees of degradation (ranging from 0.25 to 0.51), which resulted in very poor agreement. Recall for medical history and treatment was similar for the USP and VP tests. It is worth noting that the physical examination module and the laboratory and imaging test module had results for the USPs that were nearly 3 times higher than for the VPs. All the CCCs for recall

were poor. The F1 score and its CCC were close to the recall values, except for the CCC for physical examination. The weighted κ for diagnosis was 0.32 (95% CI 0.13-0.51), which was unsatisfactory. Details for the weighted κ are shown in Multimedia Appendix 3.

Table 4. Test outcomes and concordance correlation coefficients. Precision and F1 score could not be calculated for unannounced standardized patients for medical history due to missing consultation records.

Test modules	Precision (SD)			Recall (SD)			F1 score (SD)		
	USPs[a]	VPs[b]	CCC[c] (95% CI)	USPs	VPs	CCC (95% CI)	USPs	VPs	CCC (95% CI)
Medical history[a]	—[d]	0.51 (0.35)	—	0.19 (0.15)	0.13 (0.13)	0.37 (0.24 to 0.49)	—	0.18 (0.16)	—
Physical examination	0.47 (0.50)	0.25 (0.30)	0.13 (0.01 to 0.26)	0.11 (0.15)	0.34 (0.31)	0.04 (–0.05 to 0.13)	0.17 (0.21)	0.20 (0.19)	0.27 (0.12 to 0.42)
Laboratory and imaging tests	0.47 (0.49)	0.45 (0.34)	0.21 (0.03 to 0.38)	0.18 (0.20)	0.57 (0.36)	–0.06 (–0.15 to 0.03)	0.25 (0.27)	0.43 (0.27)	–0.03 (–0.20 to 0.14)
Treatment	0.77 (0.41)	0.45 (0.48)	0.07 (–0.06 to 0.20)	0.21 (0.18)	0.20 (0.25)	0.24 (0.10 to 0.39)	0.31 (0.23)	0.26 (0.31)	0.22 (0.07 to 0.37)

[a]USP: unannounced standardized patient.

[b]VP: virtual patient.

[c]CCC: concordance correlation coefficient.

[d]Not available.

Factors Influencing VP-USP Agreement

To explore factors that affected VP-USP agreement, we used multiple linear regression. For medical history, there was a significant correlation between VP and USP scores that remained stable after adjustment (ranging from 0.32 to 0.34, $P<.001$). In contrast, despite factor adjustment, the correlation between the VP and USP scores was not significant ($P=.74$) for the laboratory and imaging test module. The correlation was significantly weakened after adjusting the cases for the physical examination and treatment modules. Details are listed in Multimedia Appendix 4.

Using stepwise variable selection in the fully adjusted model, all the correlations between VP and USP scores became weaker after adjustment. The partial correlation coefficients were 0.314 (95% CI 0.183-0.445) for recall for the USPs for medical history; 0.071 (95% CI –0.090 to 0.023) for F1 score for physical examination; –0.025 (95% CI –0.169 to 0.118) for F1 score for laboratory and imaging tests; and 0.045 (95% CI –0.133 to 0.223) for F1 score for treatment. Furthermore, for medical history, female sex (versus male) and Changsha and Lanzhou (versus Guangzhou) were negatively associated with recall for VPs, while test time was positively associated with recall for VPs. The F1 scores for the physical examination module and the laboratory and imaging test module were only associated with case design. The F1 score for treatment was only associated with cases and the cities where the PHC providers worked. Combining the results of these models revealed that the major influencing factors were case design and city. Details are shown in Table 5.

Table 5. The association between assessments using virtual patients and unannounced standardized patients using stepwise regression for each module.

Test modules	β (95% CI)	*P* value	Standardized β
Medical history			
Recall for USPs[a]	.314 (.183 to .445)	<.001	.351
Female sex	–.049 (–.089 to –.009)	.02	–.366
City			
Guangzhou	0 (ref)		
Changsha	–.067 (–.126 to –.009)	.03	–.506
Lanzhou	–.065 (–.129 to –.001)	.049	–.489
Test time	.002 (.001 to .004)	.006	.205
Physical examination			
F1 score for USPs	.071 (–.090 to .023)	.39	.080
Case design			
Low back pain	0 (ref)		
Cold	.203 (.100 to .306)	<.001	1.086
Gastritis	.169 (.028 to .311)	.02	.907
Laboratory and imaging			
F1 score for USPs	–.025 (–.169 to .118)	.74	–.025
Case design			
Low back pain	0 (ref)		
Cold	.206 (.074 to .337)	.003	.768
Stress urinary incontinence	–.269 (–.401 to –.137)	<.001	–1.005
Type 2 diabetes	–.289 (–.416 to –.161)	<.001	–1.079
Gastritis	–.386 (–.543 to –.228)	.003	–1.440
Treatment			
F1 score for USPs	.045 (–.133 to .223)	.62	.034
Case design			
Low back pain	0 (ref)		
Cold	.432 (.300 to .563)	<.001	1.404
Hypertension	.419 (.283 to .555)	<.001	1.363
Type 2 diabetes	.350 (.223 to .477)	<.001	1.138
Postpartum depression	.173 (.003 to .344)	.05	.564
Gastritis	–.176 (–.347 to –.006)	.05	–.573
Stress urinary incontinence	–.189 (–.301 to –.078)	.001	–.615
Migraine	–.198 (–.374 to –.023)	.03	–.645
City			
Guangzhou	0 (ref)		
Lanzhou	–.190 (–.299 to –.082)	<.001	–.619
Xi’an	–.178 (–.313 to –.042)	.01	–.577

[a]USP: unannounced standardized patient.

Discussion

Principal Findings

Our study examined the agreement between using VPs and USPs to assess the quality of PHC in China. We found that the agreement between VP and USP results was low in general, which may result from the "know-do" gap. The VP tests might also have been influenced by different case conditions, different interface designs of the VPs, and the usability of the VPs.

We found that the agreement between VP and USP scores was low in our study sample. The USP scores were low in terms of recall, indicating that our study participants performed only some of the necessary actions, especially for the physical examination module and the laboratory and imaging test module. This suggests that PHC providers only partially performed the guideline-recommended checklist items in actual practice, which might be the result of a lack of incentives or limited time and resources [5,41,42]. In contrast, the VP scores showed relatively high recall scores, with module-specific variation, compared to the USP scores. One possible explanation for the low agreement is the "know-do" gap [43-45]. The USP tests assessed how the PHC providers performed in real clinical practice with evidence-based indicators [43,46]. However, the VP testing was more likely to assess whether the providers knew how to use their knowledge, which is defined as their competency [43]. With unlimited time and resources, the PHC providers might have performed extra actions to exclude other diseases, even more than required for differential diagnosis. Thus, VP testing may be more akin to examinations in a medical training setting and therefore indicate the competence of the examinees, whereas USP testing may be more likely to assess the quality of care in actual practice [17,47]. Similar findings have been reported by previous studies of medical education, which found that examinees in VP tests tended to explore all possible information [14]. Therefore, despite our intention to use the interactive VPs as an alternative to USPs as a quality assessment tool [17], our current findings on the agreement between VP and USP test results are insufficient to provide strong evidence for such a substitution.

Further analysis using multiple linear regression suggested that VP-based performance varied with case design, indicating that PHC providers' competency differed with clinical case design. Specifically, variations for type 2 diabetes and gastritis were observed for the test modules and low scores were observed for laboratory and imaging tests for both case designs, while higher scores were seen for treatment for type 2 diabetes than for gastritis. This finding may indicate that PHC providers were more familiar with type 2 diabetes, which is commonly seen in PHC and has a distinctive medical history and physical signs. As a result, laboratory and imaging tests for type 2 diabetes were more likely to be omitted, while appropriate treatment was more likely to be conducted in the VP tests. In contrast, PHC providers might prefer to conduct simple physical examinations for the symptoms of abdominal pain, but they were reluctant to conduct the complex laboratory and imaging testing and treatment that should be offered in accordance with the clinical guidelines for gastritis [48].

Furthermore, the VP interface and usability also influenced the VP testing. By and large, 2 types of interface were used in the VP testing: searching for keywords for consultation for medical history and selecting from multiple choices for physical examinations (for other modules, a mixture of both interface formats was adopted). Specifically, our results showed that the recall score for medical history for the VP testing was two-thirds of the score for USP testing, while the corresponding VP score for physical examination was more than twice that of the USP score. These findings indicate that multiple choices might provide more hints, allowing PHC providers to guess a correct action more easily [14,26]. Although searching for keywords for consultation or actions leads to less bias than the hints provided by multiple choices, this interface might decrease the usability of VPs, particularly when the interaction does not provide enough options or fuzzy searches. Many of the PHC providers in the study found this interaction was not user-friendly and that the consultation questions needed were often not retrievable. Frustrated by the poor interface, they tended to suddenly end the consultation and even drop out of the VP testing entirely. Although searching for keywords decreased usability, this interface should be used for the purpose of quality assessment, albeit with modifications. The influence of the study city may reflect differences in the attitude and capacity for digital adaptation of the PHC providers; those from developed regions with wider use of digital information systems may be more receptive to digitalized medical practice than their counterparts from less-developed regions [41,49]. VP usability may also be influenced by the digital adaptation attitude of PHC providers. Previous studies have found that examinees who are more open to digital innovation, better educated, and younger are more enthusiastic about using and completing digital device–based programs [50,51] like VP testing. Due to missing information on sociodemographic characteristics, our study could not examine the statistical significance of variations in these characteristics other than the city where the participants were located. However, we did find that the agreement between VP and USP results was lower in Changsha, Lanzhou, and Xi'an.

Study Limitations

The study had several limitations. First, as a purposive sampling approach was used, our research sample may be more likely to have included PHC providers who were receptive to technological innovations, and the extent to which our findings apply to providers who are less receptive needs verification. Although our user experience analysis showed promising results, only a few participants answered the user experience questionnaire. Second, due to substantial missing data for the sociodemographic characteristics of the PHC providers, we failed to identify any remaining influence of these factors on the agreement between VP and USP scores. Third, although we found that the VP interface was a key factor influencing the VP testing, we did not perform a direct comparison of different interfaces with the same disease module. Last, we used a summary score for each module to indicate the providers' performance, assuming that individual consultation or action items had equal importance. Nevertheless, a hierarchic order may exist among consultation and action items that is specific

to the disease conditions under consideration, such that a weighted score may have been better suited for quantifying the providers' performance [52].

Study Implications for Further Studies

Our findings highlight the need for further modifications to the VP platform. To improve the design of the VPs to bring them as close as possible to real clinical conditions, strict testing time limits should be implemented to enhance the sense of time pressure. Besides this, the interactive design of the VPs should opt for keyword searching over multiple choices to minimize hints. The creation of clinical settings and the application of keyword searching can be enhanced via advanced technologies, such as virtual simulation, voice input, and fuzzy retrieval [12,53], for a better, user-centered experience [54].

Moreover, to facilitate the implementation of VPs, an add-on program to widely used social software such as WeChat would be preferable to a separate application that requires installation. A short demonstration (of less than 5 minutes) of the main action steps of the VP should be embedded in the program and shown as a mandatory preview for first-time users. If needed, initial training with VP cases should also be provided, so that users can become familiar with the platform.

To better understand the agreement between VP and USP testing, future studies would benefit from systematically collecting information on potential factors contributing to differences between VPs and USPs, using both quantitative and qualitative approaches. For instance, the preferences of PHC providers for different VP interfaces could be examined with questionnaires or interviews to assess differences in perceived authenticity, cognitive load, and motivation [19]. Moreover, potential reasons for the adherence of PHC providers to clinical guidelines and associated influencing factors need to be further explored using mixed methods based on the Theoretical Domains Framework, structured questionnaires, or focus groups [55-57].

Conclusion

The agreement between VP and USP testing for PHC quality assessment was low. This low agreement may mainly reflect the "know-do" gap, while the VP test results were also influenced by different case conditions, interface design, and usability. To improve VP usability in the resource-limited settings found in PHC, VPs should be modified to be more user-centered, paying attention to the balance between enhancing usability and avoiding hints. Factors influencing the agreement between VP and USP testing need further study.

Acknowledgments

The project received support from the China Medical Board 18-300, National Natural Science Foundation of China (71974211), the Swiss Agency for Development and Cooperation (81067392), and the National Key R&D Program of China (2021ZD0113401). The authors thank the unannounced standardized patient team for the case designs and field tests and JL for proofreading the manuscript. The funders of this study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability

Due to data safety and property rights, the data sets generated and analyzed during the current study are not publicly accessible but are available upon request to the corresponding author.

Authors' Contributions

JL, YC, HL, YL, XW, and DX co-designed the study. JL, DX, and YC requested the funding. JL, YC, and DX take responsibility for project administration. All authors conducted or managed the research and investigation process. MZ, YC, and JL took part in virtual patient development. MZ, YC, and JL designed the methodology. MZ, JL, YC, and JC carried out or supported the data analysis, including the statistical analyses. YC, MZ, JL, and JC take responsibility for the integrity of the data and the accuracy of the data analysis, with each author having been responsible for specific parts of the raw data set. MZ wrote the manuscript. QH and JL reviewed the manuscript. All authors critically reviewed the manuscript and contributed to the interpretation of the results. All authors had access to the data, critically reviewed the paper before publication, and take final responsibility for the decision to submit the research for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Case development, modification, and validity.
[DOCX File , 20 KB - jmir_v24i12e40082_app1.docx]

Multimedia Appendix 2
Example of REDCap.

[DOCX File , 1544 KB - jmir_v24i12e40082_app2.docx]

Multimedia Appendix 3
Classification of diagnosis.
[DOCX File , 18 KB - jmir_v24i12e40082_app3.docx]

Multimedia Appendix 4
Regression result of different adjustment.
[DOCX File , 24 KB - jmir_v24i12e40082_app4.docx]

References

1. Binagwaho A, Adhanom Ghebreyesus T. Primary healthcare is cornerstone of universal health coverage. BMJ 2019 Jun 03;365:l2391. [doi: 10.1136/bmj.l2391] [Medline: 31160322]
2. Chan M. Ten years in public health, 2007–2017: report by Dr Margaret Chan, Director-General, World Health Organization. 2007. URL: https://apps.who.int/iris/bitstream/handle/10665/255355/9789241512442-eng.pdf?sequence=1&isAllowed=y [accessed 2022-11-14]
3. So CF, Chung JW. Telehealth for diabetes self-management in primary healthcare: A systematic review and meta-analysis. J Telemed Telecare 2018 Jun;24(5):356-364. [doi: 10.1177/1357633X17700552] [Medline: 28463033]
4. Kieny M, Evans DB. Universal health coverage. East Mediterr Health J 2013 Apr;19(4):305-306 [FREE Full text] [Medline: 23882953]
5. Sylvia S, Shi Y, Xue H, Tian X, Wang H, Liu Q, et al. Survey using incognito standardized patients shows poor quality care in China's rural clinics. Health Policy Plan 2015 Apr;30(3):322-333. [doi: 10.1093/heapol/czu014] [Medline: 24653216]
6. Hu R, Liao Y, Du Z, Hao Y, Liang H, Shi L. Types of health care facilities and the quality of primary care: a study of characteristics and experiences of Chinese patients in Guangdong Province, China. BMC Health Serv Res 2016 Aug 02;16(a):335 [FREE Full text] [doi: 10.1186/s12913-016-1604-2] [Medline: 27484465]
7. Rethans J, Gorter S, Bokken L, Morrison L. Unannounced standardised patients in real practice: a systematic literature review. Med Educ 2007 Jun;41(6):537-549. [doi: 10.1111/j.1365-2929.2006.02689.x] [Medline: 17518833]
8. Daniels B, Kwan A, Satyanarayana S, Subbaraman R, Das RK, Das V, et al. Use of standardised patients to assess gender differences in quality of tuberculosis care in urban India: a two-city, cross-sectional study. Lancet Glob Health 2019 May;7(5):e633-e643 [FREE Full text] [doi: 10.1016/S2214-109X(19)30031-2] [Medline: 30928341]
9. Wu Y, Zhou H, Ma X, Shi Y, Xue H, Zhou C, et al. Using standardised patients to assess the quality of medical records: an application and evidence from rural China. BMJ Qual Saf 2020 Jun;29(6):491-498 [FREE Full text] [doi: 10.1136/bmjqs-2019-009890] [Medline: 31776199]
10. Burnett AM, Lee L, McGuinness M, Varga B, Perez Hazel Y, Ho SM. Quality of refractive error care (Q.REC) in Cambodia, Malaysia and Pakistan: protocol for a cross-sectional unannounced standardised patient study. BMJ Open 2022 Mar 14;12(3):e057594 [FREE Full text] [doi: 10.1136/bmjopen-2021-057594] [Medline: 35288395]
11. Shah R, Edgar D, Evans BJW. Measuring clinical practice. Ophthalmic Physiol Opt 2007 Mar;27(2):113-125. [doi: 10.1111/j.1475-1313.2006.00481.x] [Medline: 17324200]
12. Triola M, Feldman H, Kalet AL, Zabar S, Kachur EK, Gillespie C, et al. A randomized trial of teaching clinical skills using virtual and live standardized patients. J Gen Intern Med 2006 May;21(5):424-429 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00421.x] [Medline: 16704382]
13. 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]
14. Isaza-Restrepo A, Gómez MT, Cifuentes G, Argüello A. The virtual patient as a learning tool: a mixed quantitative qualitative study. BMC Med Educ 2018 Dec 06;18(1):297 [FREE Full text] [doi: 10.1186/s12909-018-1395-8] [Medline: 30522478]
15. Rakofsky JJ, Talbot TB, Dunlop BW. A virtual standardized patient-based assessment tool to evaluate psychiatric residents' psychopharmacology proficiency. Acad Psychiatry 2020 Dec;44(6):693-700 [FREE Full text] [doi: 10.1007/s40596-020-01286-x] [Medline: 32681418]
16. 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]
17. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
18. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res 2019 Jul 02;21(7):e14676 [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]

19. Fink MC, Reitmeier V, Stadler M, Siebeck M, Fischer F, Fischer MR. Assessment of diagnostic competences with standardized patients versus virtual patients: experimental study in the context of history taking. J Med Internet Res 2021 Mar 04;23(3):e21196 [FREE Full text] [doi: 10.2196/21196] [Medline: 33661122]
20. Kron FW, Fetters MD, Scerbo MW, White CB, Lypson ML, Padilla MA, et al. Using a computer simulation for teaching communication skills: A blinded multisite mixed methods randomized controlled trial. Patient Educ Couns 2017 Apr;100(4):748-759 [FREE Full text] [doi: 10.1016/j.pec.2016.10.024] [Medline: 27939846]
21. Deladisma AM, Cohen M, Stevens A, Wagner P, Lok B, Bernard T, Association for Surgical Education. Do medical students respond empathetically to a virtual patient? Am J Surg 2007 Jun;193(6):756-760. [doi: 10.1016/j.amjsurg.2007.01.021] [Medline: 17512291]
22. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA 2000 Apr 05;283(13):1715-1722. [doi: 10.1001/jama.283.13.1715] [Medline: 10755498]
23. Shah R, Edgar D, Evans B. A comparison of standardised patients, record abstraction and clinical vignettes for the purpose of measuring clinical practice. Ophthalmic Physiol Opt 2010 May;30(3):209-224. [doi: 10.1111/j.1475-1313.2010.00713.x] [Medline: 20444128]
24. Dresselhaus TR, Peabody JW, Luck J, Bertenthal D. An evaluation of vignettes for predicting variation in the quality of preventive care. J Gen Intern Med 2004 Oct;19(10):1013-1018 [FREE Full text] [doi: 10.1007/s11606-004-0003-2] [Medline: 15482553]
25. Schittek Janda M, Mattheos N, Nattestad A, Wagner A, Nebel D, Färbom C, et al. Simulation of patient encounters using a virtual patient in periodontology instruction of dental students: design, usability, and learning effect in history-taking skills. Eur J Dent Educ 2004 Aug;8(3):111-119. [doi: 10.1111/j.1600-0579.2004.00339.x] [Medline: 15233775]
26. Jacklin S, Maskrey N, Chapman S. Shared decision-making with a virtual patient in medical education: mixed methods evaluation study. JMIR Med Educ 2021 Jun 10;7(2):e22745 [FREE Full text] [doi: 10.2196/22745] [Medline: 34110299]
27. Hege I, Zary N, Kononowicz AA. Criteria to assess the quality of virtual patients. Stud Health Technol Inform 2012;180:954-957. [Medline: 22874334]
28. Liao J, Chen Y, Cai Y, Zhan N, Sylvia S, Hanson K, et al. Using smartphone-based virtual patients to assess the quality of primary healthcare in rural China: protocol for a prospective multicentre study. BMJ Open 2018 Jul 11;8(7):e020943 [FREE Full text] [doi: 10.1136/bmjopen-2017-020943] [Medline: 29997138]
29. Xu D, Pan J, Dai X, Hu M, Cai Y, He H, et al. Comparing quality of primary healthcare between public and private providers in China: study protocol of a cross-sectional study using unannounced standardised patients in seven provinces of China. BMJ Open 2021 Jan 12;11(1):e040792 [FREE Full text] [doi: 10.1136/bmjopen-2020-040792] [Medline: 33436467]
30. Xu D, Hu M, He W, Liao J, Cai Y, Sylvia S, et al. Assessing the quality of primary healthcare in seven Chinese provinces with unannounced standardised patients: protocol of a cross-sectional survey. BMJ Open 2019 Feb 13;9(2):e023997 [FREE Full text] [doi: 10.1136/bmjopen-2018-023997] [Medline: 30765399]
31. Zhou M, Wang H, Zhu J, Chen W, Wang L, Liu S, et al. Cause-specific mortality for 240 causes in China during 1990-2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. Lancet 2016 Jan 16;387(10015):251-272. [doi: 10.1016/S0140-6736(15)00551-6] [Medline: 26510778]
32. McCabe D, Langer KG, Borod JC, Bender HA. Practice Effects. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer; 2011.
33. Tamblyn RM. Use of standardized patients in the assessment of medical practice. CMAJ 1998 Jan 27;158(2):205-207 [FREE Full text] [Medline: 9469142]
34. Xu DR, Cai Y, Wang X, Chen Y, Gong W, Liao J, et al. Improving data surveillance resilience beyond COVID-19: experiences of primary heAlth Care quAlity Cohort In ChinA (ACACIA) using unannounced standardized patients. Am J Public Health 2022 Jun;112(6):913-922. [doi: 10.2105/AJPH.2022.306779] [Medline: 35483014]
35. Gehanno JF, Rollin L, Le Jean T, Louvel A, Darmoni S, Shaw W. Precision and recall of search strategies for identifying studies on return-to-work in Medline. J Occup Rehabil 2009 Sep;19(3):223-230. [doi: 10.1007/s10926-009-9177-0] [Medline: 19381789]
36. Chon S, Hilgers S, Timmermann F, Dratsch T, Plum PS, Berlth F, et al. Web-based immersive patient simulator as a curricular tool for objective structured clinical examination preparation in surgery: development and evaluation. JMIR Serious Games 2018 Jul 04;6(3):e10693 [FREE Full text] [doi: 10.2196/10693] [Medline: 29973333]
37. Olgers TJ, Bij de Weg AA, Ter Maaten JC. Serious games for improving technical skills in medicine: scoping review. JMIR Serious Games 2021 Jan 25;9(1):e24093 [FREE Full text] [doi: 10.2196/24093] [Medline: 33492234]
38. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018 Sep;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
39. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
40. Jakobsson U, Westergren A. Statistical methods for assessing agreement for ordinal data. Scand J Caring Sci 2005 Dec;19(4):427-431. [doi: 10.1111/j.1471-6712.2005.00368.x] [Medline: 16324069]

41. Li X, Lu J, Hu S, Cheng KK, De Maeseneer J, Meng Q, et al. The primary health-care system in China. Lancet 2017 Dec 09;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]
42. Sylvia S, Xue H, Zhou C, Shi Y, Yi H, Zhou H, et al. Tuberculosis detection and the challenges of integrated care in rural China: A cross-sectional standardized patient study. PLoS Med 2017 Oct;14(10):e1002405 [FREE Full text] [doi: 10.1371/journal.pmed.1002405] [Medline: 29040263]
43. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]
44. Das J, Kwan A, Daniels B, Satyanarayana S, Subbaraman R, Bergkvist S, et al. Use of standardised patients to assess quality of tuberculosis care: a pilot, cross-sectional study. Lancet Infect Dis 2015 Nov;15(11):1305-1313 [FREE Full text] [doi: 10.1016/S1473-3099(15)00077-8] [Medline: 26268690]
45. Mohanan M, Vera-Hernández M, Das V, Giardili S, Goldhaber-Fiebert JD, Rabin TL, et al. The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India. JAMA Pediatr 2015 Apr 01;169(4):349-357 [FREE Full text] [doi: 10.1001/jamapediatrics.2014.3445] [Medline: 25686357]
46. Weiner SJ, Schwartz A. Directly observed care: can unannounced standardized patients address a gap in performance measurement? J Gen Intern Med 2014 Aug;29(8):1183-1187 [FREE Full text] [doi: 10.1007/s11606-014-2860-7] [Medline: 24756945]
47. Muse K, McManus F. A systematic review of methods for assessing competence in cognitive-behavioural therapy. Clin Psychol Rev 2013 Apr;33(3):484-499. [doi: 10.1016/j.cpr.2013.01.010] [Medline: 23454222]
48. Chinese Medical Association, Chinese Medical Journals Publishing House, Chinese Society of Gastroenterology, Chinese Society of General Practice, Editorial Board of Chinese Journal of General Practitioners of Chinese Medical Association, Expert Group of Guidelines for Primary Care of Gastrointestinal Disease. Guidelines for primary care in chronic gastritis (2019). Chinese Journal of General Practitioners 2020 Sep 04;19(09):768-775. [doi: 10.3760/cma.j.cn114798-20200706-00785]
49. Phichitchaisopa N, Naenna T. Factors affecting the adoption of healthcare information technology. EXCLI J 2013;12:413-436 [FREE Full text] [Medline: 26417235]
50. Thompson J, White S, Chapman S. Interactive clinical avatar use in pharmacist preregistration training: design and review. J Med Internet Res 2020 Nov 06;22(11):e17146 [FREE Full text] [doi: 10.2196/17146] [Medline: 33155983]
51. Watari T, Tokuda Y, Owada M, Onigata K. The utility of virtual patient simulations for clinical reasoning education. Int J Environ Res Public Health 2020 Jul 24;17(15):5325 [FREE Full text] [doi: 10.3390/ijerph17155325] [Medline: 32722097]
52. Ben-Assuli O, Kumar N, Arazy O, Shabtai I. The use of analytic hierarchy process for measuring the complexity of medical diagnosis. Health Informatics J 2020 Mar;26(1):218-232 [FREE Full text] [doi: 10.1177/1460458218824708] [Medline: 30672359]
53. Shorey S, Ang E, Ng ED, Yap J, Lau LST, Chui CK. Communication skills training using virtual reality: A descriptive qualitative study. Nurse Educ Today 2020 Nov;94:104592. [doi: 10.1016/j.nedt.2020.104592] [Medline: 32942248]
54. Benda NC, Kellogg KM, Hoffman DJ, Fairbanks RJ, Auguste T. Lessons learned from an evaluation of serious gaming as an alternative to mannequin-based simulation technology: randomized controlled trial. JMIR Serious Games 2020 Sep 28;8(3):e21123 [FREE Full text] [doi: 10.2196/21123] [Medline: 32985993]
55. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999 Oct 20;282(15):1458-1465. [doi: 10.1001/jama.282.15.1458] [Medline: 10535437]
56. Paksaite P, Crosskey J, Sula E, West C, Watson M. A systematic review using the Theoretical Domains Framework to identify barriers and facilitators to the adoption of prescribing guidelines. Int J Pharm Pract 2021 Feb 10;29(1):3-11. [doi: 10.1111/ijpp.12654] [Medline: 32656929]
57. Almazrou SH, Alfaifi SI, Alfaifi SH, Hakami LE, Al-Aqeel SA. Barriers to and facilitators of adherence to clinical practice guidelines in the Middle East and North Africa region: a systematic review. Healthcare (Basel) 2020 Dec 15;8(4):564 [FREE Full text] [doi: 10.3390/healthcare8040564] [Medline: 33333843]

Abbreviations

CCC: concordance correlation coefficient
PHC: primary health care
USP: unannounced standardized patient
VP: virtual patient

Edited by T Leung; submitted 05.06.22; peer-reviewed by H Zheng, M Jones; comments to author 06.09.22; revised version received 27.09.22; accepted 29.10.22; published 02.12.22.

Please cite as:
Zeng M, Cai Y, Cao J, He Q, Wang X, Lu Y, Liang H, Xu D, Liao J
The Agreement Between Virtual Patient and Unannounced Standardized Patient Assessments in Evaluating Primary Health Care Quality: Multicenter, Cross-sectional Pilot Study in 7 Provinces of China
J Med Internet Res 2022;24(12):e40082
URL: https://www.jmir.org/2022/12/e40082
doi:10.2196/40082
PMID:36459416

Original Paper

Interpretable Machine Learning Prediction of Drug-Induced QT Prolongation: Electronic Health Record Analysis

Steven T Simon[1], MD; Katy E Trinkley[2], PharmD; Daniel C Malone[3], PhD; Michael Aaron Rosenberg[4], MD

[1]Division of Cardiology, University of Colorado School of Medicine, Aurora, CO, United States

[2]Department of Clinical Pharmacy, School of Pharmacy, University of Colorado, Aurora, CO, United States

[3]College of Pharmacy, University of Utah, Salt Lake City, UT, United States

[4]Division of Cardiac Electrophysiology, University of Colorado School of Medicine, Aurora, CO, United States

Corresponding Author:
Michael Aaron Rosenberg, MD
Division of Cardiac Electrophysiology
University of Colorado School of Medicine
12631 E. 17th Ave
Aurora, CO, 80045
United States
Phone: 1 720 500 3621
Email: michael.a.rosenberg@cuanschutz.edu

Abstract

Background: Drug-induced long-QT syndrome (diLQTS) is a major concern among patients who are hospitalized, for whom prediction models capable of identifying individualized risk could be useful to guide monitoring. We have previously demonstrated the feasibility of machine learning to predict the risk of diLQTS, in which deep learning models provided superior accuracy for risk prediction, although these models were limited by a lack of interpretability.

Objective: In this investigation, we sought to examine the potential trade-off between interpretability and predictive accuracy with the use of more complex models to identify patients at risk for diLQTS. We planned to compare a deep learning algorithm to predict diLQTS with a more interpretable algorithm based on cluster analysis that would allow medication- and subpopulation-specific evaluation of risk.

Methods: We examined the risk of diLQTS among 35,639 inpatients treated between 2003 and 2018 with at least 1 of 39 medications associated with risk of diLQTS and who had an electrocardiogram in the system performed within 24 hours of medication administration. Predictors included over 22,000 diagnoses and medications at the time of medication administration, with cases of diLQTS defined as a corrected QT interval over 500 milliseconds after treatment with a culprit medication. The interpretable model was developed using cluster analysis (K=4 clusters), and risk was assessed for specific medications and classes of medications. The deep learning model was created using all predictors within a 6-layer neural network, based on previously identified hyperparameters.

Results: Among the medications, we found that class III antiarrhythmic medications were associated with increased risk across all clusters, and that in patients who are noncritically ill without cardiovascular disease, propofol was associated with increased risk, whereas ondansetron was associated with decreased risk. Compared with deep learning, the interpretable approach was less accurate (area under the receiver operating characteristic curve: 0.65 vs 0.78), with comparable calibration.

Conclusions: In summary, we found that an interpretable modeling approach was less accurate, but more clinically applicable, than deep learning for the prediction of diLQTS. Future investigations should consider this trade-off in the development of methods for clinical prediction.

(J Med Internet Res 2022;24(12):e42163) doi:10.2196/42163

KEYWORDS

drug-induced QT prolongation; predictive modeling; interpretable machine learning; ML; artificial intelligence; AI; electronic health records; EHR; prediction; risk; monitoring; deep learning

Introduction

Drug-induced long-QT syndrome (diLQTS) [1,2] is a major concern for inpatients worldwide and has been identified as a key target for clinical decision support tools [3-7]. Importantly, although certain medications have been implicated as having significant clinical risk [8,9], for others, despite a known risk of diLQTS, clinical validation has been lacking [10-12]. In the past few years, several groups have sought to apply prediction models using electronic health record (EHR) data to model risk [13-17] toward the goal of developing an automated approach that leverages innovations in data science and machine learning. In prior work [18], we performed a comparative evaluation of machine learning methods to predict diLQTS using EHR data, in which we found that the most accurate prediction method was a deep learning model (6-layer neural network). However, each of the models carried the limitation of lacking interpretability for its predictions [19], as we were unable to assess which clinical features were the most predictive. As such, we were unable to construct a meaningful decision support approach based on these models to reduce the risk of diLQTS or determine whether our model could be easily exported to other systems.

Beyond the role of increasing trust [20] in a prediction model, interpretability plays a critical role in the assessment of prediction models [21], particularly in the age of artificial intelligence, where increasingly complex models can be created using relatively raw, or unprocessed, clinical features. Limitations in interpretability are critical not only because the users may not understand why a model makes the recommendations that it does but also because a lack of interpretability increases the risk of bias in the form of data shifts [22-24]. Data shifts occur when a model is developed in one population and then applied in a different population; note that this effect could also occur within the same hospital system if the treatment paradigm changes dynamically over time. The inclusion of interpretable models also allows a detailed investigation to uncover confounding and identify situations where a critical factor was excluded from the prediction framework and to assess for reverse causality, a critical consideration in big data models. Although “interpretability” itself cannot be well quantified in the same manner as accuracy or calibration, it remains a critical consideration in the development of predictive models.

The promise of EHR data is that it provides a scale (ie, power) to draw clinical inferences across thousands of patients and potentially millions of data points, at the cost of lacking the ability for facile clinical validation. With this power comes the ability to predict clinical outcomes across a large number of heterogenous subjects, integrating the breadth of the clinical record and, with it, the range of possible diagnoses and medications that could have nonlinear associations that cannot be as easily detected using standard (ie, regression-based) methods. However, methods to leverage EHR data using machine learning have been limited by the ability to include interpretability along with predictive accuracy.

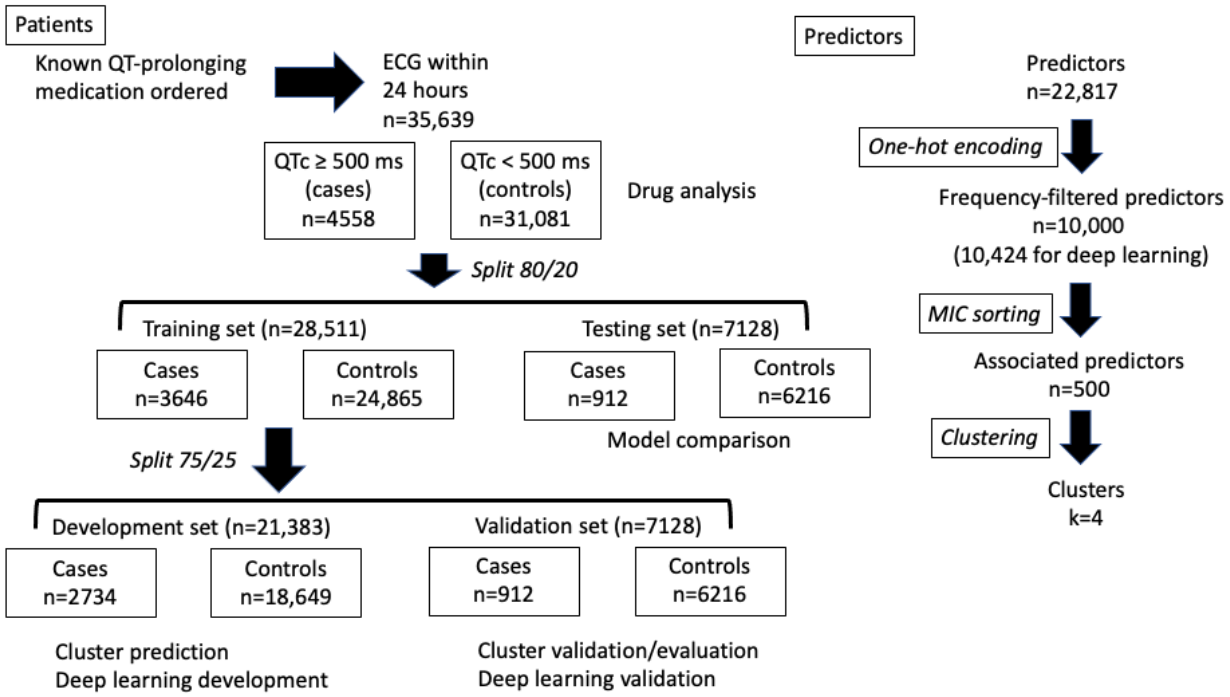
In this follow-up investigation to our previous work [18], we examined the application of an interpretable approach to predictive modeling applied at scale to EHR data to predict diLQTS. We specifically examined the use of clustering as a bridge to interpretability and compared this approach with a deep learning, noninterpretable method previously identified as providing superior predictive accuracy within our health care system.

Methods

Data Source and Study Population

The data for this investigation have been previously described [19]. Briefly, we examined EHR data from 35,639 inpatients within the UCHealth system treated between 2003 and 2018 with at least 1 of 39 medications associated with the risk of drug-induced QT prolongation and who had an electrocardiogram (ECG) in the system performed within 24 hours of medication administration (Figure 1). The primary outcome of drug-induced QT prolongation was based on any corrected QT interval over 500 milliseconds during the encounter, after the exclusion of ECGs with conduction disease (eg, bundle branch block, intraventricular conduction disease, and ventricular pacing). Predictors included any medication or diagnosis (International Classification of Diseases, Ninth or Tenth Edition) listed in the medical record that was present at the time of medication administration.

Figure 1. Data management schema. Left: patient data ascertained by order for known QT-prolonging medication with an electrocardiogram (ECG) performed within 24 hours to define cases (QTc ≥500 ms) and controls (QTc <500 ms), followed by subsequent splitting for models and validation. All splits stratified by case status. Right: processing of predictors using frequency filters, information coefficient, and clustering. MIC: maximum information coefficient; QTc: corrected QT interval.



Initial Drug Analysis

Varying formulations for each of the 39 culprit medications were combined (ie, oral and intravenous amiodarone were analyzed together). We first performed an unadjusted association analysis with each medication and the risk of diLQTS using a chi-square calculation. Those with significant associations after adjustment for multiple comparison (Bonferroni correction, *P* value for significance = .05/29 = .0017) were categorized as "high risk" for a combined analysis, as well as further model development (see below).

Predictor Filtering and Data Splitting

The medications and diagnoses in the raw data set were extracted from the EHR for each subject as a string array, following which we performed one-hot encoding (*keras.Tokenizer* [25]; version 2.8.0) to create a separate variable for each, labeled as 0 if the diagnosis or medication was absent at the time of QT-associated medication administration and 1 if it was present. As such, missing values were coded as 0, under the assumption that if the medication or diagnosis was not present in the EHR, the patient was not taking the medication or did not have that diagnosis. This process resulted in a data set containing 22,817 unique medications and diagnostic codes, from which we filtered the top 10,000 based on frequency. Of note, the 10,000th most frequent predictor was present in only 5 of 36,639 subjects. The unadjusted association for each of these 10,000 predictors with diLQTS was examined using the maximum information coefficient (MIC; *minepy.MIC*; version 1.2.6), which examines both linear and nonlinear associations based on mutual information [26]. After sorting by MIC, the top 500 most associated diagnoses and medications were selected for cluster analysis (see below). For deep learning analysis, the top 10,424 predictors after one-hot encoding were directly inputted into the model. Data splitting (Figure 1) was performed by subject index, stratified by the diagnosis of diLQTS (*sklearn.train_test_split*; version 1.1.2). The data were first split into training (28,511/35,639, 80%) and testing (7128/35,639, 20%) sets; the training set was then further split into development (21,383/28,511, 75%) and validation (7128/28,511, 25%) sets. The development set was used to fit clusters (cluster analysis) as well as to train the deep neural network. The validation set was used to examine cluster patterns and predictive accuracy, as well as to examine the training of deep learning. The testing set was used for comparative testing of cluster and deep learning models as outlined below.

Cluster Development and Evaluation

Clustering was performed using only diagnostic codes to facilitate comparisons of risk by drugs. To identify the optimal number of clusters, we first applied KMean clustering (*sklearn*; version 1.1.2) to the development set to create clusters from K=2 to K=50 and then examined inertia plot and silhouette scores (Figures S1A and S1B in Multimedia Appendix 1). After identification of K=4 as the optimal cluster number, we fitted the validation set with cluster assignments. To identify which diagnoses were the most overrepresented in each cluster (ie, which were the most different from other clusters), we calculated the proportion of each diagnosis for each cluster and assigned a value based on the product of the proportion within that cluster and the difference between this proportion and the cluster with the next highest proportion (termed the "proportion product"). The clinical interpretation of each cluster was performed by a clinician expert (MAR) after ranking the proportion product

within each cluster. Clinical interpretation included evaluating each cluster for themes of diagnoses (eg, critical care–related diagnoses and gastrointestinal-related diagnoses) to provide an overarching framework of the "types" of patients that each cluster was composed of. Clusters were examined using chi-square test for independent association the risk of diLQTS, as well as using logistic regression (unpenalized) for the proportionate risk of any high-risk medication or combinations of high-risk medications. Margin plots were created using Stata IC software (version 16; StataCorp).

Deep Learning Model Development

Hyperparameters for the deep learning model (deep neural network) were applied from our prior investigation [19]. Specifically, the deep neural network was composed of 6 layers, with 1024 neurons in the first layer and 512 neurons in the subsequent 5 layers; sigmoid activation function; 50% dropout for each layer; and batch normalization between layers. The final output was a binary prediction (the presence of diLQTS), with a binary cross-entropy loss function (*RMSprop optimizer*; learning rate=1×10^{-5}; ρ=0.9), and a validation metric of area under the receiver operating characteristic curve (AUC). The model was run over 500 planned epochs, with early stopping (*keras.callbacks.EarlyStopping*) if no improvement over 50 epochs, resulting in 118 total epochs of training. Training was monitored using learning curves (Figures S2A and S2B in Multimedia Appendix 1). The development set was used for training, and the validation set was used for validation after each epoch. In total, the deep learning model had 12,265,473 total parameters, with 12,258,305 trainable parameters and 7168 nontrainable parameters.

Model Comparison

Prediction from the cluster model was performed on the held-out testing set using logistic regression by cluster and the number of high-risk medications to obtain a predicted probability. Prediction from the deep learning model was performed through the application of the trained model to the testing set to obtain a predicted probability of diLQTS. Models were first compared using AUC, average precision score (*sklearn.metrics.average_precision_score*), and area under precision recall curve to obtain a threshold-independent comparison. The optimal probability cutoff was selected for each using the method of Youden [27]. After the selection of a cutoff, models were then compared on classification accuracy using F_1-score, recall, precision, and contingency tables. Calibration was assessed using calibration curves. Platt rescaling was performed on neural network predictions through the creation of a logistic regression model to predict actual labels.

Analysis

All analyses were conducted using Python (version 3.9.7; Python Software Foundation), run on Jupyter Notebook (Anaconda). Graphs for margin plots for cluster analysis and rescaling was performed using Stata IC software (version 16). The final script is available in Table S1 in Multimedia Appendix 1.

Ethics Approval

This project was approved by the University of Colorado Internal Review Board (COMIRB #18-0251).

Results

Initial Drug Analysis

In the initial medication evaluation, we found that amiodarone, dofetilide, fluconazole, propofol, and sotalol were significantly associated with unadjusted increased risk for diLQTS (Table 1 and Table S1 in Multimedia Appendix 1). Interestingly, medications previously highly associated with inpatient diLQTS, such as haldoperidol [5], methadone [8], citalopram [28], and azithromycin [29], were either borderline or not significantly associated with diLQTS. Additionally, it was noteworthy that ondansetron [30] was significantly associated with a decreased risk of diLQTS (P=1.12×10^{-39}).

Table 1. Association with drug-induced long-QT syndrome for selected medications. Statistically significant associations emphasized with italics.

QT-associated medication	Odds ratio (95% CI)	Chi-square (*df*)	*P* value
Dofetilide	*5.75 (4.68-7.06)*	*354.80 (4)*	*1.61×10^{-75}*
Amiodarone	*4.41 (4.0-4.87)*	*1010.70 (4)*	*1.69×10^{-217}*
Sotalol	*2.88 (2.28-3.65)*	*85.04 (4)*	*1.49×10^{-17}*
Propofol	*2.71 (2.49-2.96)*	*541.36 (4)*	*7.58×10^{-116}*
Fluconazole	*1.39 (1.21-1.59)*	*22.25 (4)*	*1.78×10^{-4}*
Methadone	1.39 (1.10-1.76)	7.45 (4)	.11
Citalopram	1.19 (1.00-1.40)	4.46 (4)	.35
Haloperidol	1.10 (1.00-1.21)	3.54 (4)	.47
Azithromycin	0.99 (0.88-1.12)	0.0085 (4)	.99
Ondansetron	*0.65 (0.61-0.69)*	*188.49 (4)*	*1.12×10^{-39}*

Association With diLQTS

Among the top 10,000 most common diagnoses and medications, the 100 with the highest MIC for association with the label of diLQTS are listed in Table S2 in Multimedia Appendix 1, with the top 500 kept for cluster analysis (minimum MIC 0.000443). The top diagnoses associated with diLQTS included long-QT syndrome, acidosis, cardiogenic shock, atrial fibrillation, and acute respiratory failure; the top medications associated included potassium chloride, furosemide, amiodarone, magnesium, and albumin (Table S2 in Multimedia Appendix 1). These results highlight the potential for possible reverse causation, as it seems more likely that potassium chloride and magnesium would be administered as treatment of or to prevent diLQTS, rather than themselves being causative. The strong association with a prior diagnosis of long-QT syndrome provides a meaningful proof of principle, as congenital long-QT syndrome is a well-known risk factor for diLQTS [1,31-34].

Cluster Analysis

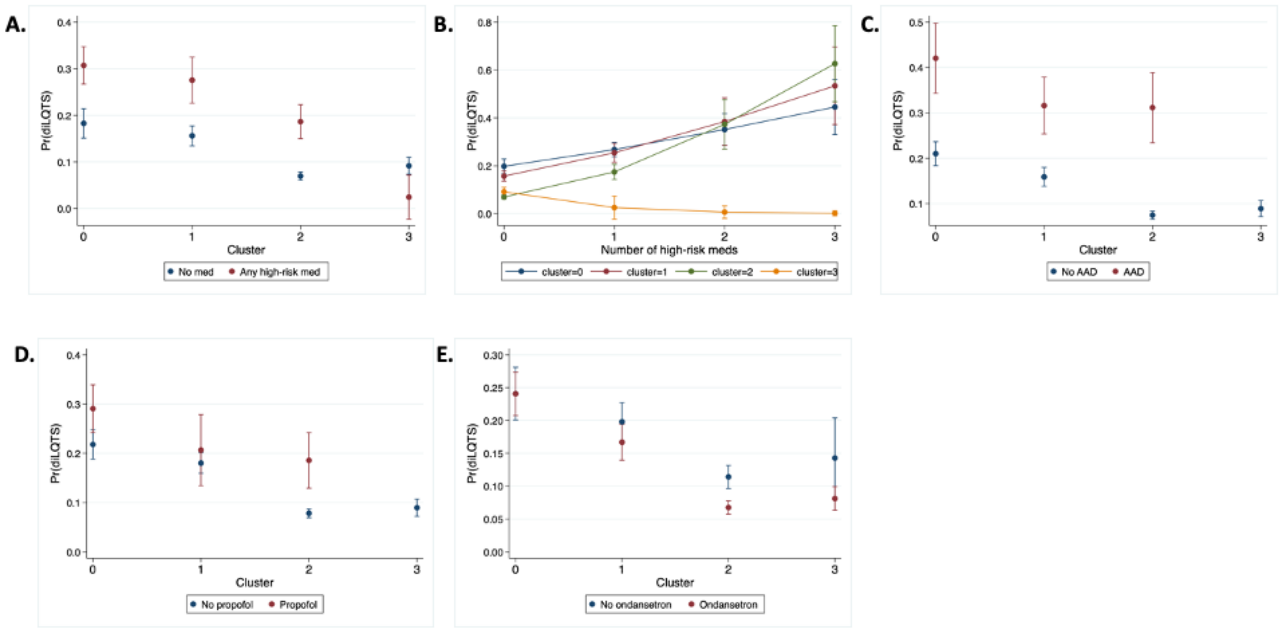
Cluster number optimization identified 4 clusters as the highest silhouette score (Figure S1A in Multimedia Appendix 1), which was validated using the elbow method applied to the inertia score (Figure S1B in Multimedia Appendix 1). Manual inspection of the cluster components (Table 2 and Table S3 in Multimedia Appendix 1) indicated that cluster 0 seemed to include a large number of critical care diagnoses; cluster 1 included diagnoses suggestive of cardiovascular disease; cluster 2 included diagnoses consistent with drug intoxication and injuries; and cluster 3 included diagnoses of nausea, abdominal pain, and headaches. In the validation set, we found that clusters 0 and 1 had an increased baseline risk of diLQTS compared with clusters 2 and 3 (Table 2), which increased with exposure to high-risk medications (Figure 2A) and combinations of high-risk medications (Figure 2B). Subjects in cluster 3 were not treated with any of the high-risk antiarrhythmic medications (amiodarone, sotalol, or dofetilide), but for all 3 other clusters, treatment with one of these agents increased the risk of diLQTS (Figure 2C). Interestingly, the use of propofol was only significantly (P=.0002) associated with risk of diLQTS for subjects in cluster 2 (Figure 2D) but not clusters 0 (P=.0161) or 1 (P=.4920; cluster 3 was not exposed), and the use of ondansetron was significantly associated with decreased risk of diLQTS in cluster 2 (P=6.371 × 10^{-6}) but not the other clusters (0: P=.996, 1: P=.129, and 3: P=.0577; Figure 2E). These results indicate that although antiarrhythmic drugs increased the risk of diLQTS broadly across all clusters, for non-antiarrhythmic medications, the impact was primarily seen in cluster 2, where propofol increased the risk of diLQTS and ondansetron decreased risk.

Table 2. Cluster composition and association with drug-induced long-QT syndrome (diLQTS). Cluster 3 represents baseline comparator group (odds ratio for the risk of diLQTS are compared with cluster 3).

Cluster	Representative diagnoses	Odds ratio (95% CI)	*P* value
0	Kidney failure, sepsis, respiratory failure, and anemia	3.25 (2.51-4.21)	<.001
1	Coronary artery disease, hypertension, hyperlipidemia, diabetes, and myocardial infarction	2.29 (1.77-2.95)	<.001
2	Live birth, motor vehicle accident, drug overdose, and alcohol intoxication	0.94 (0.73-1.20)	.61
3	Nausea, abdominal pain, and headache	1	N/A[a]

[a]N/A: not applicable.

Figure 2. Probability of diLQTS. (A) Probability of diLQTS for each cluster with treatment with high-risk medication. (B) Probability of diLQTS with increasing numbers of high-risk meds, by cluster. (C) Probability of diLQTS for each cluster with treatment with antiarrhythmic medication (AAD). (D) Probability of diLQTS for each cluster with treatment with propofol. (E) Probability of diLQTS for each cluster with treatment with ondansetron. diLQTS: drug-induced long-QT syndrome.



Comparison of Predictive Accuracy

The AUC for deep learning was 0.776 (Figure 3A) compared with the AUC of the cluster analysis of 0.636 (Figure 3B); the area under precision recall curve was 0.373 for deep learning (Figure 3C) compared with 0.322 for cluster analysis (Figure 3D); and the average precision score for deep learning was 0.379 and 0.193 for cluster analysis. Based on the Youden's method for cutoff selection, the optimal cutoff for the prediction of diLQTS from deep learning was Pr(diLQTS) of 0.12, and for cluster analysis, it was 0.15. Based on these cutoffs, the F_1-score for deep learning was 0.39, and for cluster analysis, it was 0.29. Contingency tables for both are in Tables S4A and S4B in Multimedia Appendix 1, with classification comparison in Table 3 demonstrating an agreement of 71.4% for the 2 approaches. Calibration comparison is provided in Figure 4, in which we noted that the neural network was poorly calibrated and generally overpredicted the risk of diLQTS (ie, actual proportion of diLQTS cases less than predicted probability), which had been described with these models in our previous work [18]. With Platt rescaling (Figures S3A and S3B in Multimedia Appendix 1), calibration of the neural network was improved and was similar to calibration of the cluster analysis (Figure S3B in Multimedia Appendix 1).

Figure 3. Accuracy assessment of models. (A) receiver operating characteristic (ROC) curve for neural network. (B) ROC curve for cluster model. (C) Precision-recall for neural network. (D) Precision-recall for cluster model. AUC: area under ROC curve; NN: neural network.

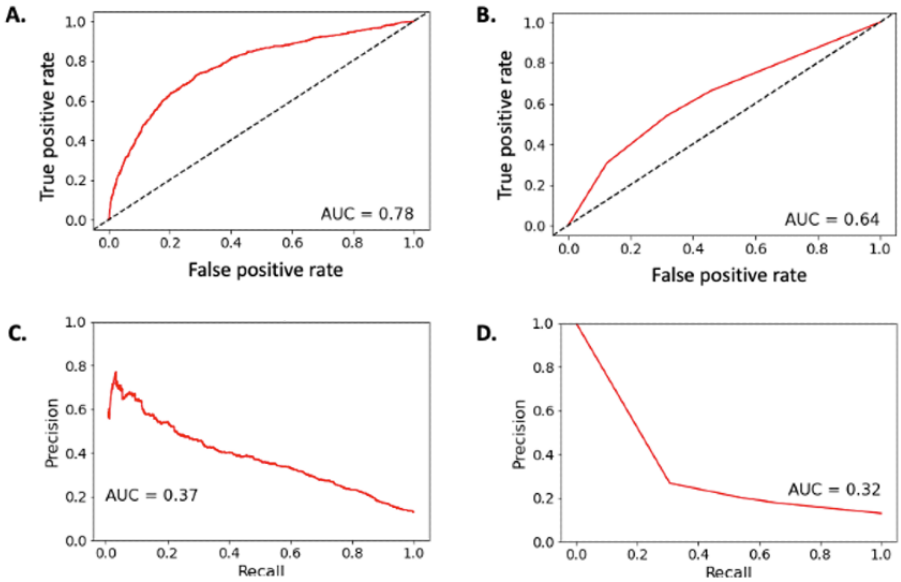
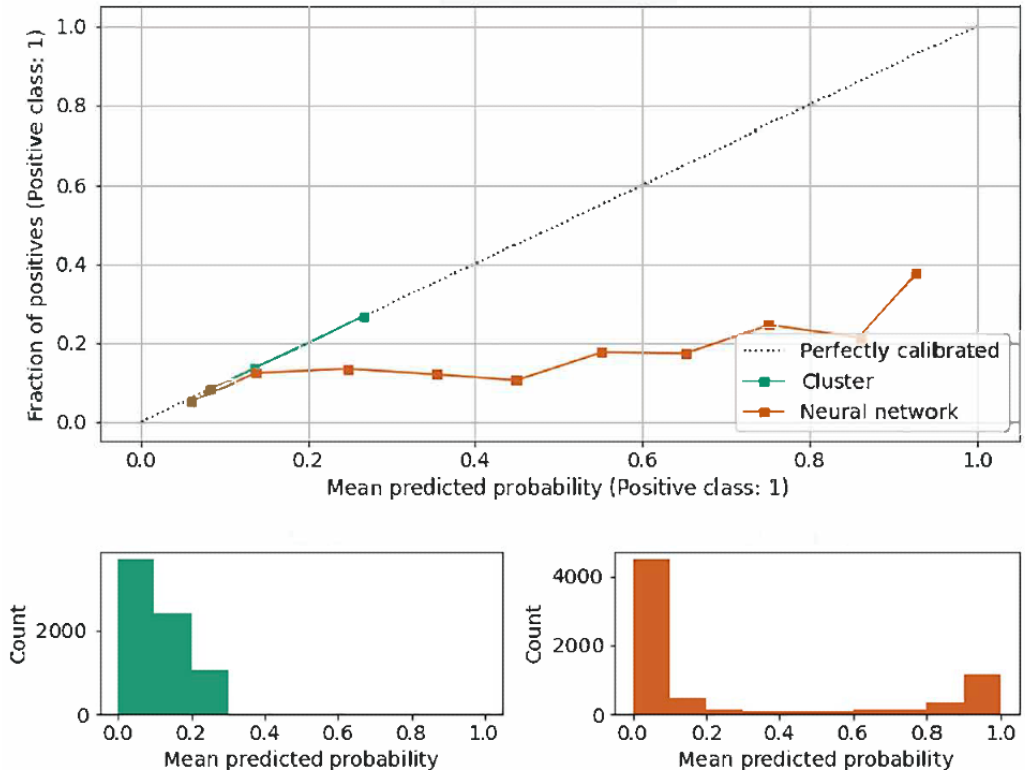


Table 3. A 2 × 2 table of comparative predictions at selected cutoffs. For deep learning models, the cutoff was probability of drug-induced long-QT syndrome (diLQTS) of 0.12, and for cluster analysis, it was 0.15. These values are based on predictive models for which the probability of diLQTS is produced for each individual, and the cutoff represents the probability above, in which an individual would be predicted to be at risk, and below, in which one would not be at risk.

	Cluster model (N=7128)		
	Predicted low risk, n (%)	Predicted high risk, n (%)	Total, n (%)
Neural network model			
Predicted low risk	3653 (51.2)	1018 (14.3)	4671 (65.6)
Predicted high risk	1017 (14.3)	1440 (20.2)	2457 (34.4)
Total	4670 (65.5)	2458 (34.5)	7128 (100)

Figure 4. Calibration analysis of neural network and cluster-based models. Top: Calibration plot for each model, with abscissa corresponding to the binned predicted probability of diLQTS (positive class) from the model and ordinate corresponding to the proportion of actual positives (diLQTS cases) within each bin. Bottom: Histogram of predicted probability for each model (left: cluster, right: neural network). Note that cluster-based model did not predict probability over 0.5 for any individual. diLQTS: drug-induced long-QT syndrome.



Discussion

Principal Findings

In this EHR-based follow-up analysis, we sought to compare 2 divergent methods for the integration of machine learning to guide clinical decisions to prevent diLQTS, with a focus on clinical interpretability and predictive accuracy. In one, we applied cluster analysis to group individuals by patterns of diagnostic codes to identify potentially recognizable clinical subgroups from which a treating clinician could identify patients who might be at risk for diLQTS to guide future decision-making. For comparison, we applied a deep learning algorithm that was identified based on prior work in this same population to obtain a "gold standard" level of predictive accuracy, to quantify the potential loss in predictive accuracy with the use of a more interpretable methodology. From a clinical perspective, our findings revealed some interesting insights regarding which specific medications have the greatest risk of diLQTS, as well as which subpopulations appear to be the most susceptible. However, we also found that there was a fairly substantial loss of predictive accuracy using this interpretable method in comparison with a "black box" method, which should be considered in future work on the integration of predictive models in clinical care.

Among the clinical insights, several are noteworthy. First, we found that when examined independent of patient characteristics, certain medications such as haldoperidol or methadone, which are well established with diLQTS, were not associated with increased risk, whereas others, such as ondansetron, were actually associated with decreased risk in our population. This finding points to the multifactorial nature of diLQTS, highlighting the need to consider other relevant contextual factors in assessing risk. However, it may also suggest that in the inpatient setting, there might be more benefit than risk with

using these medications, which is also consistent with prior studies [9-12], including one where a clinical decision support tool to prevent diLQTS had a paradoxical decrease in mortality for patients in whom the treating provider ignored the alert and prescribed the known QT-prolonging medication despite risk [4]. Particularly in subjects who were not critically ill (not in cluster 0) and without a history of cardiovascular disease (not in cluster 1), there appeared to be more benefit to using ondansetron, balanced against more risk with using propofol. However, these insights should be taken with caution, as we do not know the specific timing of the administration of QT-associated medications in relation to obtaining the ECG nor whether a medication was administered once, several times, or not at all (merely listed as an as needed pro re nata medication). Such a limitation seems likely for several of the known QT-associated medications that are frequently ordered pro re nata, such as haldoperidol and ondansetron, in which we found no (former) or an inverse (latter) association with QT-prolongation. Regardless of the underlying impact, this consideration highlights the limitations of the use of clinical decision support tools applied broadly across all medications associated with diLQTS and a need to focus on the relative population risk and indication when designing future tools.

Second, we found that, perhaps not surprisingly, the cluster of patients (cluster 0) with diagnoses suggestive of critical illness were the most susceptible to use of high-risk medications for diLQTS, and that patients in clusters 2 and 3 with more benign diagnoses were less likely to have diLQTS. This finding has direct clinical implications, as it suggests that decision support tools might be the most effectively targeted toward patients in an intensive care unit, where risk is the greatest, rather than broadly across all inpatients, with the caveat that the use of propofol might need to be more closely monitored in subjects without cardiovascular disease or critical illness. Our findings also suggest that specific combinations of medications, such as amiodarone and propofol, should either be avoided or administered with close monitoring and aggressive treatment of other factors that could predispose risk of diLQTS, such as electrolyte abnormalities.

Finally, our findings highlight the critical trade-off between model interpretability and accuracy, as we found that a black-box prediction model using deep learning was significantly more accurate (greater AUC and area under precision recall curve) than the more interpretable cluster-based model. This finding raises a key question for all practitioners of predictive modeling: Is the improvement in predictive accuracy worth the lack of understanding for why the model makes the predictions it does? More specifically, without understanding how a model makes its predictions, how can it be challenged if a treating clinician believes it is less applicable for a particular patient, and what changes should be made if the predictive accuracy diminishes (a so-called “data shift” occurs [23,24]). It is not difficult for an experienced clinician to understand why patients who are critically ill (cluster 0) would be at increased risk or why combinations of medications with high risk of diLQTS would increase risk, and a method that can uncover these categories would seem to be more useful clinically than a black-box approach. Such clinical interpretation is unavailable for the deep learning model, which creates a challenge of trust in application. Further, in prior work, we demonstrated that reinforcement learning can be applied to cluster-based decision models (using a Q table) to allow a decision support tool to improve over time [35]; it is unclear whether a deep learning model could be as easily integrated with reinforcement learning or whether there would be sufficient prospective data to update the over 20 million parameters of such a model. Broadly, as increasing numbers of predictive models based on deep learning are applied to predict diLQTS, especially those applied directly to the ECG tracing itself [36,37], the trade-off with interpretability will remain a critical consideration in clinical applications.

Limitations

Principal among the limitations of this investigation is the high degree of noise inherent in studies of EHR data at scale and the challenges with having a lack of ability to perform detailed validation of diagnoses, medications, or outcomes, beyond what can be performed in silico without manual chart review. Several of these limitations related to reverse causation or lack of temporal granularity with medication administration are highlighted above. On the one hand, this common limitation of big data science limits what can be done in terms of granular validation; on the other hand, it provides both the improvement in statistical power for modeling and some protection against population bias, as might occur with studies at a single clinic or single provider level. With the increased expansion of EHR use worldwide, it is likely that methods to explore interpretability within these large data models will be increasingly relevant, for which our investigation should provide some foundation for how interpretability can be balanced against predictive accuracy.

Future Directions

Importantly, our findings provide the opportunity for direct clinical implementation of “smart” clinical decision tools that incorporate patient characteristics along with an understanding of patient risk to improve the accuracy of predictions of diLQTS, as well as guide clinical decisions including monitoring for those at high risk or selecting alternative agents where they are available. When combined with dynamic learning models, such as Q learning [35], our approach offers the opportunity to improve overall patient safety and clinical outcomes.

Conclusion

In summary, we found that interpretable methods to predict diLQTS allow for evaluation in a manner that facilitates deeper inspection of specific medication interactions and the identification of meaningful clinical populations to target for prevention. This interpretability comes at the expense of predictive accuracy, which must be considered among organizations seeking to integrate predictive modeling into clinical decision support tools to prevent diLQTS.

Acknowledgments

This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (MAR: R01HL146824), as well as Google, Inc. (MAR). We would also like to thank Michelle Edelmann and Premanand Tiwari of the University of Colorado Health Data Compass for assistance with data acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental Documents.
[PDF File (Adobe PDF File), 459 KB - jmir_v24i12e42163_app1.pdf]

References

1. Roden DM. Drug-induced prolongation of the QT interval. N Engl J Med 2004 Mar 04;350(10):1013-1022. [doi: 10.1056/NEJMra032426] [Medline: 14999113]
2. Mehlhorn AJ, Brown DA. Safety concerns with fluoroquinolones. Ann Pharmacother 2007 Nov;41(11):1859-1866. [doi: 10.1345/aph.1K347] [Medline: 17911203]
3. Sharma S, Martijn Bos J, Tarrell RF, Simon GJ, Morlan BW, Ackerman MJ, et al. Providers' response to clinical decision support for QT prolonging drugs. J Med Syst 2017 Sep 02;41(10):161. [doi: 10.1007/s10916-017-0803-7] [Medline: 28866768]
4. Trinkley KE, Pell JM, Martinez DD, Maude NR, Hale G, Rosenberg MA. Assessing prescriber behavior with a clinical decision support tool to prevent drug-induced long QT syndrome. Appl Clin Inform 2021 Jan;12(1):190-197 [FREE Full text] [doi: 10.1055/s-0041-1724043] [Medline: 33694143]
5. Pell JM, Cheung D, Jones MA, Cumbler E. Don't fuel the fire: decreasing intravenous haloperidol use in high risk patients via a customized electronic alert. J Am Med Inform Assoc 2014 Jun 13;21(6):1109-1112 [FREE Full text] [doi: 10.1136/amiajnl-2014-002777] [Medline: 24928174]
6. Cheung D, Cumbler E, Hale G, Pell J. Reining in the QTc: reducing the risk of Torsades de Pointes across a major health system. J Am Med Inform Assoc 2018 Sep 01;25(9):1202-1205 [FREE Full text] [doi: 10.1093/jamia/ocy081] [Medline: 29961858]
7. Gallo T, Heise CW, Woosley RL, Tisdale JE, Tan MS, Gephart SM, et al. Clinician responses to a clinical decision support advisory for high risk of Torsades de Pointes. J Am Heart Assoc 2022 Jun 07;11(11):e024338 [FREE Full text] [doi: 10.1161/JAHA.122.024338] [Medline: 35656987]
8. Ehret GB, Voide C, Gex-Fabry M, Chabert J, Shah D, Broers B, et al. Drug-induced long QT syndrome in injection drug users receiving methadone: high frequency in hospitalized patients and risk factors. Arch Intern Med 2006 Jun 26;166(12):1280-1287. [doi: 10.1001/archinte.166.12.1280] [Medline: 16801510]
9. Girardin FR, Gex-Fabry M, Berney P, Shah D, Gaspoz J, Dayer P. Drug-induced long QT in adult psychiatric inpatients: the 5-year cross-sectional ECG Screening Outcome in Psychiatry study. Am J Psychiatry 2013 Dec;170(12):1468-1476. [doi: 10.1176/appi.ajp.2013.12060860] [Medline: 24306340]
10. Goldstein LH, Gabin A, Fawaz A, Freedberg NA, Schwartz N, Elias M, et al. Azithromycin is not associated with QT prolongation in hospitalized patients with community-acquired pneumonia. Pharmacoepidemiol Drug Saf 2015 Oct;24(10):1042-1048. [doi: 10.1002/pds.3842] [Medline: 26238864]
11. Aakjaer M, Werther SK, de Bruin ML, Andersen M. Serious arrhythmia in initiators of citalopram, escitalopram, and other selective serotonin reuptake inhibitors: a population-based cohort study in older adults. Clin Transl Sci 2022 Sep;15(9):2105-2115 [FREE Full text] [doi: 10.1111/cts.13319] [Medline: 35733364]
12. Tabrizi S, Heidari S, Rafiei H. Investigation role of ondansetron on long QT interval among non-cardiac patients. Ann Med Surg (Lond) 2021 Nov;71:102971 [FREE Full text] [doi: 10.1016/j.amsu.2021.102971] [Medline: 34712483]
13. Hincapie-Castillo JM, Staley B, Henriksen C, Saidi A, Lipori GP, Winterstein AG. Development of a predictive model for drug-associated QT prolongation in the inpatient setting using electronic health record data. Am J Health Syst Pharm 2019 Jul 02;76(14):1059-1070. [doi: 10.1093/ajhp/zxz100] [Medline: 31185072]
14. Tisdale JE, Jaynes HA, Kingery JR, Mourad NA, Trujillo TN, Overholser BR, et al. Development and validation of a risk score to predict QT interval prolongation in hospitalized patients. Circ Cardiovasc Qual Outcomes 2013 Jul;6(4):479-487 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.113.000152] [Medline: 23716032]
15. Tomaselli Muensterman E, Tisdale JE. Predictive analytics for identification of patients at risk for QT interval prolongation: a systematic review. Pharmacotherapy 2018 Aug;38(8):813-821. [doi: 10.1002/phar.2146] [Medline: 29882591]
16. Vandael E, Vandenberk B, Vandenberghe J, van den Bosch B, Willems R, Foulon V. A smart algorithm for the prevention and risk management of QTc prolongation based on the optimized RISQ-PATH model. Br J Clin Pharmacol 2018 Dec;84(12):2824-2835 [FREE Full text] [doi: 10.1111/bcp.13740] [Medline: 30112769]

17. Berger FA, van der Sijs H, Becker ML, van Gelder T, van den Bemt PMLA. Development and validation of a tool to assess the risk of QT drug-drug interactions in clinical practice. BMC Med Inform Decis Mak 2020 Jul 23;20(1):171 [FREE Full text] [doi: 10.1186/s12911-020-01181-3] [Medline: 32703198]
18. Kim RS, Simon S, Powers B, Sandhu A, Sanchez J, Borne RT, et al. Machine learning methodologies for prediction of rhythm-control strategy in patients diagnosed with atrial fibrillation: observational, retrospective, case-control study. JMIR Med Inform 2021 Dec 06;9(12):e29225 [FREE Full text] [doi: 10.2196/29225] [Medline: 34874889]
19. Simon ST, Mandair D, Tiwari P, Rosenberg MA. Prediction of drug-induced long QT syndrome using machine learning applied to harmonized electronic health record data. J Cardiovasc Pharmacol Ther 2021 Jul;26(4):335-340 [FREE Full text] [doi: 10.1177/1074248421995348] [Medline: 33682475]
20. Rosenberg MA. Trusting magic: interpretability of predictions from machine learning algorithms. Circulation 2021 Mar 30;143(13):1299-1301 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.121.053733] [Medline: 33779269]
21. Lipton ZC. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue 2018 Jun;16(3):31-57. [doi: 10.1145/3236386.3241340]
22. Stacke K, Eilertsen G, Unger J, Lundstrom C. Measuring domain shift for deep learning in histopathology. IEEE J Biomed Health Inform 2021 Feb;25(2):325-336. [doi: 10.1109/JBHI.2020.3032060] [Medline: 33085623]
23. Kheterpal S, Singh K, Topol EJ. Digitising the prediction and management of sepsis. Lancet 2022 Apr 16;399(10334):1459. [doi: 10.1016/S0140-6736(22)00658-4] [Medline: 35430013]
24. 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 [FREE Full text] [doi: 10.1056/NEJMc2104626] [Medline: 34260843]
25. Chollet F. Keras. Github. 2015. URL: https://github.com/fchollet/keras [accessed 2022-11-23]
26. Jiang YH, Long J, Zhao ZB, Li L, Lian ZX, Liang Z, et al. Gene co-expression network based on part mutual information for gene-to-gene relationship and gene-cancer correlation analysis. BMC Bioinformatics 2022 May 24;23(1):194 [FREE Full text] [doi: 10.1186/s12859-022-04732-9] [Medline: 35610556]
27. Youden WJ. Index for rating diagnostic tests. Cancer 1950 Jan;3(1):32-35. [doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3] [Medline: 15405679]
28. Wang J, Sun Z, Tao S. Torsade de pointes caused by citalopram during the pacemaker battery-depletion phase: a case report. Ann Noninvasive Electrocardiol 2022 Jul;27(4):e12936 [FREE Full text] [doi: 10.1111/anec.12936] [Medline: 35322503]
29. Ray WA, Murray KT, Hall K, Arbogast PG, Stein CM. Azithromycin and the risk of cardiovascular death. N Engl J Med 2012 May 17;366(20):1881-1890 [FREE Full text] [doi: 10.1056/NEJMoa1003833] [Medline: 22591294]
30. Tiver KD, Dharmaprani D, Quah JX, Lahiri A, Waddell-Smith KE, Ganesan AN. Vomiting, electrolyte disturbance, and medications; the perfect storm for acquired long QT syndrome and cardiac arrest: a case report. J Med Case Rep 2022 Jan 11;16(1):9 [FREE Full text] [doi: 10.1186/s13256-021-03204-7] [Medline: 35012656]
31. Singh P, Maldonado-Duran JM. Drug-induced QT prolongation as a result of an escitalopram overdose in a patient with previously undiagnosed congenital long QT syndrome. Case Rep Med 2014 Jul 01;2014:917846 [FREE Full text] [doi: 10.1155/2014/917846] [Medline: 25101129]
32. Lopez-Medina AI, Chahal CAA, Luzum JA. The genetics of drug-induced QT prolongation: evaluating the evidence for pharmacodynamic variants. Pharmacogenomics 2022 Jun;23(9):543-557. [doi: 10.2217/pgs-2022-0027] [Medline: 35698903]
33. Strauss DG, Vicente J, Johannesen L, Blinova K, Mason JW, Weeke P, et al. Common genetic variant risk score is associated with drug-induced QT prolongation and Torsade de Pointes risk: a pilot study. Circulation 2017 Apr 04;135(14):1300-1310 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.116.023980] [Medline: 28213480]
34. Roden DM, Viswanathan PC. Genetics of acquired long QT syndrome. J Clin Invest 2005 Aug;115(8):2025-2032 [FREE Full text] [doi: 10.1172/JCI25539] [Medline: 16075043]
35. Levy AE, Biswas M, Weber R, Tarakji K, Chung M, Noseworthy PA, et al. Applications of machine learning in decision analysis for dose management for dofetilide. PLoS One 2019 Dec 31;14(12):e0227324 [FREE Full text] [doi: 10.1371/journal.pone.0227324] [Medline: 31891645]
36. Prifti E, Fall A, Davogustto G, Pulini A, Denjoy I, Funck-Brentano C, et al. Deep learning analysis of electrocardiogram for risk prediction of drug-induced arrhythmias and diagnosis of long QT syndrome. Eur Heart J 2021 Oct 07;42(38):3948-3961. [doi: 10.1093/eurheartj/ehab588] [Medline: 34468739]
37. Bos JM, Attia ZI, Albert DE, Noseworthy PA, Friedman PA, Ackerman MJ. Use of artificial intelligence and deep neural networks in evaluation of patients with electrocardiographically concealed long QT syndrome from the surface 12-lead electrocardiogram. JAMA Cardiol 2021 May 01;6(5):532-538 [FREE Full text] [doi: 10.1001/jamacardio.2020.7422] [Medline: 33566059]

Abbreviations

AUC: area under the receiver operating characteristic curve
diLQTS: Drug-induced long-QT syndrome
ECG: electrocardiogram

EHR: electronic health record
MIC: maximum information coefficient

Edited by G Eysenbach; submitted 24.08.22; peer-reviewed by H Sun, D Albert; comments to author 07.10.22; revised version received 31.10.22; accepted 17.11.22; published 01.12.22.

Please cite as:
Simon ST, Trinkley KE, Malone DC, Rosenberg MA
Interpretable Machine Learning Prediction of Drug-Induced QT Prolongation: Electronic Health Record Analysis
J Med Internet Res 2022;24(12):e42163
URL: https://www.jmir.org/2022/12/e42163
doi:10.2196/42163
PMID:36454608

Original Paper

A Machine Learning-Based Approach to Predict Prognosis and Length of Hospital Stay in Adults and Children With Traumatic Brain Injury: Retrospective Cohort Study

Cheng Fang[1*], MM; Yifeng Pan[2*], MSE; Luotong Zhao[1], MM; Zhaoyi Niu[1], MM; Qingguo Guo[1], MM; Bing Zhao[1], MD

[1]Department of Neurosurgery, The Second Affiliated Hospital of Anhui Medical University, Anhui Medical University, Hefei, China

[2]The School of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei, China

[*]these authors contributed equally

Corresponding Author:
Bing Zhao, MD
Department of Neurosurgery
The Second Affiliated Hospital of Anhui Medical University
Anhui Medical University
No. 678, Furong Raod
Hefei, 230601
China
Phone: 86 138 6611 2073
Fax: 86 0551 63869400
Email: aydzhb@126.com

Abstract

Background: The treatment and care of adults and children with traumatic brain injury (TBI) constitute an intractable global health problem. Predicting the prognosis and length of hospital stay of patients with TBI may improve therapeutic effects and significantly reduce societal health care burden. Applying novel machine learning methods to the field of TBI may be valuable for determining the prognosis and cost-effectiveness of clinical treatment.

Objective: We aimed to combine multiple machine learning approaches to build hybrid models for predicting the prognosis and length of hospital stay for adults and children with TBI.

Methods: We collected relevant clinical information from patients treated at the Neurosurgery Center of the Second Affiliated Hospital of Anhui Medical University between May 2017 and May 2022, of which 80% was used for training the model and 20% for testing via screening and data splitting. We trained and tested the machine learning models using 5 cross-validations to avoid overfitting. In the machine learning models, 11 types of independent variables were used as input variables and Glasgow Outcome Scale score, used to evaluate patients' prognosis, and patient length of stay were used as output variables. Once the models were trained, we obtained and compared the errors of each machine learning model from 5 rounds of cross-validation to select the best predictive model. The model was then externally tested using clinical data of patients treated at the First Affiliated Hospital of Anhui Medical University from June 2021 to February 2022.

Results: The final convolutional neural network–support vector machine (CNN-SVM) model predicted Glasgow Outcome Scale score with an accuracy of 93% and 93.69% in the test and external validation sets, respectively, and an area under the curve of 94.68% and 94.32% in the test and external validation sets, respectively. The mean absolute percentage error of the final built convolutional neural network–support vector regression (CNN-SVR) model predicting inpatient time in the test set and external validation set was 10.72% and 10.44%, respectively. The coefficient of determination (R^2) was 0.93 and 0.92 in the test set and external validation set, respectively. Compared with back-propagation neural network, CNN, and SVM models built separately, our hybrid model was identified to be optimal and had high confidence.

Conclusions: This study demonstrates the clinical utility of 2 hybrid models built by combining multiple machine learning approaches to accurately predict the prognosis and length of stay in hospital for adults and children with TBI. Application of these models may reduce the burden on physicians when assessing TBI and assist clinicians in the medical decision-making process.

(J Med Internet Res 2022;24(12):e41819) doi:10.2196/41819

KEYWORDS

convolutional neural network; machine learning; neurosurgery; support vector machine; support vector regression; traumatic brain injury

Introduction

Background

More than 50 million people worldwide suffer from traumatic brain injury (TBI) each year, which reduces patient quality of life and leads to high morbidity and mortality. Approximately half of the global population is likely to experience one or more brain injuries in their lifetime [1,2]. The greatest burden of TBI has been reported in low- and middle-income countries [3], where medical resources are limited and medical experience is lacking and patients often have a poor prognosis, further adding to the medical burden on society. Therefore, creating a tool that can predict patient prognosis and length of stay to aid clinician medical decisions is essential to achieve precision medicine [4].

With the popularity of computers and the rapid development of computer science, people are increasingly using computer knowledge to solve practical problems, and machine learning methods are gaining more and more attention from scientists. Machine learning is a scientific discipline that focuses on how computers learn from data and has been widely used in military and civilian applications [5]. Research incorporating computer algorithms into medicine has also been reported [6-10]. Currently, common machine learning methods include artificial neural networks and back propagation (BP) neural networks, some of the classical algorithms that have been widely used, but their drawback of easy overfitting is difficult to solve. Novel algorithms such as convolutional neural networks (CNN), support vector machine (SVM), and support vector regression (SVR) have solved this drawback well, allowing for more accurate machine learning models to be built. There has been very little research into the integration of these algorithms into clinical practice, let alone into complex studies such as predicting patient prognosis and length of stay. CNNs are a class of feedforward neural networks that incorporate convolutional computation, have deep structure, and are one of the representative algorithms for deep learning [11]. CNNs are built to mimic biological visual perception mechanisms and can directly process 2D images, hence their wide application in image processing [12,13]. Considering that CNN has achieved great success in the image field, we would like to see if CNN can also have good prediction and classification results when the input data is structured data. SVM is a class of generalized linear classifier that performs binary classification of data in a supervised learning fashion, where the decision boundary is a maximum-margin hyperplane solved for the learned samples. While SVM itself is proposed for classification problems, SVR is an important application branch of SVM, which is an application of SVM for regression prediction problems, both of which are applicable to our study. Compared with traditional machine learning methods, CNN, SVM, and SVR have faster learning speed, better network generalization, and more accurate classification and prediction of variables.

Aim

The aim of this study is to apply the latest algorithms in machine learning to medicine for outcome prediction based on relevant clinical data. Machine learning methods create a mapping relationship between input and output data through the analytical processing of raw data and the application of various prediction algorithms. CNN, SVM, and SVR are among the representative new generation algorithms. Therefore, we combine the advantages of these methods and build two hybrid models for patient prediction. In this study, we aim to demonstrate that both methods are effective in predicting patients with traumatic injuries, and we hope to provide inspiration to future researchers working in health care information analysis.

Methods

Ethics Approval

This retrospective cohort study was approved by the Ethics Committee of the Second Affiliated Hospital of Anhui Medical University in December of 2020 (approval number S20210098). Participants or proxies signed the relevant informed consent forms within 24 hours of admission.

Participants

The predictive model was developed based on relevant data from 1001 patients registered at the Neurosurgery Center of the Second Affiliated Hospital of Anhui Medical University with traumatic craniocerebral injury between May 2017 and May 2022 and at the First Hospital of Anhui Medical University between June 2021 and February 2022. By random splitting, we used 80% of the data in the training model and the remaining 20% to test model performance. We also collected clinical data from 111 patients at the First Affiliated Hospital of Anhui Medical University as a test cohort for external validation of the model.

We included all patients with complete demographic, clinical, and radiological data during this period, with inclusion criteria of (1) craniocerebral trauma as a result of external forces; (2) clinical diagnosis of craniocerebral trauma; and (3) complete clinical data including cases, course records, imaging examinations, and test reports were available. Our model prediction results included the length of stay of the patient; the decision regarding hospital discharge requires discussion by the treatment team and assessment by experienced neurosurgeons. Hence, the length of stay results were inaccurate when a patient died, when treatment was abandoned by the subjective will of the patient's family, or when follow-up was lost due to transfer to other departments. Exclusion criteria were as follows: (1) patients died during hospitalization, (2) patients' family requested to abandon treatment (including financial factors) and patients were automatically discharged, (3) patients with other severe injuries who required transfer to relevant

departments for further treatment, and (4) patients with a history of craniocerebral injury (Figures 1 and 2).

Figure 1. Training set and test set. TBI: traumatic brain injury.

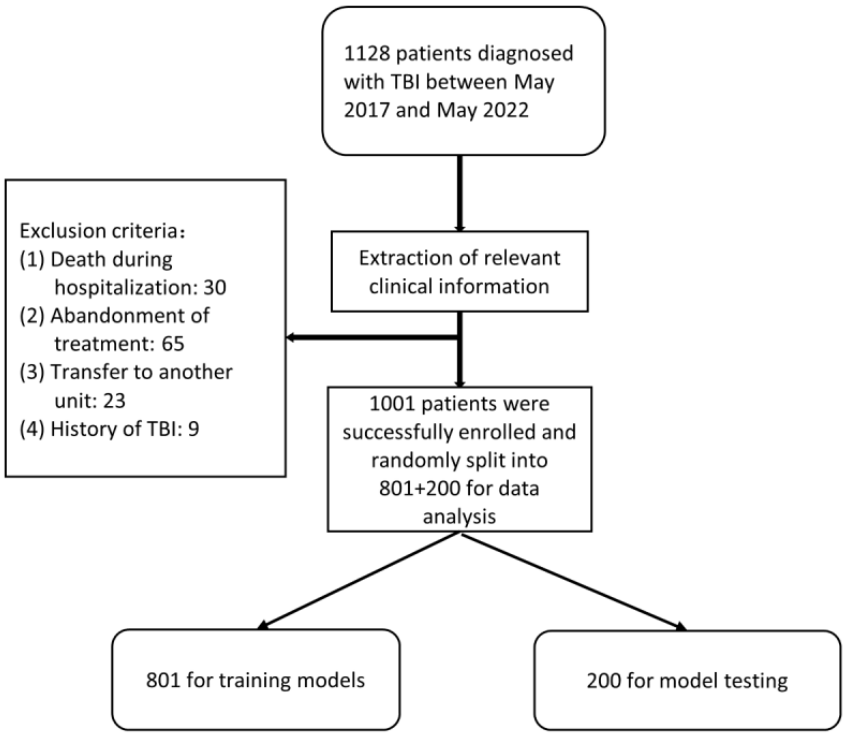
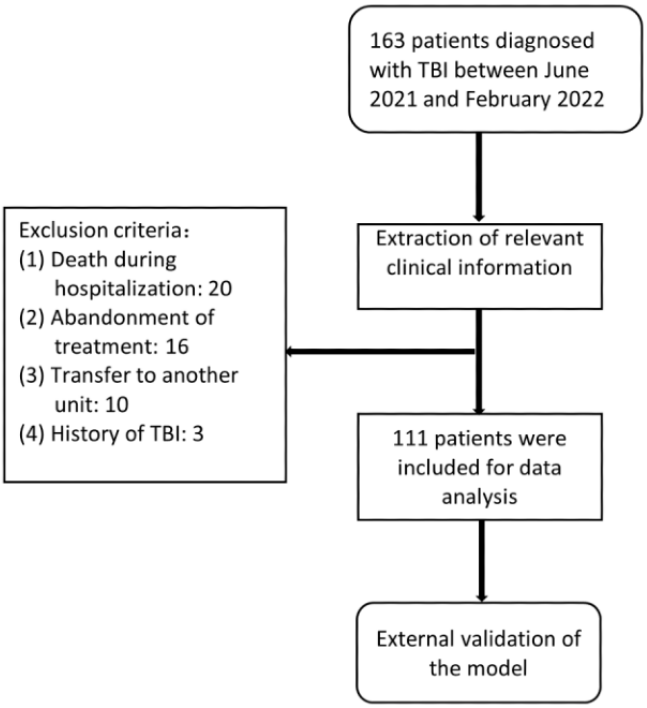


Figure 2. External validation set. TBI: traumatic brain injury.



Data Collection

By reviewing relevant papers, formally trained neurologists extracted the necessary data for modeling from the electronical medical records of enrolled patients, including patients' general characteristics (sex, age, and previous medical history), clinical and imaging data of patients with TBI (mechanism of TBI, loss of consciousness after injury, Glasgow Coma Scale [GCS] score, cranial computed tomography [CT] findings as jointly diagnosed by radiologists and neurosurgeons, other site injuries, treatment, admission to intensive care unit [ICU], and complications), and length of stay. The classification and definition of variables used to construct the predictive model are listed in Table 1.

Table 1. Variables used to construct the model.

Variable	Total (n=801), n (%)	Data type
Age (years)		
≥17	49 (6.1)	Floating point data
18-44	220 (27.5)	Floating point data
45-64	321 (40.1)	Floating point data
65-74	147 (18.3)	Floating point data
≤75	64 (8.0)	Floating point data
Gender		
Male	577 (72.0)	Binary data
Female	224 (28.0)	Binary data
Past medical history		
Hypertension	146 (18.2)	Binary data
Diabetes	42 (5.2)	Binary data
Coronary artery disease	16 (2.0)	Binary data
Chronic renal failure	6 (0.7)	Binary data
Cerebral infarction	18 (2.2)	Binary data
Respiratory disorders	10 (1.2)	Binary data
Mechanism of injury		
Fall on the same plane	214 (26.7)	Binary data
Fall from high place	140 (17.5)	Binary data
Road accident	415 (51.8)	Binary data
Object striking the head	32 (4.0)	Binary data
Loss of consciousness		
Yes	385 (48.0)	Binary data
No	416 (52.0)	Binary data
Glasgow Coma Scale score		
13-15	480 (59.9)	Binary data
9-12	123 (15.4)	Binary data
3-8	198 (24.7)	Binary data
Neuroimaging results		
Epidural hematoma	244 (30.5)	Binary data
Subdural hematoma	434 (54.2)	Binary data
Subarachnoid hemorrhage	411 (51.3)	Binary data
Skull fracture	509 (63.5)	Binary data
Diffuse axonal injury	13 (1.6)	Binary data
Brain herniation	20 (2.5)	Binary data
Treatment		
Conservative	180 (22.5)	Binary data
Neurological surgery	621 (77.5)	Binary data
Other site injuries		
Fractures in other areas	232 (29.0)	Binary data
Visceral contusions	18 (2.2)	Binary data
Traumatic wet lung	103 (12.9)	Binary data

Variable	Total (n=801), n (%)	Data type
Pneumothorax	16 (2.0)	Binary data
Duration of intensive care unit stay (days)		
≤5	103 (12.9)	Floating point data
6-15	94 (11.7)	Floating point data
≥16	30 (3.7)	Floating point data
Complications		
Infections	191 (23.8)	Binary data
Tracheotomy	133 (16.6)	Binary data
Electrolyte disorders	230 (28.7)	Binary data
Impaired organ function	256 (32.0)	Binary data
Anemia	118 (14.7)	Binary data
Abnormal blood clotting	36 (4.5)	Binary data
Cerebrospinal fluid leakage	18 (2.2)	Binary data
Glasgow Outcome Scale score		
1	0(0)	Binary data
2	85 (10.6)	Binary data
3	97 (12.1)	Binary data
4	419 (52.3)	Binary data
5	200 (25.0)	Binary data
Length of stay in hospital (days)		
≤10	225 (28.1)	Floating point data
11-20	365 (45.6)	Floating point data
21-30	152 (19.0)	Floating point data
31-40	53 (6.6)	Floating point data
≥41	6 (0.7)	Floating point data

The Glasgow Outcome Scale (GOS) score published by Jennett and Bond in 1975 [14] has emerged as one of the most widely used prognostic tools for assessing recovery after disability and TBI worldwide (Table 2). Patients with scores of 1 who died were excluded, those with scores of 5 and 4 were considered to have recovered, and patients with scores of 2 and 3 were considered to have a poorer prognosis. This supports our use of this tool as a criterion to evaluate patient prognosis.

Table 2. Descriptions of the categories of the Glasgow Outcome Scale.

Score		
1	Dead	As a direct result of brain trauma, or...due to secondary complications or other complications
2	Vegetative state	Patients who remain unresponsive and speechless…
3	Severe disability	Patient is conscious but needs the assistance of another person for some activities of daily living every day...
4	Moderate disability	Patient can look after themself at home, get out and about to the shops, and travel by public transport. However, some previous activities, either at work or in social life, are now no longer possible by reason of either physical or mental deficit...
5	Good recovery	Patient has the capacity to resume normal occupational and social activities, although there may be minor physical or mental deficits...social outcome should be included in the assessment here, such as leisure activities and family relationships

Modeling

Neurologists were involved throughout the model development process and supervised the clinical application of the algorithm to ensure that the model's predictions are meaningful and the research process meets the requirements of the ethics committee.

CNN-SVM Hybrid Model for Predicting GOS Score

The ability of CNN to extract data features is applicable to the processing of multidimensional input data in this study [15]. SVM can automatically identify support vectors that have better differentiation ability for classification. The resulting classifier can maximize the class-to-class interval, thus demonstrating better adaptability and higher classification accuracy [16], which is applicable to the prediction of GOS classification results. Therefore, this study combined CNN and SVM to build a hybrid CNN-SVM model for predicting the prognosis of patients with TBI. This model combined the respective advantages of CNN and SVM to improve model prediction accuracy and therefore exhibited greater advantages [17].

In the CNN-SVM hybrid model, the input layer consisted of 11 classes of input parameters, and the output layer was divided into 5 classes of GOS scores. Of the original 1001 data sets, 80% were randomly selected as the training set and 20% as the test set. Five rounds of model learning and validation were performed, and the average GOS classification accuracy of the 5 training and testing sessions was finally obtained. We developed the CNN-SVM hybrid model using the Pyrorch framework and Python 3.9 programming language.

The CNN-SVM model was used to make classification predictions for GOS score. Cross entropy was selected as the loss function of the model. Hyperparameters were selected through training, and the full training data were then retrained with the optimal parameters of the optimal model. After several attempts, rectified linear unit was selected as the activation function of the model [18]. The optimizer used momentum gradient descent [19]. The learning rate was set to 10-3, and the batch size was set to 64 according to the number of samples in the training set to ensure memory utilization and enhance processing speed for the same amount of data. The SVM model used an radical basis function kernel to avoid falling into local optimal solution. Penalty factor (P=100) and kernel parameter (γ=0.02) of the SVM were finalized using the grid search method [20]. The process of building the CNN-SVM hybrid model is shown in Figure 3; regarding the setup in CNN, details about our convolutional and pooling layers for structured data are shown in the Table 3.

Figure 3. Convolutional neural network–support vector machine hybrid model building process. SVM: support vector machine; GOS: Glasgow Outcome Score.

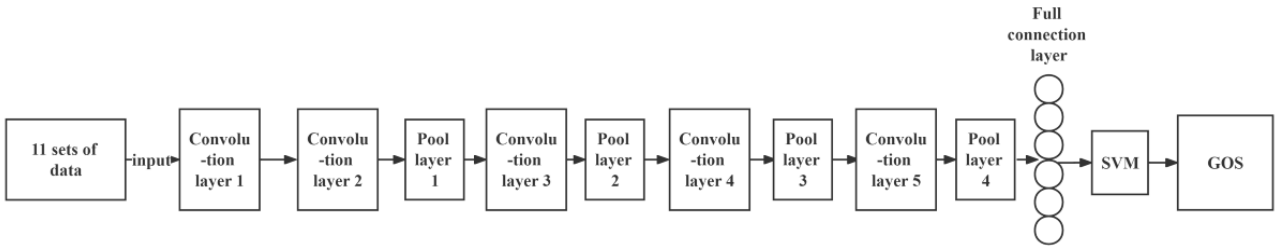


Table 3. Parameter setting of CNN.

Network layer	Model parameter setting
Input layer	Data matrix
Convolution layer 1	64 1×1 convolution kernels; kernel_size = 5
Convolution layer 2	128 1×1 convolution kernels; kernel_size = 5
Pool layer 1	MaxPool; kernel_size =1; stride = 2
Convolution layer 3	128 1×1 convolution kernels; kernel_size = 5
Pool layer 2	MaxPool; kernel_size =1; stride = 2
Convolution layer 4	256 1×1 convolution kernels; kernel_size = 5
Pool layer 3	MaxPool; kernel_size =1; stride = 2
Convolution layer 5	516 1×1 convolution kernels; kernel_size = 5
Pool layer 4 (adaptive pooling layer)	Output 1D vector
Full connection layer	Output

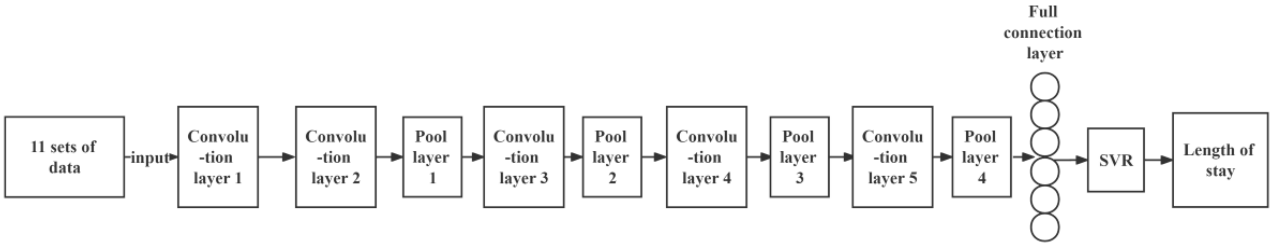
CNN-SVR Hybrid Model to Predict Length of Stay

SVMs are used for classification problems. SVR is a key application branch of SVMs. SVR and SVM are distinct in that SVM aims to maximize the distance to the nearest sample point in the hyperplane, while SVR aims to minimize the distance to the farthest sample point in the hyperplane. Therefore, SVR is applicable to the regression prediction of length of stay in this study but not to the prediction of classification problems. We combined CNN and SVR to build a hybrid CNN-SVR model for predicting the hospital stay of patients with TBI. The input layer consisted of 11 input parameters, and the output layer was length of stay. We randomly selected 80% of the original 1001 data sets as the training set and 20% as the test set. In this study, the mean absolute percentage error (MAPE) was used to measure the error between the real hospitalization time and predicted hospitalization time in the model, as shown in equation

(1), where y_i represents the real data and $\hat{y}_i$ represents the predicted data. We developed the CNN-SVR hybrid model using the Pyrorch framework and Python 3.9 programming language.

The CNN-SVR model was used to predict the length of stay of patients with TBI. The mean square error was selected as the loss function for the regression prediction model. Hyperparameters were selected through training, and the full training data were retrained with the optimal parameters of the optimal model. After several attempts, rectified linear unit was selected as the activation function of the model. The optimizer was used as reported by Kingma and Ba [21]. The learning rate was set to 10-3, and the batch size was set to 32 according to the number of samples in the training set. The radical basis function kernel was used for the SVR model. The penalty factor of P=100 for SVR was finally determined using a grid search method with the kernel parameter γ=0.01. The process of building the CNN-SVR hybrid model shows in Figure 4, and details about our convolutional and pooling layers for structured data are shown in the Table 3.

Figure 4. Convolutional neural network–support vector regression hybrid model building process. SVR: support vector regression.



Results

Evaluation Indicators

All 1001 valid samples were divided into training and test sets. The data were divided according to the rule of having similar statistical characteristics. The training set was divided into 0.8 of the total sample size and included the cross-validation data. Training and testing were repeated 5 times and averaged. For the classification prediction of GOS scores, the metric of precision was used to measure classification accuracy. For the prediction model of length of stay in patients with TBI, coefficient of determination (R^2) and MAPE were used to examine model performance.

Predicting the Prognosis of Patients With TBI: GOS Scores

To establish the optimal GOS classification model, this study compared the CNN-SVM model design with the construction of CNN, SVM, and BP neural network models. Accuracy indicates the proportion of correctly classified samples in the test set to the total sample size and can evaluate the predictive accuracy of the model, receiver operating characteristic curve is a composite indicator of sensitivity and specificity continuous variables, and area under the curve (AUC) can evaluate the generalization ability of the model. We used the accuracy and AUC to evaluate the advantages and disadvantages of the classification models. The accuracy and AUC of the 4 models (BP, CNN, SVM, and CNN-SVM) in the test data set are presented in Table 4. The classification results of the CNN-SVM hybrid model exhibited the highest accuracy and AUC; accuracy was 16.50%, 9.00%, and 5.50% higher than the BP, CNN, and SVM models, respectively, and AUC was 15.75%, 10.47%, and 6.33% higher than the BP, CNN, and SVM models, respectively. These results indicate that the classification prediction of GOS score using the hybrid CNN-SVM model is optimal.

Table 4. Accuracy and area under the curve of the four models.

Model	ACC[a], %		AUC[b], %	
	Training set	Testing set	Training set	Testing set
BP[c]	83.63	76.50	86.57	78.93
CNN[d]	87.63	84.00	87.18	84.21
SVM[e]	90.13	87.50	91.24	88.35
CNN-SVM	94.13	93.00	96.89	94.68

[a]ACC: accuracy.

[b]AUC: area under the curve.

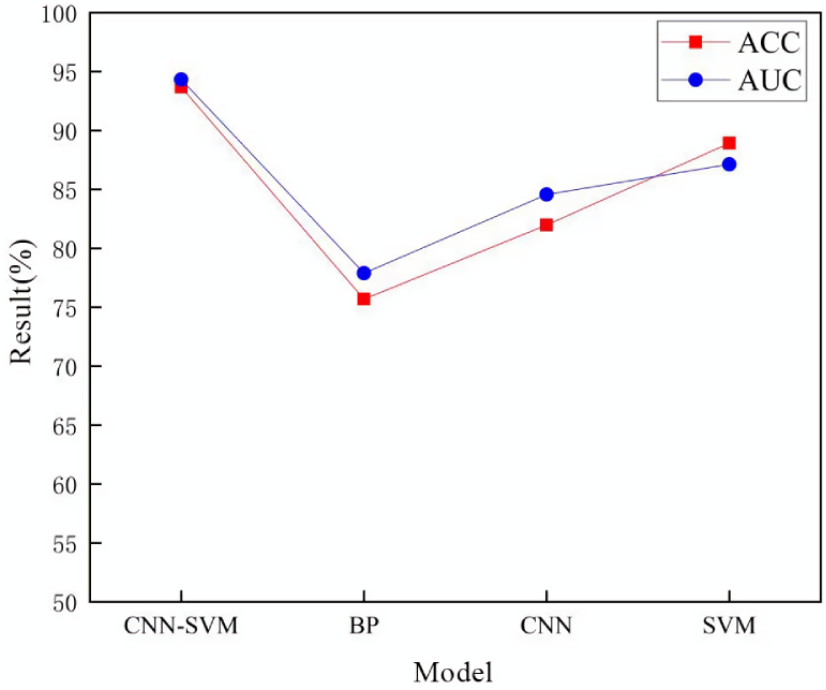
[c]BP: back propagation.

[d]CNN: convolutional neural network.

[e]SVR: support vector machine.

To further validate the reliability and merit of CNN-SVM, 111 data records from the First Affiliated Hospital of Anhui Medical University were used for validation. Eleven types of input variables collected were fed into the 4 models to compare the classification results; the experimental results are shown in Figure 5. The accuracy of GOS score classified by the 4 models (CNN-SVM, BP, CNN, and SVM) was 93.69%, 75.68%, 81.98%, and 88.29%, respectively, and the AUC was 94.32%, 77.89%, 84.57%, and 87.12%, respectively. The accuracy and AUC of CNN-SVM model were the best. Therefore, the CNN-SVM model is still the optimal model for predicting the GOS classification model through the validation of external hospital data.

Figure 5. Accuracy and area under the curve of the 4 models in the external validation set. ACC: accuracy; AUC: area under the curve; CNN-SVM: convolutional neural network–support vector machine; BP: back propagation; CNN: convolutional neural network; SVM: support vector machine.



Predicting Length of Hospital Stay in Patients With TBI

To establish the optimal model for predicting the length of hospital stay for patients with TBI, this study compared the construction of CNN, SVR, and BP neural network models based on the design of a CNN-SVR model. The R^2 and MAPE between the 4 models (CNN-SVR, BP, CNN, and SVR) for predicting length of stay and the true length of stay in the training and test sets are presented in Table 5. The CNN-SVR hybrid model exhibited the smallest error in the prediction results. Compared with that of the BP, CNN, and SVR models, MAPE was reduced by 7.61%, 10.15%, and 3.65%, respectively, indicating that the hybrid CNN-SVR model optimally predicted the length of hospital stay with higher prediction accuracy. The R^2 of the CNN-SVR model was higher than that of the other 3 models, with values of 0.96 and 0.93 for the training and test sets, respectively, indicating that the CNN-SVR had the best fit.

These results indicated that the CNN-SVR model had high regression fit and regression accuracy for the training samples and good learning ability. The model could be trained to the maximum extent with the existing data while accurately approximating the actual values of the training samples. In general, the regression fit and prediction accuracy of the model for the predicted samples were lower than that of the modeled samples. A smaller difference between these parameters indicated better generalization ability of the model. The evaluation data of the models revealed that the CNN-SVR model most closely approximated the actual prediction ability and modeling effects with the smallest difference. This indicated better robustness and actual generalization performance of the CNN model and that this model was most suitable as the prediction model.

Table 5. Mean absolute percentage error and coefficient of determination for four model predictions.

Model	MAPE[a], %		R^2	
	Training error	Testing error	Training error	Testing error
BP[b]	13.18	18.33	0.82	0.79
CNN[c]	15.29	20.69	0.76	0.73
SVR[d]	10.86	14.37	0.89	0.85
CNN-SVR	8.12	10.72	0.96	0.93

[a]MAPE: mean absolute percentage error.

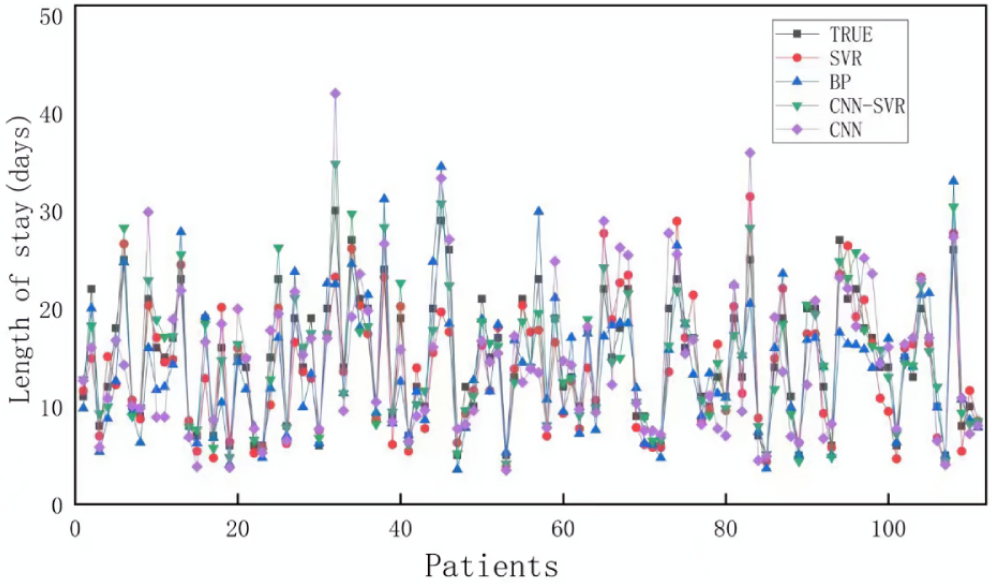
[b]BP: back propagation.

[b]CNN: convolutional neural network.

[b]SVR: support vector regression.

To further compare the reliability of the algorithms, 111 data records from the First Affiliated Hospital of Anhui Medical University were used for external validation. In total, 11 types of input variables from the collected data were input into the 4 models to compare the strengths and weaknesses of the models' prediction effects. The experimental results are presented in Figure 6. Overall, the CNN-SVR predicted hospital length-of-stay contours were generally consistent with the patients' true length-of-stay contours. And we further calculated the MAPE and R^2 for CNN-SVR, BP, CNN, and SVR models as 10.44%, 17.60%, 20.71%, and 16.00%, and 0.92, 0.76, 0.73, and 0.79, respectively. The CNN-SVR model performed better with regard to both MAPE and R^2 metrics. Therefore, the CNN-SVR model was the optimal model for predicting the length of stay of patients with TBI, as validated by external hospital data.

Figure 6. Comparison of the predicted results of the 4 models with the true values. SVR: support vector regression; BP: back propagation; CNN-SVR: convolutional neural network–support vector regression; CNN: convolutional neural network.



Discussion

Principal Findings

It is feasible to apply CNN, SVM, and SVR to the development of hybrid prediction models, which outperform traditional algorithms. In this study, we compared and constructed 4 models for each of the 2 prediction results, and the 2 hybrid models, CNN-SVM and CNN-SVR, performed the best in all metrics in the prediction results. The first was a hybrid CNN-SVM model for predicting GOS scores, combining the respective strengths of CNN and SVM, with accuracies of 94.13%, 93.00%, and 93.69% and AUCs of 96.89%, 94.68%, and 94.32% in the training, testing, and external validation sets, respectively. The second model was a hybrid CNN-SVR model for predicting actual length of stay with MAPE and R^2 of 8.12% and 0.96, 10.72% and 0.93, and 10.44% and 0.92 in the training, test, and external validation sets, respectively. The data were optimal, indicating that our prediction model has high reliability and the results hold clinical utility. To our knowledge, this is the first study to build hybrid prediction models based on clinical data for prognosis and length of stay in hospital for patients with TBI.

Comparison With Prior Work

Previous studies have proposed a linear regression (LR) scoring system for clinical studies, but its specificity and sensitivity are low, and its predictive performance is inferior to that of

multivariate prediction models. Moreover, when LR is applied to describe complex multivariate nonlinear relationships, it may have low robustness and often requires complex transformations due to multicollinearity between variables [22]. Our machine learning models represent a new generation of multivariate statistical methods that can deal more effectively with multidimensional factors and are suitable for incorporating a wider range of risk factors for prediction. This ability reduces the reliance on practitioner experience and ensures objective results. For example, prediction of delayed graft function after kidney transplantation revealed that SVM-based machine learning exhibited better performance compared to LR [23]. Indeed, Feng et al [24] compared 22 machine learning methods with LR and reported that the AUC of LR was 0.83 with an accuracy of 88%, while almost all machine learning algorithms achieved higher AUC than that of LR. Compared to traditional methods, machine learning methods offer advantages in feature selection; the more factors considered, the more accurate the predictions become. In fact, machine learning methods have been applied to the field of TBI [9,25-27] using classical algorithms like artificial neural networks and BP neural networks, but these algorithms have serious shortcomings. This has also inspired us to apply the next generation of machine learning methods to clinical applications.

We used CNN in our study because it is a powerful machine learning model commonly used in the field of neurosurgery to analyze cranial CT scans [28-30]. With the development of computational power, the network depth of CNNs is increasing, enabling more accurate approximations of nonlinearly increasing objective functions. However, this is accompanied by increasing complexity of networks, making them difficult to optimize and prone to overfitting. Therefore, we introduced SVM and combined the two to build a more reliable model for a classification problem like predicting a patient's prognosis (GOS score); on the contrary, a patient's length of stay constitutes a regression prediction problem and SVR, a branch of SVM, was proposed precisely to solve the regression problem; therefore, we combined SVR and CNN to build a hybrid model for predicting the length of stay. The comparative validation shows that our model outperformed CNN, SVM or SVR, BP models built separately in training, testing and external validation sets. This confirms our conjecture that it is feasible and effective to use novel machine learning methods and combine their respective strengths in building hybrid models to solve prediction problems, and the results are clinically relevant.

Patients with TBI are generally sicker and have longer hospital stays than other patients, and this new form of hybrid predictive model could provide a reliable reference for health care decision makers in their work and help in managing patients more accurately. The 11 categories of input variables are available in previous studies, and these are all commonly used clinical data in the field of TBI that are easy to collect, which also demonstrates the operational and practical nature of our study.

Limitations

Although our study lays the groundwork for the use of machine learning–based modeling in the field of TBI, several limitations should be acknowledged. First, machine learning methods are a computational construct unfamiliar to most physicians and may be dismissed as esoteric or unproven. However, with the rapid advancement of technology, artificial intelligence and machine learning will inevitable become widely used tools in the future. Second, due to the location of the Children's Hospital of Anhui Province in our area, the sample size for severe TBI in children may be insufficient. In this regard, it may be necessary to cooperate with the children's hospital at later stages to collect data from as many children with TBI as possible to further refine the model. In this study, we collected patients' past medical history and classified them into 6 categories, which may not have a large enough sample size, and there are other types of past history that may affect the length of stay; later studies need to further expand the sample size to improve the accuracy of the model. In addition, there may also be small significant relationships between the input data; for example, a patient's GOS score on admission and a head CT suggestive of brain herniation may have a significant impact on the model. This involves multicollinearity between input variables and feature selection, which is the focus of our next phase of research.

There has been an increasing number of recent reports on the detection of body fluid markers in patients to predict patient prognosis [31,32]. Although prediction performance in these studies did not supersede that of machine learning–based models, these studies have provided insight with regard to collection of relevant predictors as input data for machine learning models. Despite the limitations of this study, it is the first to use a next generation algorithm to build hybrid models to predict prognosis and length of stay in hospital for patients with TBI, and our models work better compared with traditional algorithms, demonstrating that CNN-SVM and CNN-SVR models can be useful in clinical work.

Conclusions

In summary, our study is the first to combine multiple novel machine learning methods to develop hybrid models for application to TBI. Our hybrid models achieved excellent results and predicted target values quickly and accurately, with more stable performance. Further replication of the model may enable clinical teams and hospital managers to work collaboratively to provide optimal clinical care and may assist inexperienced practitioners in small remote or rudimentary facilities. We believe that our approach will provide more robust and accurate predictions and these can be updated in real time, with crucial implications for clinical work.

Acknowledgments

This research was funded by grant 31804h08020270 from the Key Research and Development Plan Project of Anhui Province, grant gxypZD2019019 from the College Excellent Youth Talent Support Program in Anhui Province, grant KJ2019A0267 from

the Key Projects of Natural Science Research in Anhui Province, and grant gxbjZD10 from the Academic Funding Project for Top Talents in Colleges and Universities in Anhui Province.

Authors' Contributions

BZ designed the study and revised the manuscript and was responsible for the whole study. LZ, ZN, and QG collected the data. CF screened and checked the data. YP built the models. CF and YP drafted the manuscript. All authors made substantial contributions to the study and provided the approval of the submitted version.

Conflicts of Interest

None declared.

References

1. Maas AIR, Menon DK, Adelson PD, Andelic N, Bell MJ, Belli A, TBIR Participants and Investigators. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 2017 Dec;16(12):987-1048. [doi: 10.1016/S1474-4422(17)30371-X] [Medline: 29122524]
2. Pickelsimer EE, Selassie AW, Gu JK, Langlois JA. A population-based outcomes study of persons hospitalized with traumatic brain injury: operations of the South Carolina Traumatic Brain Injury Follow-up Registry. J Head Trauma Rehabil 2006;21(6):491-504. [doi: 10.1097/00001199-200611000-00004] [Medline: 17122680]
3. Dewan M, Rattani A, Gupta S, Baticulon R, Hung Y, Punchak M, et al. Estimating the global incidence of traumatic brain injury. J Neurosurg 2018 Apr 01:1-18. [doi: 10.3171/2017.10.JNS17352] [Medline: 29701556]
4. Agarwal V, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of predicting health care utilization via web search behavior: a data-driven analysis. J Med Internet Res 2016 Sep 21;18(9):e251 [FREE Full text] [doi: 10.2196/jmir.6240] [Medline: 27655225]
5. Deo RC. Machine learning in medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
6. Monteiro M, Newcombe VFJ, Mathieu F, Adatia K, Kamnitsas K, Ferrante E, et al. Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head CT using deep learning: an algorithm development and multicentre validation study. Lancet Digit Health 2020 Jun;2(6):e314-e322 [FREE Full text] [doi: 10.1016/S2589-7500(20)30085-6] [Medline: 33328125]
7. Wang X, Zhong J, Lei T, Chen D, Wang H, Zhu L, et al. An artificial neural network prediction model for posttraumatic epilepsy: retrospective cohort study. J Med Internet Res 2021 Aug 19;23(8):e25090 [FREE Full text] [doi: 10.2196/25090] [Medline: 34420931]
8. Shahjouei S, Ghodsi SM, Zangeneh Soroush M, Ansari S, Kamali-Ardakani S. Artificial neural network for predicting the safe temporary artery occlusion time in intracranial aneurysmal surgery. J Clin Med 2021 Apr 02;10(7):1464 [FREE Full text] [doi: 10.3390/jcm10071464] [Medline: 33918168]
9. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. Stroke 2019 May;50(5):1263-1265. [doi: 10.1161/STROKEAHA.118.024293] [Medline: 30890116]
10. Tunthanathip T, Sae-Heng S, Oearsakul T, Sakarunchai I, Kaewborisutsakul A, Taweesomboonyat C. Machine learning applications for the prediction of surgical site infection in neurological operations. Neurosurg Focus 2019 Aug 01;47(2):E7. [doi: 10.3171/2019.5.FOCUS19241] [Medline: 31370028]
11. Wang P, Fan E, Wang P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recog Let 2021 Jan;141:61-67. [doi: 10.1016/j.patrec.2020.07.042]
12. Shen D, Wu G, Suk H. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017 Jun 21;19:221-248 [FREE Full text] [doi: 10.1146/annurev-bioeng-071516-044442] [Medline: 28301734]
13. Kattenborn T, Leitloff J, Schiefer F, Hinz S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS J Photogrammetry Remote Sens 2021 Mar;173:24-49. [doi: 10.1016/j.isprsjprs.2020.12.010]
14. Jennett B, Bond M. Assessment of outcome after severe brain damage. Lancet 1975 Mar 01;1(7905):480-484. [doi: 10.1016/s0140-6736(75)92830-5] [Medline: 46957]
15. Zhang W, Yu Y, Qi Y, Shu F, Wang Y. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. Transportmetrica Transport Sci 2019 Jul 08;15(2):1688-1711. [doi: 10.1080/23249935.2019.1637966]
16. Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008 Mar 07;131(Pt 3):681-689 [FREE Full text] [doi: 10.1093/brain/awm319] [Medline: 18202106]
17. Niu X, Suen CY. A novel hybrid CNN–SVM classifier for recognizing handwritten digits. Pattern Recog 2012 Apr;45(4):1318-1325. [doi: 10.1016/j.patcog.2011.09.021]
18. Jarrett K, Kavukcuoglu K, Ranzato M, LeCun Y. What is the best multi-stage architecture for object recognition? 2009 Presented at: IEEE 12th International Conference on Computer Vision (ICCV); 2009; Kyoto p. 2146-2153. [doi: 10.1109/iccv.2009.5459469]

19. Polyak B. Some methods of speeding up the convergence of iteration methods. USSR Computat Math Math Physics 1964 Jan;4(5):1-17. [doi: 10.1016/0041-5553(64)90137-5]
20. Hou K, Shao G, Wang H, Zheng L, Zhang Q, Wu S, et al. Research on practical power system stability analysis algorithm based on modified SVM. Prot Control Mod Power Syst 2018 May 8;3(1):1. [doi: 10.1186/s41601-018-0086-0]
21. Kingma DP, Ba J. Adam: a method for stochastic optimization. ArXiv. Preprint posted online on January 30, 2017 2017:1.
22. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 1996 Nov;49(11):1225-1231. [doi: 10.1016/s0895-4356(96)00002-9] [Medline: 8892489]
23. Decruyenaere A, Decruyenaere P, Peeters P, Vermassen F, Dhaene T, Couckuyt I. Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods. BMC Med Inform Decis Mak 2015 Oct 14;15:83 [FREE Full text] [doi: 10.1186/s12911-015-0206-y] [Medline: 26466993]
24. Feng J, Wang Y, Peng J, Sun M, Zeng J, Jiang H. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. J Crit Care 2019 Dec;54:110-116. [doi: 10.1016/j.jcrc.2019.08.010] [Medline: 31408805]
25. Gholipour C, Rahim F, Fakhree A, Ziapour B. Using an artificial neural networks (ANNs) model for prediction of intensive care unit (ICU) outcome and length of stay at hospital in traumatic patients. J Clin Diagn Res 2015 Apr;9(4):OC19-OC23 [FREE Full text] [doi: 10.7860/JCDR/2015/9467.5828] [Medline: 26023581]
26. Hale A, Stonko D, Lim J, Guillamondegui O, Shannon C, Patel M. Using an artificial neural network to predict traumatic brain injury. J Neurosurg Pediatr 2018 Nov 02;23(2):219-226 [FREE Full text] [doi: 10.3171/2018.8.PEDS18370] [Medline: 30485240]
27. Majdan M, Brazinova A, Rusnak M, Leitgeb J. Outcome prediction after traumatic brain injury: comparison of the performance of routinely used severity scores and multivariable prognostic models. J Neurosci Rural Pract 2017 Sep 04;8(1):20-29 [FREE Full text] [doi: 10.4103/0976-3147.193543] [Medline: 28149077]
28. Chang P, Kuoy E, Grinband J, Weinberg B, Thompson M, Homo R, et al. Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. AJNR Am J Neuroradiol 2018 Sep;39(9):1609-1616 [FREE Full text] [doi: 10.3174/ajnr.A5742] [Medline: 30049723]
29. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018 Dec 01;392(10162):2388-2396. [doi: 10.1016/S0140-6736(18)31645-3] [Medline: 30318264]
30. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. NPJ Digit Med 2018;1:9 [FREE Full text] [doi: 10.1038/s41746-017-0015-z] [Medline: 31304294]
31. Chen W, Yang J, Li B, Peng G, Li T, Li L, et al. Neutrophil to lymphocyte ratio as a novel predictor of outcome in patients with severe traumatic brain injury. J Head Trauma Rehabil 2018;33(1):E53-E59. [doi: 10.1097/HTR.0000000000000320] [Medline: 28520670]
32. Zadora P, Dabrowski W, Czarko K, Smolen A, Kotlinska-Hasiec E, Wiorkowski K, et al. Preoperative neutrophil-lymphocyte count ratio helps predict the grade of glial tumor: a pilot study. Neurol Neurochir Pol 2015;49(1):41-44. [doi: 10.1016/j.pjnns.2014.12.006] [Medline: 25666772]

Abbreviations

AUC: area under the curve
BP: back propagation
CNN: convolutional neural network
CT: computed tomography
GCS: Glasgow Coma Scale
GOS: Glasgow Outcome Scale
ICU: intensive care unit
LR: linear regression
MAPE: mean absolute percentage error
SVM: support vector machine
SVR: support vector regression
TBI: traumatic brain injury

Edited by G Eysenbach; submitted 09.08.22; peer-reviewed by H Ellethy, M Kapsetaki, D Barsasella, H Veldandi; comments to author 19.10.22; revised version received 04.11.22; accepted 15.11.22; published 09.12.22.

Please cite as:
Fang C, Pan Y, Zhao L, Niu Z, Guo Q, Zhao B
A Machine Learning-Based Approach to Predict Prognosis and Length of Hospital Stay in Adults and Children With Traumatic Brain Injury: Retrospective Cohort Study
J Med Internet Res 2022;24(12):e41819
URL: https://www.jmir.org/2022/12/e41819
doi:10.2196/41819
PMID:36485032

Original Paper

Social Media Recruitment Strategies to Recruit Pregnant Women Into a Longitudinal Observational Cohort Study: Usability Study

Chloe Pekarsky[1]; Janice Skiffington[1], MSc; Lara M Leijser[2], MSc, MD, PhD; Donna Slater[1,3], PhD; Amy Metcalfe[1,4], PhD

[1]Department of Obstetrics and Gynecology, University of Calgary, Calgary, AB, Canada

[2]Section of Neonatology, Department of Pediatrics, University of Calgary, Calgary, AB, Canada

[3]Department of Physiology and Pharmacology, University of Calgary, Calgary, AB, Canada

[4]Departments of Community Health Sciences and Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Chloe Pekarsky
Department of Obstetrics and Gynecology
University of Calgary
2500 University Dr NW
Calgary, AB, T2N 1N4
Canada
Phone: 1 4035606717
Email: chloe.pekarsky@ucalgary.ca

Abstract

Background: Use of social media for study recruitment is becoming increasingly common. Previous studies have typically focused on using Facebook; however, there are limited data to support the use of other social media platforms for participant recruitment, notably in the context of a pregnancy study.

Objective: Our study aimed to evaluate the effectiveness of Facebook, Twitter, and Instagram in recruiting a representative sample of pregnant women in a longitudinal pregnancy cohort study in Calgary, Alberta, between September 27, 2021, and April 24, 2022.

Methods: Paid advertisements were targeted at 18- to 50-year-old women in Calgary, with interests in pregnancy. Data regarding reach, link clicks, and costs were collected through Facebook Ads Manager (Meta Platforms, Inc) and Twitter Analytics (Twitter, Inc). The feasibility of each platform for recruitment was assessed based on the recruitment rate and cost-effectiveness. The demographic characteristics of the participants recruited through each source were compared using the chi-square test.

Results: Paid advertisements reached 159,778 social media users, resulting in 2390 link clicks and 324 participants being recruited. Facebook reached and recruited the highest number of participants (153/324, 47.2%), whereas Instagram saw the highest number of link clicks relative to the number of users who saw the advertisement (418/19,764, 2.11%). Facebook and Instagram advertisements were cost-effective, with an average cost-per-click of CAD $0.65 (US $0.84; SD $0.27, US $0.35) and cost-per-completer of CAD $7.89 (US $10.25; SD CAD $4.08, US $5.30). Twitter advertisements were less successful in terms of recruitment and costs. Demographic characteristics of participants did not differ based on recruitment source, except for education and income, where more highly educated and higher-income participants were recruited through Instagram or Twitter. Many issues related to fraudulent responses were encountered throughout the recruitment period.

Conclusions: Paid social media advertisements (especially Facebook and Instagram) are feasible and cost-effective methods for recruiting a large sample of pregnant women for survey-based research. However, future research should be aware of the potential for fraudulent responses when using social media for recruitment and consider strategies to mitigate this problem.

(J Med Internet Res 2022;24(12):e40298) doi:10.2196/40298

KEYWORDS

social media; Facebook; Twitter; Instagram; recruitment; pregnancy; surveys; questionnaires; fraudulent responses

Introduction

Background

Recruitment is crucial to the success of prospective studies; however, it remains one of the most challenging aspects of conducting research owing to its time-consuming and expensive nature [1]. To address the challenges of recruitment, researchers are increasingly turning to social media as a tool to reach potential participants [1,2]. Using social media for recruitment is particularly beneficial in the context of the COVID-19 pandemic where many health care services have moved to web-based formats, thus decreasing the likelihood that participants encounter study advertisements through traditional methods such as posters or postcards.

Approximately 85% of Canadians reported using social media in the past year, with Facebook, Twitter, and Instagram being 3 of the 4 most commonly used platforms [1,3]. Women between the ages of 25 and 34 years account for the greatest proportion of social media users in the country [3], demonstrating the potential of social media to reach pregnant women for study recruitment. Previous studies focusing on pregnancy have found that using traditional recruitment methods in conjunction with paid social media advertisements is an effective way of recruiting desired number of participants within a short period [1,4-6]. However, most literature on this topic focuses exclusively on comparing traditional methods with Facebook. Therefore, it is unknown whether findings from Facebook are generalizable to other platforms such as Twitter and Instagram.

Objectives

By examining the success of multiple paid advertisement campaigns across Facebook, Instagram, and Twitter simultaneously, this study aimed to determine (1) which social media platform is most effective for recruitment in terms of recruitment rate and cost; (2) what kind of advertisements lead to the most engagement with our study and the most number of participants; and (3) whether participants recruited through each platform differed from each other and from participants recruited through traditional methods.

Methods

The P3 Cohort Study

The P3 Cohort Study (Prediction, Prevention and Interventions for Preterm Birth) is a longitudinal cohort study aimed at recruiting 4000 pregnant women and their partners in Calgary, Alberta, to better understand preterm birth [7]. The study comprises 5 web-based surveys to be completed during pregnancy and the first year postpartum. In addition, the partner may choose to participate in 2 surveys. As of April 2022, participant pregnancy status and identity are self-reported but, following completion of recruitment, will be verified by medical records. Participants are compensated with a CAD $10 (US $13) electronic gift card for every survey they complete. Participants are eligible for this study if they are <32 weeks pregnant with a singleton pregnancy, living in the Calgary Zone of Alberta Health Services, and ≥16 years old.

Recruitment Platforms and Study Advertisements

Beginning in September 2021, paid advertisements targeting women between the ages of 18 and 50 years living within a 20-mile radius of Calgary with specific interests in pregnancy and parenting were used to facilitate recruitment. Facebook- and Instagram-specific targeting features included “motherhood or baby shower and parents: parents (all),” and Twitter-specific targeting features included “family and parenting- babies and toddlers, family and parenting- daycare and preschool, family and parenting- parenting K-6 kids, life stages- moms.” The terms of service of the social media platforms did not allow us to use these specific targeting features for users <18 years of age. For nearly every advertisement campaign, the appearance of study advertisements included (1) a title (ie, “Are you less than 32 weeks pregnant?”), (2) a description (ie, “Help UCalgary researchers and join our study to understand preterm birth!”), (3) an image or a graphic (ie, a pregnant person and a baby in the neonatal intensive care unit), (4) institutional logos to establish the credibility of the study (ie, the University of Calgary, the Calgary Health Foundation, and the Alberta Children’s Hospital Foundation), and (5) a link to our website.

Between September 27, 2021, and April 24, 2022, a total of 13 campaigns were run, with each advertisement being manipulated in terms of budget, duration, and content (Table 1). Each campaign involved the same advertisement running on multiple social media platforms simultaneously. Most of the advertisements contained only the essential information, while other advertisements were themed (eg, Halloween and Prematurity Awareness Month). In addition, some advertisements mentioned the study incentive of CAD $10 (US $13; although participants received the incentive for each survey completed regardless of whether or not this was mentioned in the advertisement). During this 7-month recruitment period, traditional methods (eg, posters and postcards) were distributed in the community and through health care providers to potential participants.

Table 1. Duration, budget, and image for each of the 13 advertisement campaigns used throughout recruitment.

Campaign	Duration, day	Budget (per day), CAD $ (US $)	Image
0[a]	N/A[b]	N/A	Standard[c]
1[a]	N/A	N/A	Standard
2	5	25 (32.5)	Standard
3	3	10 (13)	Halloween[d]
4	5	10 (13)	Prematurity Awareness Month[e]
5	5	35 (45.5)	Standard
6	5	50 (65)	Standard
7	3	25 (32.5)	Standard
8	3	25 (32.5)	Incentive mentioned[f]
9	5	25 (32.5)	Incentive mentioned
10	5	50 (65)	Incentive mentioned
11	8	25 (32.5)	Incentive mentioned
12	1[g]	35 (45.5)	Incentive mentioned

[a]Campaigns 0 and 1 were unpaid to address the issues associated with fraudulent participants. Unpaid advertisements are not boosted by the social media platforms and are consequently shown to fewer users. These campaigns can still mention the incentive as participants who see the advertisement will still be compensated for their participation. Campaign 0 mentioned the incentive in the caption of the campaign but not in its image.

[b]N/A: not applicable.

[c]The standard advertisement refers to a post that included a title (“Less than 32 weeks pregnant?”), a brief description of the study, our study website, a relevant cartoon, and logos of affiliated institutions. There was no mention of incentives.

[d]The Halloween advertisements compared the size of a baby at different gestational age to the Halloween candy. There were no logos of the institutions with which we were affiliated and no mention of the incentive. Information regarding the study was provided in the caption.

[e]The Prematurity Awareness Month advertisement was posted in November and included a title (“November is Prematurity Awareness Month”), an image of an infant in the neonatal intensive care unit, and the logos of the affiliated institutions. There was no mention of any incentives.

[f]The incentive advertisements were the same as the standard advertisement but mention the incentive directly in the photo of the advertisement.

[g]Campaign 12 was scheduled to run for 5 days but was discontinued after only a day owing to a high number of fraudulent responses.

Data Collection

Data regarding reach, link clicks, and cost for each campaign were collected through Facebook Ads Manager (which includes data for both Facebook and Instagram) and Twitter Analytics (Textbox 1 for definitions). Data regarding where participants learned about the study (Facebook, Twitter, Instagram, or a traditional source) and demographic data, including age, education, income, race, country of birth, and marital status, were obtained from the baseline survey to determine whether demographic characteristics varied between recruitment methods.

Textbox 1. Definitions of study outcomes.

Reach

- Number of users who saw the advertisement on their feed

Link clicks

- Number of users who clicked on the advertisement link

Click-through-rate

- Number of link clicks divided by reach

Completion to click ratio

- Number of completers divided by number of link clicks

Cost-per-click

- Total cost of the advertisement divided by the number of link clicks for each advertisement

Cost-per-completer

- Total cost of the advertisement divided by the number of people who completed the baseline survey

Program Evaluation Method

The effectiveness of Facebook, Twitter, and Instagram was evaluated based on the recruitment rate and cost-effectiveness. The analysis was broken down by each advertisement campaign, which spanned the day the campaign started to the day before the next campaign. The number of participants who consented to each campaign was used, as it was assumed that the consent date would more accurately reflect the day that the participant saw the study advertisement compared with the date that they completed the baseline survey (eg, participants could have consented days before completing the baseline survey).

Recruitment Rate

The recruitment rate was calculated using two metrics: the number of link clicks and number of completers per campaign. Using both metrics provided an understanding of whether the participants who clicked on our advertisement were doing so out of curiosity or whether they truly intended to participate in the study. The detailed breakdown provided by Facebook mobile allowed us to differentiate the reach and number of link clicks obtained through each Facebook and Instagram campaign, whereas Twitter gave this information directly. Chi-square tests were used to compare the demographic profile of individuals recruited via social media versus traditional means and to compare the profiles of individuals who were recruited across the 3 social media platforms.

Cost-effectiveness

Cost-effectiveness was measured using (1) cost-per-click and (2) cost-per-completer (Textbox 1). It was not possible to differentiate cost data between Facebook and Instagram; thus, data for these platforms were analyzed together.

Ethics Approval

Template study advertisements were approved by the University of Calgary's Conjoint Research Ethics Board (REB 20-1635) to ensure that they would not mislead participants regarding the purpose of the study.

Results

Recruitment Results

Between September 2021 and April 2022, a total of 324 participants enrolled in the ongoing P3 Cohort Study and completed the baseline survey across 13 separate advertisement campaigns. For Facebook and Instagram, 11 (85%) of the 13 campaigns were paid, and for Twitter, 5 (38%) of the 13 campaigns were paid. We reported the results of these paid campaigns. Of the 324 participants recruited, 153 (47%) heard about the study through Facebook, 79 (24%) through Instagram, 10 (3%) through Twitter, and 82 (25%) from other sources (eg, traditional methods including postcards and posters). Our paid advertisements reached 159,778 social media users that translated into 2390 link clicks.

Throughout our recruitment efforts, we encountered several issues regarding fraudulent responses. Of the 2390 link clicks, we initially had 1572 consents; however, upon further inspection, 1220 (78%) of these were deemed fraudulent, as indicated by made-up names and email addresses, as well as IP addresses outside the Calgary Zone of Alberta Health Services. Furthermore, out of 561 baseline surveys, 237 (42%) were deemed fraudulent because of inconsistencies in survey answers (eg, gestation age not matching the due date), nonsensical email addresses (eg, the email address primarily consisting of numbers or inconsistencies between a participant's name and the name in their email address), and phone numbers and IP addresses outside of Calgary. These issues were most salient during our pilot campaign (labeled campaign 0). As such, we decided to stop mentioning the incentive in our caption and discontinued the paid advertisement to stop the circulation on social media. Subsequently, we implemented new security measures, including logic checks, monitoring for fake email addresses and duplicate IP addresses, and changing the landing page such that the

website did not link directly to the consent form. Instead, when participants clicked on the link to join our study, they were required to answer some screening questions and were only then sent the consent form manually by the study team if the participant responses seemed legitimate. Screening questions asked for participant contact information (name, email address, and phone number) and basic demographic information to confirm eligibility (age, pregnancy status, gestational age, and place of residence). Members of the research team reviewed the responses and contacted participants if they were deemed eligible. These new measures led to an appreciable decrease in the rate of fraudulent responses; however, throughout the remainder of the recruitment period selected for this study, the team had to be diligent in monitoring survey response rates.

Recruitment Rate per Platform

Facebook

Link Clicks

Our paid Facebook advertisements reached 124,515 users through 11 paid campaigns, which translated into 1916 link clicks, resulting in a click-through-rate (CTR) of 1.54% (1916/12,451) and a completion to click ratio of 7.99% (153/1916; Table 2). Campaign 6 was the most successful in generating link clicks on Facebook. Campaigns 8 to 10, which mentioned the incentive directly in the image of the advertisement, also generated much traffic to our study website. Campaign 4, the prematurity awareness advertisement, and campaign 12, which was discontinued after only a day of recruitment owing to fraudulent responses, were the least successful in generating link clicks.

Table 2. Recruitment rates for paid advertisements on Facebook, Instagram, and Twitter.

Campaign	Facebook			Instagram			Twitter		
	Reach, n	Link clicks (CTR[a]), n (%)	Completed surveys, n (%[b])	Reach, n	Link clicks (CTR), n (%)	Completed surveys, n (%[b])	Reach, n	Link clicks (CTR), n (%)	Completed surveys, n (%[b])
0	__[c]	—	2	—	—	1	—	—	5
1	—	—	—	—	—	—	—	—	—
2	8698	134 (1.54)	10 (7.46)	146	2 (1.37)	1 (50)	549	3 (0.54)	—
3	6945	220 (3.16)	2 (0.91)	—	—	—	929	6 (0.65)	—
4	5313	50 (0.94)	8 (16)	2645	12 (0.45)	1 (8.33)	1798	12 (0.67)	1 (8.33)
5	10,956	167 (1.52)	14 (8.38)	288	8 (2.78)	1 (12.5)	7639	17 (0.23)	2 (11.76)
6	27,732	378 (1.36)	25 (6.61)	3416	85 (2.49)	9 (10.59)	4584	24 (0.52)	2 (8.33)
7	9016	89 (0.99)	4 (4.49)	864	15 (1.74)	5 (33.33)	__[d]	__[d]	—
8	8395	149 (1.77)	21 (14.09)	1879	69 (3.67)	14 (20.29)	__[d]	__[d]	—
9	12,147	274 (2.26)	23 (8.39)	1532	38 (2.48)	7 (18.42)	__[d]	__[d]	—
10	20,800	272 (1.31)	23 (8.46)	3480	62 (1.78)	16 (25.81)	__[d]	__[d]	—
11	12,599	154 (1.22)	6 (3.9)	5012	115 (2.29)	16 (13.91)	__[d]	__[d]	—
12	1914	29 (1.52)	15 (51.72)	502	12 (2.39)	1 (8.33)	__[d]	__[d]	—
Total	124,515	1916 (1.54)	153 (7.99)	19,764	418 (2.11)	79 (18.9)	15,499	62 (0.4)	10 (16.13)

[a]CTR: click-through-rate.

[b]Percentage refers to the ratio of the number of individuals who completed the baseline survey divided by the number of individuals who clicked on the advertisement (completer-to-click ratio).

[c]Data on these metrics were unavailable.

[d]Campaigns 7 to 12 were unpaid on Twitter; therefore, no information on reach or link clicks was collected.

Completion

Campaign 6, the advertisement with the highest budget, led to the most number of completed baseline surveys (25/153, 16.3%), whereas campaigns 8 and 9, which were the first to introduce the study's incentive into the advertisement's image and caption, were also highly successful in terms of recruitment (Table 2). Campaigns with lower budgets and duration (campaigns 3 and 7) led to the fewest number (2/153, 1.3%, and 4/153, 2.6%, respectively) of completed baseline surveys.

Instagram

Link Clicks

Our paid Instagram advertisements reached 19,764 users through the 11 paid campaigns, which translated into 418 link clicks, resulting in a CTR of 2.11% (418/19,764) and completion to click ratio of 18.9% (79/418; Table 2). As with Facebook advertisements, campaigns 6 and 8 to 10 were also the most successful in generating link clicks.

Completion

Campaigns 6, 10, and 11 led to the highest number of completed surveys (9/79, 11.4%; 16/79, 20.3%; and 16/79, 20.3%, respectively), on Instagram (Table 2). However, campaigns 2, 3, 4, 5, and 12 led to only one or no completed baseline surveys.

Twitter

Link Clicks

Our paid Twitter advertisements reached 15,499 users through the 5 paid campaigns, which translated into 62 link clicks, resulting in a CTR of 0.4% (62/15,499) and completion to click ratio of 16.13% (10/62; Table 2).

Completion

Although there was a higher proportion (10/62, 16.13%) of completers on Twitter compared with Facebook (153/1,916, 7.99%), the overall yield of the advertisements was low. Therefore, paid Twitter advertisements were discontinued after campaign 5 because of the low recruitment rate.

Cost-effectiveness

Facebook and Instagram

The total cost for Facebook and Instagram advertisements was CAD $1430 (US $1859) throughout the 11 paid campaigns. Cost-per-click for Facebook and Instagram advertisements was consistently under CAD $1 (US $1.3) across campaigns, with an average cost-per-click of CAD $0.65 (US $0.84; SD $0.27; US $0.35) throughout the 11 campaigns. Campaign 3 saw the lowest cost-per-click at CAD $0.14 (US $0.18), whereas the highest cost-per-click was for campaigns 4 and 5, both at CAD $1 (US $1.3; Table 3).

The cost-per-completer on Facebook and Instagram remained <CAD $15 (US $19.5; Table 3). Campaign 12, which mentioned the incentive and had the second highest budget per day, saw the lowest cost-per-completer at CAD $1.44 (US $1.87) per participant. Overall, the average cost-per-completer was CAD $7.89 (US $10.25; SD $4.08, US $5.30).

Table 3. Cost-per-click and cost-per-completer for Facebook and Instagram (combined) and Twitter.

Campaign	Facebook and Instagram		Twitter	
	Cost-per-click, CAD $ (US$)	Cost-per-completer, CAD $ (US$)	Cost-per-click, CAD $ (US$)	Cost-per-completer, CAD $ (US$)
0[a]	—[b]	—[b]	—[b]	0.00[c]
1[a]	—[b]	—[b]	—[b]	—[b]
2	0.92 (1.19)	11.36 (14.7)	11.71 (15.22)	—[b]
3	0.14 (0.18)	15.00 (19.5)	4.18 (5.43)	—[b]
4	1.00 (1.3)	11.11 (14.44)	—[d]	0.00[c]
5	1.00 (1.3)	11.67 (15.08)	9.49 (12.33)	69.00 (89.7)
6	0.54 (0.70)	6.10 (7.93)	4.63 (6.01)	55.61 (72.2)
7	0.72 (0.93)	8.33 (10.8)	—[e]	—[b]
8	0.34 (0.44)	2.14 (2.78)	—[e]	—[b]
9	0.40 (0.52)	4.17 (5.41)	—[e]	—[b]
10	0.75 (0.97)	6.41 (5.11)	—[e]	—[b]
11	0.74 (0.96)	9.08 (11.8)	—[e]	—[b]
12	0.57 (0.74)	1.44 (1.87)	—[e]	—[b]

[a]Campaigns 0 and 1 were unpaid across all platforms.

[b]Data not available.

[c]Twitter campaigns 0 and 4 recruited a participant despite not using any budget, so they were likely to have heard about the advertisement through organic methods.

[d]Twitter did not use any of our budget for campaign 4.

[e]Campaigns 7 to 12 were unpaid on Twitter.

Twitter

The total cost of Twitter advertisements throughout the 5 paid campaigns was CAD $299 (US $388.7). Depending on the campaign, only some (or sometimes none) of the prespecified budget was used by Twitter to display our advertisements.

Cost-per-click for the advertisements that used a partial or full amount of the prespecified budget ranged from CAD $4.18 (US $4,72; campaign 3) to CAD $11.71 (US $15.22; campaign 2), with an average cost-per-click of CAD $7.50 (US $9.75; SD $3.20, US $4.16; Table 3). Cost-per-completer for paid Twitter advertisements ranged from CAD $55.61 (US $72.29; campaign 6) to CAD $69 (US $89.7; campaign 5).

Representativeness

The demographic profiles of individuals recruited via social media did not differ from those recruited through traditional means in terms of education, household income, marital status, immigration status, race, or age (Table 4). However, among individuals who were recruited via social media, those with educational backgrounds of university graduation and above were more likely to be recruited via Instagram or Twitter (P<.001) and those with a higher household income (above CAD $100,000/year; US $13,000) were also more likely to be recruited through Instagram or Twitter (P<.001).

Table 4. Demographic characteristics of the participants recruited through Facebook, Instagram, Twitter, and other sources (N=307)[a].

Characteristics	Facebook, n (%)	Instagram, n (%)	Twitter, n (%)	Other sources, n (%)	*P* value	
					For differences between 3 social media sources	For differences between social media and other sources
Education, (n=306[a])					<.001	.53
Did not graduate university	26 (18)	7 (10)	0 (0)	9 (11)		
Graduated university	92 (64)	44 (60)	2 (20)	47 (59)		
Graduated school	25 (17.48)	22 (30)	8 (80)	24 (30)		
Income in CAD $, (n=303[a])					<.001	.54
<99,999 (US $129,998)	49 (35)	10 (14)	0 (0)	24 (30)		
≥99,999 (US $129,998)	91 (65)	63 (86)	10 (100)	56 (70)		
Marital status, (n=307[a])					.57	.34
Married or common law	132 (92.31)	69 (95)	10 (100)	78 (96)		
Other	11 (7.69)	4 (5)	0 (0)	3 (4)		
Country of origin, (n=307[a])					.75	.20
Canada	121 (85)	60 (82)	9 (100)	63 (78)		
Other	22 (15)	13 (18)	1 (10)	18 (22)		
Race, (n=306[a])					.17	.36
White	107 (75)	53 (73)	10 (100)	57 (70)		
Other	35 (25)	20 (27)	0 (0)	24 (30)		
Age in years, (n=300[a])					.31	.64
<35	100 (71)	44 (61)	6 (60)	54 (70)		
≥35	41 (29)	28 (39)	4 (40)	23 (30)		

[a]Owing to item nonresponse, the total of the n values does not add up to 324.

Discussion

Principal Findings

Using social media as a recruitment strategy proved to be an effective method to reach and recruit a sample of pregnant women in our longitudinal study although the recruitment rate and cost-effectiveness did vary by platform. Facebook and Instagram were highly effective in generating traffic in our study survey. The CTR of 1.54% (1916/124,515) for Facebook and 2.11% (418/19,764) for Instagram were consistent with previous literature on this topic, with most studies finding a CTR of approximately 2% for Facebook advertisements [1,2,8,9]. The higher CTR on Instagram can likely be attributed to the fact that Instagram is more commonly used by our target demographic (especially among women aged 18-29 years [10]), while the success garnered by Facebook advertisements is likely related to the regular use of this platform by pregnant women to connect with other pregnant women and to find answers to pregnancy- or parenting-related questions [1,5,11]. Twitter was much less effective in generating traffic to our website and recruiting participants, despite being commonly used by younger and middle-aged women [10]. A possible explanation is that the Twitter algorithm is less effective at targeting and reaching the population of interest. Our Twitter account was contacted mainly by other researchers as opposed to accounts related to our target population.

The campaigns with the most success in terms of link clicks and completed surveys were those with either a high budget (campaign 6) or those that mentioned the study's incentive directly in the image of the advertisement (campaigns 8-11). Campaign 12 was an exception because the advertisement had to be discontinued after only a day to a high influx of fraudulent

responses, thus generating fewer link clicks. It is important to note that although campaign 3 had a high number of link clicks even with a low budget, the advertisement was mistakenly targeted to all of Canada, likely resulting in many link clicks and few completed surveys. Interestingly, campaign 4 was less successful in terms of generating link clicks and completed surveys, demonstrating that standard advertisements were more successful than themed ones. Advertisements posted for a shorter duration (3 days rather than 5 days) and without the incentive were also less successful in generating link clicks and completed surveys (campaign 7). The most successful advertisements contained a short title and description of the study, the logos of affiliated institutions, the incentive of the study, the study website link, and a relevant image or cartoon.

Paid Facebook and Instagram advertisements proved to be cost-effective, with an average cost-per-click of CAD $0.65 (US $0.84; SD $0.27; US $0.35) and an average cost-per-completer of CAD $7.89 (US $10.25; SD $4.08, US $5.30). This cost-per-completer was lower than what has been found in previous pregnancy studies that have used social media for recruitment [1,12-14]. The cost-per-completer in these studies ranged from CAD $14.63 (US $19.01) [14] to CAD $51.27 (US $66.65) [13]. The average cost-per-click of CAD $0.65 (US $0.84; SD $0.27, US $0.35) was similar to findings in other pregnancy studies [1,12]. Overall, Facebook and Instagram advertisements provided a cost-effective method to reach our target population, especially in comparison to the cost normally incurred by traditional methods where researchers must spend time and money designing, printing, and distributing posters and brochures [1,14]. An important consideration, however, is that it was quite time-consuming for our study team to monitor and sort through fraudulent participants. Paid advertisements on Twitter were less cost-effective than Facebook and Instagram advertisements, with an average cost-per-click of CAD $7.50 (US $9.75; SD $3.20, US $4.16) and cost-per-completer ranging from CAD $55.61 (US $72.2) to CAD $69 (US $89.7). Focusing on using organic (ie, unpaid) recruitment methods rather than paid advertisements on Twitter may help overcome some of the challenges we faced when using this platform.

Overall, the characteristics of the participants recruited in this study reflected those of participants in other studies that focused on pregnancy or other studies examining the effectiveness of social media recruitment [1,2,9,15,16]. Our sample mainly comprised highly educated and higher-income White women born in Canada. On the basis of the results of our analysis, the demographic characteristics of the participants recruited through social media did not differ significantly from those of participants recruited through traditional methods. Therefore, the use of social media for recruitment will not leave out important demographics that would otherwise have been obtained using only traditional methods. In addition, we found that all 3 social media platforms recruited participants with similar demographic characteristics; however, Facebook was more effective than Twitter and Instagram in recruiting individuals with lower education and income levels. This might be related to the fact that our Twitter page was primarily followed by researchers, although there is less obvious explanation for Instagram. Importantly, however, in our overall sample, individuals with lower levels of education and income remained underrepresented.

An increasing number of studies using social media for recruitment have reported issues regarding fraudulent responses. One study that used Facebook and Twitter for recruitment to examine patient perceptions of patient-provider communication in the ovarian cancer care setting found that most of their survey respondents were illegitimate [17]. They suggested indicators of low-quality data, including evidence of inattention (completing the survey in an unrealistic amount of time), duplicate or unusual responses to open-ended survey items, inconsistent responses to verifiable items (eg, the location of the survey respondent and the time zone not matching), and evidence of bot automation [17]. Similar indicators were used to flag fraudulent participants in this study. Studies suggest that web-based private servers and server farms may be at the root of this issue, as they allow one individual to complete many surveys simultaneously, each with a unique IP address, purely for financial gain [17,18]. Consequently, research funds meant to compensate legitimate study participants are used for bot responses [18]. As IP addresses are not tied to physical locations, but rather are assigned by internet service providers when users access the internet, they can be manipulated, and IP addresses alone cannot be used to identify real versus fraudulent responses. Many other strategies exist to mitigate the fraudulent responses, including lowering the value of the incentive, collecting verifiable information (phone number), sending each respondent a unique survey link, having items that can be compared for consistency, capturing time stamps for start and stop times, requiring open-text questions, or limiting the visibility of the advertisement on social media platforms [17-21]. In addition, having participants check a box stating that they understand that ineligible responses will not receive the incentive and that researchers may contact them by phone to confirm eligibility may drive illegitimate participants away from the study [18]. Ultimately, researchers must find a balance between making the study easy and convenient for legitimate respondents to participate and establishing sufficient security measures to preserve the integrity of the data [21]. Too many steps (ie, screening questions, CAPTCHA [Completely Automated Public Turing test to tell Computers and Humans Apart] tests, etc) may cause frustration among real participants and lead to a lower recruitment rate. More research is needed in this area to determine the best approach for mitigating fraudulent responses while maintaining the convenience and cost-effectiveness of social media recruitment.

Strengths and Limitations

This study has both strengths and limitations. To the best of our knowledge, this is one of the first studies to directly compare the ability of multiple social media platforms to recruit participants. This provides valuable information that other researchers can use to help determine the best recruitment source for their needs and to accurately estimate their budget. This study also has several limitations. First, owing to the nature of our study design, we were unable to know where consenting participants heard about the study. This metric would have allowed us to determine whether participants who clicked on

our website and consented followed through with participating in the study (or if they just clicked on the advertisement out of curiosity) based on which platform they heard about the study. In addition, we could not determine whether there was a major difference in the cost-per-consent and cost-per-completer. A solution to this would be to include a question asking where they heard about the study on the consent form as opposed to only on the baseline survey. Evaluation of social media as a recruitment method can be challenging owing to the ever-changing policies and algorithms, making it difficult to maintain consistency over a long recruitment period and to compare with other studies of a similar nature [2]. The potential for recall bias should also be considered, as participants may have seen the study advertisement in many places or forgotten how they heard about the study, thereby influencing our results. Similar to other prospective pregnancy cohort studies, this study underrepresented individuals with lower education, lower income, and racial minorities. This poses a problem in generalizing the findings of the broader cohort study to the general population. To counter this, studies have suggested targeting advertisements to lower-income postal codes or neighborhoods within a city or targeting advertisements to the interests of minority populations [22,23]. Future research should focus on these strategies. By focusing on social media strategies aimed at recruiting underrepresented populations, studies could identify common factors of pregnancy in these populations that make them susceptible to preterm birth. This could ultimately lead to more targeted interventions to reduce health disparities within the community. In addition, the advertisements in this study were specifically targeted at social media users with an interest in pregnancy and parenting (because the social media algorithm did not permit us to directly target advertisements to pregnant women). Therefore, our sample likely consists of individuals who either discuss their pregnancy openly on social media or individuals who like and follow pregnancy-related pages. Future research could investigate the effectiveness of generic advertisements (ie, advertisements targeted at all women) compared with more targeted advertisements such as those used in this study. Finally, the advertisements and content used to promote this study were targeted toward women; however, we recognize that not all individuals who are pregnant identify themselves as women. In addition, much of the content focused on the positive and exciting aspects of pregnancy; however, pregnancy and the transition to parenthood can be challenging. Future content can aim to overcome this by being more inclusive of different gender identities and experiences of pregnancy.

Conclusions

In conclusion, the results of this study demonstrate that social media is a feasible and cost-effective way to reach and recruit pregnant women to a longitudinal study. Paid advertisements on Facebook and Instagram, specifically, were highly practical and cost-effective methods of reaching and recruiting participants. Researchers can turn to this work to gain an understanding of what to expect in terms of recruitment rate, cost, and representativeness when using social media to recruit pregnant women or other populations. However, researchers should be aware of the potential fraudulent responses and identify mitigation strategies for such issues. With ever-changing technology and the competitive nature of obtaining research funding, researchers should use social media to their advantage as an effective and low-cost means of recruitment. Ultimately, this work feeds into the broader P3 Cohort Study and is the first step toward gaining a better understanding of preterm births. As of April 2022, the P3 Cohort Study is ongoing, and, as such, the findings from this study will help inform our recruitment strategy going forward. Recruiting an appropriate number of participants is crucial to a study of this nature; therefore, the findings of this kind of work cannot be overlooked.

Acknowledgments

The P3 Cohort Study was jointly funded by the Alberta Children's Hospital Foundation and the Calgary Health Foundation. We would also like to acknowledge the entire P3 Cohort team and participants for their contributions to this study.

Conflicts of Interest

None declared.

References

1. Adam LM, Manca DP, Bell RC. Can Facebook be used for research? Experiences using Facebook to recruit pregnant women for a randomized controlled trial. J Med Internet Res 2016 Sep 21;18(9):e250 [FREE Full text] [doi: 10.2196/jmir.6404] [Medline: 27655184]
2. Bennetts SK, Hokke S, Crawford S, Hackworth NJ, Leach LS, Nguyen C, et al. Using paid and free Facebook methods to recruit Australian parents to an online survey: an evaluation. J Med Internet Res 2019 Mar 06;21(3):e11206 [FREE Full text] [doi: 10.2196/11206] [Medline: 30839282]
3. Statista Research Department. Social media usage in Canada - Statistics and Facts. Statista. 2021 Aug 25. URL: https://www.statista.com/topics/2729/social-networking-in-canada/#dossierKeyfigures [accessed 2021-11-21]
4. Powell R, Rosenthal J, August EM, Frey M, Garcia L, Sidibe T, et al. Ante la duda, pregunta: a social marketing campaign to improve contraceptive access during a public health emergency. Health Commun 2022 Feb;37(2):177-184 [FREE Full text] [doi: 10.1080/10410236.2020.1828534] [Medline: 33016136]
5. Shere M, Zhao XY, Koren G. The role of social media in recruiting for clinical trials in pregnancy. PLoS One 2014;9(3):e92744 [FREE Full text] [doi: 10.1371/journal.pone.0092744] [Medline: 24671210]

6. Moseson H, Wollum A, Seymour J, Zuniga C, Thompson TA, Gerdts C. Comparison of Facebook, Google Ads, and Reddit for the recruitment of people who considered but did not obtain abortion care in the United States: cross-sectional survey. JMIR Form Res 2021 Feb 24;5(2):e22854 [FREE Full text] [doi: 10.2196/22854] [Medline: 33625368]
7. P3 Cohort. 2021. URL: https://p3cohort.ca/ [accessed 2022-03-21]
8. Graham JE, Moore JL, Bell RC, Miller T. Digital marketing to promote healthy weight gain among pregnant women in Alberta: an implementation study. J Med Internet Res 2019 Feb 01;21(2):e11534 [FREE Full text] [doi: 10.2196/11534] [Medline: 30707100]
9. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young Australian women. Am J Epidemiol 2015 May 15;181(10):737-746. [doi: 10.1093/aje/kwv006] [Medline: 25883155]
10. Auxier B, Anderson M. Social Media Use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2022-04-06]
11. Baker B, Yang I. Social media as social support in pregnancy and the postpartum. Sex Reprod Healthc 2018 Oct;17:31-34. [doi: 10.1016/j.srhc.2018.05.003] [Medline: 30193717]
12. Arcia A. Facebook advertisements for inexpensive participant recruitment among women in early pregnancy. Health Educ Behav 2014 Jun;41(3):237-241 [FREE Full text] [doi: 10.1177/1090198113504414] [Medline: 24082026]
13. Guillory J, Jordan A, Paquin RS, Pikowski J, McInnis S, Anakaraonye A, et al. Using social media to conduct outreach and recruitment for expanded newborn screening. Front Commun 2020 May 6;5(21):1-11 [FREE Full text] [doi: 10.3389/fcomm.2020.00021]
14. Admon L, Haefner JK, Kolenic GE, Chang T, Davis MM, Moniz MH. Recruiting pregnant patients for survey research: a head to head comparison of social media-based versus clinic-based approaches. J Med Internet Res 2016 Dec 21;18(12):e326 [FREE Full text] [doi: 10.2196/jmir.6593] [Medline: 28003174]
15. Loxton D, Powers J, Anderson A, Townsend N, Harris ML, Tuckerman R, et al. Online and offline recruitment of young women for a longitudinal health survey: findings from the Australian longitudinal study on women's health 1989-95 cohort. J Med Internet Res 2015 May 04;17(5):e109 [FREE Full text] [doi: 10.2196/jmir.4261] [Medline: 25940876]
16. van Gelder MM, van de Belt TH, Engelen LJ, Hooijer R, Bredie SJ, Roeleveld N. Google AdWords and Facebook ads for recruitment of pregnant women into a prospective cohort study with long-term follow-up. Matern Child Health J 2019 Oct;23(10):1285-1291 [FREE Full text] [doi: 10.1007/s10995-019-02797-2] [Medline: 31222599]
17. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020 Oct 07;22(10):e23021 [FREE Full text] [doi: 10.2196/23021] [Medline: 33026360]
18. Griffin M, Martino RJ, LoSchiavo C, Comer-Carruthers C, Krause KD, Stults CB, et al. Ensuring survey research data integrity in the era of internet bots. Qual Quant 2022;56(4):2841-2852 [FREE Full text] [doi: 10.1007/s11135-021-01252-1] [Medline: 34629553]
19. Pratt-Chapman M, Moses J, Arem H. Strategies for the identification and prevention of survey fraud: data analysis of a web-based survey. JMIR Cancer 2021 Jul 16;7(3):e30730 [FREE Full text] [doi: 10.2196/30730] [Medline: 34269685]
20. Dewitt J, Capistrant B, Kohli N, Rosser BR, Mitteldorf D, Merengwa E, et al. Addressing participant validity in a small internet health survey (The Restore Study): protocol and recommendations for survey response validation. JMIR Res Protoc 2018 Apr 24;7(4):e96 [FREE Full text] [doi: 10.2196/resprot.7655] [Medline: 29691203]
21. Salinas MR. Are your participants real? Dealing with fraud in recruiting older adults online. West J Nurs Res (forthcoming) 2022 May 19:1939459221098468. [doi: 10.1177/01939459221098468] [Medline: 35587721]
22. Pechmann C, Phillips C, Calder D, Prochaska JJ. Facebook recruitment using zip codes to improve diversity in health research: longitudinal observational study. J Med Internet Res 2020 Jun 05;22(6):e17554 [FREE Full text] [doi: 10.2196/17554] [Medline: 32501274]
23. Remschmidt C, Walter D, Schmich P, Wetzstein M, Deleré Y, Wichmann O. Knowledge, attitude, and uptake related to human papillomavirus vaccination among young women in Germany recruited via a social media site. Hum Vaccin Immunother 2014;10(9):2527-2535 [FREE Full text] [doi: 10.4161/21645515.2014.970920] [Medline: 25483492]

Abbreviations

CTR: click-through-rate

Edited by R Kukafka; submitted 17.06.22; peer-reviewed by K Reuter, M Waring, R Pozzar; comments to author 06.08.22; revised version received 30.09.22; accepted 25.10.22; published 12.12.22.

Please cite as:
Pekarsky C, Skiffington J, Leijser LM, Slater D, Metcalfe A
Social Media Recruitment Strategies to Recruit Pregnant Women Into a Longitudinal Observational Cohort Study: Usability Study
J Med Internet Res 2022;24(12):e40298
URL: https://www.jmir.org/2022/12/e40298
doi:10.2196/40298
PMID:36508244

Original Paper

Challenges in Recruiting University Students for Web-Based Indicated Prevention of Depression and Anxiety: Results From a Randomized Controlled Trial (ICare Prevent)

Felix Bolinski[1,2], MSc; Annet Kleiboer[1], PhD; Koen Neijenhuijs[3], PhD; Eirini Karyotaki[1], PhD; Reinout Wiers[4], PhD; Lisa de Koning[4], MSc; Corinna Jacobi[5], PhD; Anna-Carlotta Zarski[6], PhD; Kiona K Weisel[7], PhD; Pim Cuijpers[1,8], PhD; Heleen Riper[1,9], PhD

[1]Department of Clinical, Neuro, and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Department of Mental Health and Prevention, Trimbos Institute, Netherlands Institute for Mental Health and Addiction, Utrecht, Netherlands

[3]Department of Research and Business Development, HumanTotalCare, Utrecht, Netherlands

[4]Department of Psychology and Centre for Urban Mental Health, University of Amsterdam, Amsterdam, Netherlands

[5]Department of Clinical Psychology and Psychotherapy, Technical University Dresden, Dresden, Germany

[6]Department of Sport and Health Science, Technical University Munich, Munich, Germany

[7]Department of Clinical Psychology and Psychotherapy, Friedrich Alexander University, Erlangen-Nuremberg, Germany

[8]International Institute for Psychotherapy, Babe -Bolyai University, Cluj-Napoca, Romania

[9]Department of Psychiatry, Amsterdam UMC, location VUmc, Amsterdam, Netherlands

Corresponding Author:
Felix Bolinski, MSc
Department of Clinical, Neuro, and Developmental Psychology
Amsterdam Public Health Research Institute
Vrije Universiteit Amsterdam
De Boelelaan 1105
Amsterdam, 1081HV
Netherlands
Phone: 31 0205985204
Email: f.bolinski@vu.nl

Abstract

Background: Depression and anxiety are common mental health conditions in college and university student populations. Offering transdiagnostic, web-based prevention programs such as ICare Prevent to those with subclinical complaints has the potential to reduce some barriers to receiving help (eg, availability of services, privacy considerations, and students' desire for autonomy). However, uptake of these interventions is often low, and accounts of recruitment challenges are needed to complement available effectiveness research in student populations.

Objective: The aims of this study were to describe recruitment challenges together with effective recruitment strategies for ICare Prevent and provide basic information on the intervention's effectiveness.

Methods: A 3-arm randomized controlled trial was conducted in a student sample with subclinical symptoms of depression and anxiety on the effectiveness of an individually guided (human support and feedback on exercises provided after each session, tailored to each participant) and automatically guided (computer-generated messages provided after each session, geared toward motivation) version of ICare Prevent, a web-based intervention with transdiagnostic components for the indicated prevention of depression and anxiety. The intervention was compared with care as usual. Descriptive statistics were used to outline recruitment challenges and effective web-based and offline strategies as well as students' use of the intervention. A basic analysis of intervention effects was conducted using a Bayesian linear mixed model, with Bayes factors reported as the effect size.

Results: Direct recruitment through students' email addresses via the central student administration was the most effective strategy. Data from 35 participants were analyzed (individually guided: n=14, 40%; automatically guided: n=8, 23%; care as usual: n=13, 37%). Use of the intervention was low, with an average of 3 out of 7 sessions (SD 2.9) completed. The analyses did not suggest any intervention effects other than anecdotal evidence (all Bayes factors$_{10}$≤2.7).

Conclusions: This report adds to the existing literature on recruitment challenges specific to the student population. Testing the feasibility of recruitment measures and the greater involvement of the target population in their design, as well as shifting from direct to indirect prevention, can potentially help future studies in the field. In addition, this report demonstrates an alternative basic analytical strategy for underpowered randomized controlled trials.

Trial Registration: International Clinical Trials Registry Platform NTR6562; https://tinyurl.com/4rbexzrk

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-018-2477-y

(J Med Internet Res 2022;24(12):e40892) doi:10.2196/40892

KEYWORDS

digital mental health; students; indicated prevention; recruitment; randomized trial; mobile phone

Introduction

Background

Offering transdiagnostic prevention programs for depression and anxiety to college and university students (henceforth denoted as students) is relevant for a number of reasons: these conditions are highly comorbid [1,2] and often emerge during the time individuals enter tertiary education [3-6]. Moreover, symptoms of depression and anxiety are highly correlated with various stressors specifically affecting students, such as considerable academic demands [7,8]; balancing life, studies, and student jobs [9,10]; and, more recently, the global COVID-19 pandemic and the measures implemented to reduce its spread [11,12]. In addition, some studies have reported that financial concerns, in particular student debt, might negatively affect students' mental health [13,14].

Results from the World Health Organization World Mental Health International College Student initiative [15] suggest that approximately one-third of the almost 14,000 surveyed students had had a mental health disorder in the previous year, with major depressive disorder (MDD; 18.5%) and generalized anxiety disorder (GAD; 16.7%) being the most prevalent conditions [5]. Both disorders cause adverse effects on quality of life [16] and can lead to severe interpersonal impairment. For example, research indicates that more than half of the first-year students who presented with depression or anxiety reported severe disturbances in their social lives and close relationships [17]. In addition, mental health complaints negatively affect students' academic performance and increase dropout [18,19]. A recent study among Dutch adolescents also highlighted the economic costs of subclinical depression, including health care and societal costs related to school absenteeism [20]. Considering the importance of higher education for economic growth [21], preventing the onset of mental health conditions in students matters from an individual, societal, and economic perspective.

In particular, indicated prevention approaches [22] that focus on individuals with subclinical complaints could be beneficial. These have the potential to prevent the onset of, for example, MDD [23] and—in contrast to universal approaches—reduce expenditure on costly treatment by allocating scarce health care resources to those students who are in immediate need [24]. Meta-analyses of community samples suggest that preventive interventions can reduce the incidence rate of depression by approximately 19% [25] and of anxiety by approximately 43% [26].

Despite the need for and availability of such interventions [27,28], uptake of preventive programs is particularly low compared with treatment for psychological disorders [29,30], likely as perceived symptoms are not severe enough yet to motivate help seeking. Students in particular often do not seek professional help—merely approximately a quarter of World Health Organization International College Student Initiative respondents indicated that they would definitely seek treatment for mental health complaints; of those who would not seek help, approximately half indicated that they would rather deal with the problem on their own or preferred to consult with friends or relatives [31]. Fear of stigmatization is another reason for low help-seeking behavior [32,33]. In addition, on-campus mental health services often lack the capacity to meet the needs of those students who seek help [34]. Moreover, often these services focus exclusively on study-related issues (eg, test anxiety and procrastination). As a result, studies suggest that only approximately one-third to one-sixth of students receive adequate help for their complaints [35,36].

Digital interventions provided via the internet on computers or via smartphone apps have been proposed as a way to overcome barriers to the availability of counseling services and help seeking by offering the privacy and autonomy desired by students [37]. A recent meta-analysis on the effectiveness of such interventions for mental health complaints included 48 randomized controlled trials (RCTs) on individually guided (ie, tailored human support) and unguided or automatically guided (ie, computer-generated standardized feedback) digital interventions, most of them based on cognitive behavioral therapy [28]. Small but significant differences favoring the interventions were found only when these were compared with passive controls (eg, depression: g=0.18; anxiety: g=0.27), with individual guidance having no significant effect on these results. However, slightly larger effects on depression outcomes (g=0.29) were found in studies that targeted subclinical complaints. Importantly, the prediction intervals in this meta-analysis suggest that future trials will likely include nil effects [28]. In addition, transdiagnostic components that target both conditions may be beneficial, particularly for depression outcomes (g=0.22) [38]. Although this indicates some potential for indicated transdiagnostic prevention efforts, the general focus on reduction of symptoms rather than prevention of (future) mental health conditions implies a need for further research [39]. However, an RCT has shown that a digital intervention can also prevent the onset of MDD in the general population [23].

Objectives

On the basis of these considerations, ICare Prevent—a transdiagnostic individually tailored digital intervention for the indicated prevention of depression and anxiety—was developed. We planned to conduct an RCT on the effectiveness of the intervention among students in the Netherlands. However, despite 2 years of extensive countrywide recruitment efforts, this trial was concluded without reaching the targeted sample size (N=252). Therefore, the aims of this study are 3-fold, namely to (1) describe the recruitment process for the RCT, (2) describe participants' use of the intervention, and (3) conduct a basic analysis of intervention effects on depressive and anxiety symptoms.

Methods

General Study Design and Inclusion and Exclusion Criteria

This trial was registered in the International Clinical Trials Registry Platform (NTR6562), and a detailed protocol for the planned trial has been published [40]. In summary, the design entailed a 3-arm parallel superiority trial comparing an individually guided and an automatically guided version of the intervention with care as usual (CAU). The aim was to include 252 students (84 per condition; expected effect size: d=0.35; for the power calculation, see the protocol by Bolinski et al [40]). Recruitment started in June 2017 and was concluded in July 2019.

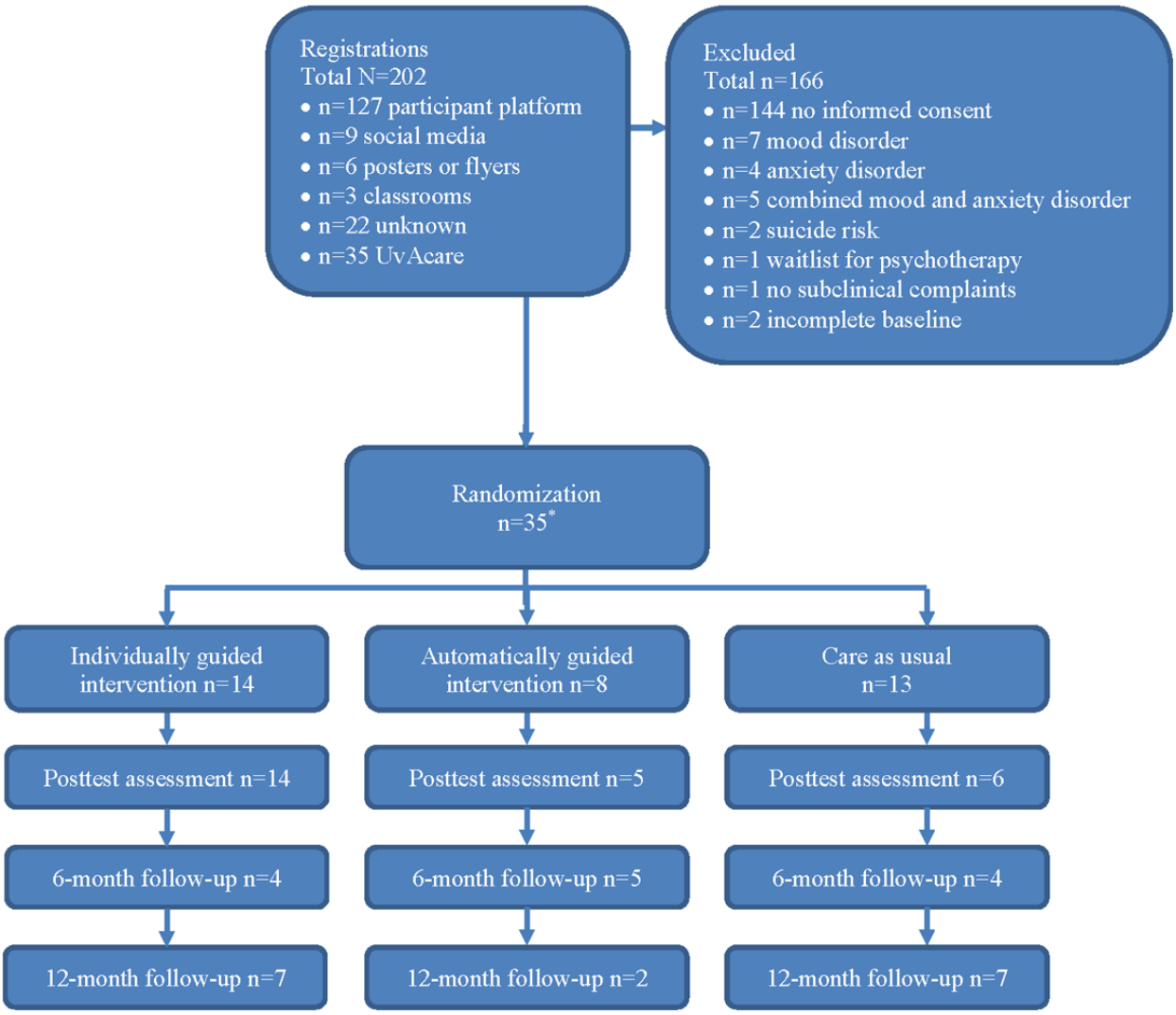
Dutch- and English-speaking students aged ≥16 years with subclinical symptoms of depression or anxiety, defined as a score of ≥16 on the Center for Epidemiological Studies Depression Scale [41] or ≥5 on the 7-item Generalized Anxiety Disorder Scale (GAD-7) [42], respectively, could participate if they provided written informed consent. They were excluded if they were in remission of a major depressive episode; currently meeting diagnostic criteria for a mood or anxiety, lifetime bipolar, or psychotic disorder; or presenting moderate to high suicide risk (all assessed by phone using the Mini International Neuropsychiatric Interview [MINI]) [43]. In addition, they were excluded if they reported currently being on a waitlist for or having received psychotherapy in the previous half year as well as participating in similar intervention studies at the time of the screening assessment.

Randomization was performed at the individual level at a 1:1:1 allocation ratio following baseline assessment. It was stratified by the type of subclinical symptoms (ie, depression, anxiety, or both according to the MINI). Subsequent assessment points were the midway point (5 weeks after randomization), posttest assessment (8 weeks after randomization), and the 6- and 12-month follow-ups.

Procedure

Participants first received information on the study. Following registration and provision of written informed consent, self-reported inclusion and exclusion criteria were assessed through web-based questionnaires, followed by the assessment of diagnostic exclusion criteria by phone. Participants were then randomized according to incoming informed consent forms by an independent researcher using a dedicated randomization platform. Owing to the nature of the study, blinding of participants and of the coaches who provided guidance in the individually guided intervention arm was not possible. However, assessors performing the MINI interviews at the posttest and follow-up assessments were blind to the participants' group allocation (see the flowchart in Figure 1).

Figure 1. Study flowchart (*=1 participant excluded because of dropout after randomization).



Recruitment Methods

A broad array of recruitment measures was used, such as social media campaigns, printed advertisements, a paid participant platform, and other project collaborations. Examples of some of the recruitment materials are provided in Multimedia Appendices 1 and 2.

Social Media

Paid social media campaigns on Facebook and Instagram were targeted to higher-education students in the Netherlands. These advertisements contained images that reflected the study aim, a brief outline of the web-based intervention, and a call to visit the study website and register for participation. In addition, information on the study and the call to participate were posted on Facebook groups relevant to students.

Print and News Media

Flyers, posters, and stickers were distributed at universities countrywide. Similar to social media posts, they contained brief information on the study and the intervention as well as contact details (study website and email address). This material was also published in 2 of the largest student newspapers.

Targeted Email Distribution Through Project Collaboration (UvAcare)

In March 2019, recruitment was extended to a student mental health project conducted at another university (UvAcare) [44]. Therein, all students were screened for mental health complaints, including those with more severe presentations of such problems. Screening generally took place at the beginning of a semester. A data-sharing agreement was set up, and data on demographic variables and primary mental health outcomes of a subsample of participating (PhD) students fulfilling the aforementioned inclusion criteria were provided. This collaboration offered the possibility of using students' email addresses—sent through the central administration—for providing information on the study and the intervention and a digital rather than written informed consent form. In this project collaboration, the use of email addresses remained the primary recruitment channel. A full description of this project is provided in the associated study protocol [44].

Participant Platform

A paid platform for participation in clinical trials was used [45]. The platform maintains a directory of users who are interested in participating in clinical trials. Information on the study was

uploaded to the platform, and interested users could sign up to receive the information letter and informed consent form. A brief questionnaire on whether a user was currently registered at a Dutch higher education institution and not currently receiving psychological treatment was implemented as an initial screener.

Other

Key individuals at the university (eg, student bodies, teachers, and counselors) were approached to create support for the study. It was then pitched in classrooms and web-based education portals. Awareness of the project was also created by participating in the largest running event in the Netherlands and wearing a shirt printed with the study website [46]. Local and international conferences were used to generate a network of contacts that could help in the dissemination of project information and, thus, recruitment.

Intervention

ICare Prevent is a web-based and mobile-supported intervention for the indicated prevention of depression and anxiety. It uses both transdiagnostic and individually tailored elements and was originally developed for the German-speaking general population following existing evidence- and cognitive behavioral therapy–based web-based modules from different digital interventions [23,47]. The intervention comprises a sequential (ie, a session has to be completed to unlock the next) 7-session web-based program with 1 booster session. Participants were advised to complete between 1 and 2 sessions per week, with each requiring approximately 45 to 60 minutes to be completed. In addition, 8 elective modules (on sleep, perfectionism, gratitude, self-esteem, alcohol use, relaxation, acceptance, and rumination) and 5 diaries (positive activities, negative thoughts, sleep, challenging situations, and alcohol use) targeted factors common to both mood and anxiety problems. All the sessions and elective modules included individual exercises that needed to be completed. In line with its transdiagnostic, individually tailored approach, users received information targeting both conditions (sessions 1-4) before they could prioritize the techniques that focused on their most prominent complaints (sessions 5 and 6).

In the individually guided arm, clinical psychology students motivated participants to continue with the sessions and provided structured and personalized feedback on the homework exercises. Those allocated to the automatically guided intervention received standard and computerized feedback after completing each session, which was geared toward motivating them to continue to the next session. Technical and usability questions could be asked by participants in either group.

As part of the study, the intervention was translated and adapted to the student context in the Netherlands. To do this, the first author worked together with a graphic designer and text editor experienced in web editing. Finally, the intervention was tested on spelling and functionality by a number of students who were employed as research assistants. Adaptations included changing the testimonials to be more representative of a diverse student population, changing the focus of exercises to problems more applicable to students, and considerably shortening the amount of text in the modules while keeping the structure and main content of the original intervention intact (see Multimedia Appendix 3 and the protocol by Bolinski et al [40] for a description of the intervention content per session and its transdiagnostic components). Students in the CAU condition were informed that they could seek help from their general practitioner (GP) or student psychologist for any mental health complaints. However, students in all trial arms were free to access CAU during the study.

Primary Outcome Variables

In total, 2 primary outcome variables were assessed by interviewers via telephone to measure changes in disorder-specific symptoms from baseline to posttest and follow-up assessments [40]. For depression, the clinician-rated version of the Quick Inventory of Depressive Symptomatology (QIDS-CR) [48] was used. This 16-item instrument assesses 9 symptom criteria for depression (sad mood, concentration difficulties, self-criticism, suicidal thoughts, general interest, fatigue, sleep, appetite, and psychomotor retardation) based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [49], with total scores ranging from 0 to 27. Higher scores indicate greater psychopathology. The QIDS-CR has well-established psychometric properties and is sensitive to the effects of treatment [50,51]. It has been shown to generate reliable scores when administered over the phone compared with in-person or paper-and-pencil versions [52]. Anxiety was assessed using the Structured Interview Guide for the Hamilton Anxiety Rating Scale (SIGH-A) [53,54]. A total of 14 items cover anxiety symptoms, resulting in a total score ranging from 0 to 56, with higher scores again indicating greater psychopathology. The SIGH-A has well-established psychometric properties and has provided sensitive diagnoses compared with, for example, in-person ratings [53,55,56]. More information on the instruments has been provided in the study protocol [40].

Other Variables

Overview

The original study protocol [40] lists a number of secondary variables. In this report, only data collected through the project collaboration (UvAcare) could be analyzed. Therefore, only a subset of secondary variables that had also been collected in the UvAcare project is reported here.

Psychopathology

In addition to during screening, the MINI was also completed over the phone during follow-up assessments. The interview has reasonably good psychometric properties [43,57,58]. In addition, 2 self-report questionnaires were administered at all assessment points after screening. The 9-item Patient Health Questionnaire (PHQ-9) [59] measures the presence of depressive symptoms during the past 2 weeks on a 4-point Likert scale (0=not at all to 3=nearly every day). The PHQ-9 is a reliable and valid screening instrument with established cutoff scores of 5, 10, and 20 referring to mild, moderate, and severe depression, respectively [59]. The GAD-7 is scored identically and measures anxiety symptoms during the past 2 weeks, with total values of 5, 10, and 15 indicating mild, moderate, and

severe manifestations of generalized anxiety, respectively [42]. The questionnaire has good psychometric properties [60].

Health Care Use

At posttest assessment and during follow-up assessments, CAU use was monitored by asking participants if they had been in contact with any health care professionals (eg, GP, student counseling services, or psychologist) in the previous 2 months. If so, they were asked to indicate how often they had had contact with the professional. However, the reasons for their visits could not be established.

Intervention Adherence

Log data were retrieved from the intervention platform. Given the sequential nature of the intervention, the last completed main module was taken as an indication of intervention progress. Furthermore, the number of diary entries and completed elective modules was retrieved.

Reporting and Statistical Analysis

Recruitment data were used to target the first aim of this study. Where this was traceable (eg, through web statistics or information from the participants), the number of potential participants reached and the number of registrations were reported per recruitment method. Subsequent progress, such as returned informed consent forms and final inclusions, was described. Study dropout was reported per condition and assessment point after baseline by calculating the percentage of included participants who did not complete the interviews and self-report questionnaires. Descriptive statistics for all assessed variables were calculated. Intervention and health care use were reported to target the second aim of this report. This comprises the average number of completed sessions, the percentage of elective modules chosen over the entirety of the sessions (as elective modules could be completed multiple times across sessions), and diary entries, as well as the number of students who visited a health care professional and their contact frequency. For the individually guided intervention condition, the median and average number of messages sent between coaches and participants was calculated.

Finally, to tackle the third aim, we carried out a basic analysis of intervention effectiveness on the intention-to-treat sample under a Bayesian framework. If not otherwise specified, all statistical analyses were run in R (R Foundation for Statistical Computing) and RStudio (version 4.1.1; RStudio, Inc [61]; references for auxiliary packages used are provided at the end of this report). A Bayesian linear mixed model using BRMS (version 2.17.0; Bürkner et al [62]) and RStan (version 2.21.5; Guo et al [63]), including both random and fixed effects (chains=4, iterations=4000, and burn-in phase=1000), was conducted separately for the 2 primary outcomes, the QIDS-CR and SIGH-A, whereby all 4 assessment points were nested within individuals. Gender was entered as a covariate. The analyses were repeated for the self-reported measures of depression (PHQ-9) and anxiety (GAD-7), which contained an additional assessment point (midway). The Bayesian approach—in addition to other advantages (see the articles by Wagenmakers et al [64] and by Verhagen and Wagenmakers [65] for a summary)—allowed us to still provide information on the distribution of the small data set without the power restrictions inherent in the frequentist framework. However, to increase the explanatory value of the data, participants from both intervention groups (ie, individually guided and automatically guided) were pooled to constitute 1 intervention condition, which was then compared with CAU in the analyses through an interaction term with assessment point (for a discussion on the results of a sensitivity analysis including all 3 groups separately, see Multimedia Appendix 4). At each assessment point, the hypothesis that the interaction effect (ie, an intervention effect) was different from zero was tested using contrasts comparing each assessment point with the baseline. Moreover, no prior information was included in the analysis, resulting in flat prior distributions [66]. The probability of the collected data emerging under the null model (ie, meaning that the data were a collection of random noise) was compared with the alternative model (ie, condition [intervention vs CAU] had an effect on outcome). The resulting conditional probability was the Bayes factor (BF), which was reported as a measure of the size of the effect for each postbaseline assessment point. Both notations of the BF are reported: BF_{10}, indicating support for the alternative over the null hypothesis, and BF_{01}, indicating support for the null over the alternative hypothesis [67]. When considering the strength of the support, a BF_{10} of approximately 1 indicates no support, and a BF_{10} of >10 indicates strong support [68]. High-density intervals were calculated to accompany credibility intervals (CIs) as a measure of the uncertainty of the estimated parameters. Finally, the results of the MINI were used to report the number of individuals who met the diagnostic criteria for a mood or anxiety disorder at the 2 follow-up assessments.

Ethics Approval

The ICare Prevent trial was registered in the International Clinical Trial Registry Platform (ICare Prevent NTR6562) and approved by the medical ethics committees of the Amsterdam University Medical Center (NL6075.029.17 and A2018.166). All participants had to provide informed consent upon registration for the study. A data-sharing agreement between the 2 universities involved in this report was set up. This report has been compiled in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines [69].

Results

Recruitment and Screening

The reach of the recruitment activities could only be determined with certainty for targeted social media advertisements (ie, Facebook and Instagram), defined as the number of individuals who were presented with the advertisement on their computer or smartphone. Reach for other channels such as printed advertisements and sports events was estimated based on publicly available information (eg, number of students and number of attendees). Table 1 provides an overview of the reach and number of registrations per recruitment method, with most

registrations stemming from the participant platform (127/202, 62.9%).

Of the 167 students who registered for participation outside of the project collaboration (UvAcare), only 23 (13.8%) returned the signed informed consent form and were assessed for eligibility. Of these 23 individuals, 18 (78%) were excluded at screening based on elevated psychopathology on the MINI (mood disorder: n=7, 39%; anxiety: n=4, 22%; combined mood and anxiety disorder: n=5, 28%; suicide risk: n=2, 11%), 1 (6%) was excluded for being on a waitlist for psychotherapy, and another (6%) was excluded because complaints were too light. This left 13% (3/23) of inclusions, one of which completed the baseline assessment and was randomized. However, this participant dropped out of the study before the midintervention assessment and, thus, only provided baseline data (not reported in this manuscript). Finally, data from 35 students were provided through the collaboration with the UvAcare student mental health project. These students were recruited through their email addresses and provided digital informed consent, making this the most successful approach to recruitment.

Table 1. Reach of potential participants and subsequent registrations for participation per recruitment method (N=202).

Recruitment method	Reach, N	Registrations, n (%)
Social media (Instagram and Facebook)	122,044[a]	9 (4.5)
Printed advertisements (flyers, posters, and newspapers)	115,000[b]	6 (3)
Classrooms and web-based education platforms	Unknown	3 (1.5)
Participant platform	Unknown	127 (62.9)
Conference presentations	Unknown	Unknown
Running event	11,000[c]	Unknown
Targeted emails (UvAcare)	Unknown	35 (17.3)
Unknown or could not be determined	Unknown	22 (10.9)

[a]Actual number based on data extracted from the Facebook advertisement platform.

[b]Estimate based on official numbers of registered students at Utrecht University, Leiden University, University of Groningen, and University of Amsterdam (2018).

[c]Estimate based on the number of participants according to the official website.

Participants and Study Dropout

The 35 participants (individually guided intervention: n=14, 40%; automatically guided intervention: n=8, 23%; CAU: n=13, 37%; descriptives in Table 2) had a mean age of 25.86 (SD 4.75) years, with a slight majority of female students (19/35, 54%). A total of 71% (25/35) were undergraduate students, and 29% (10/35) were PhD students.

Study dropout was 57% (20/35; individually guided condition: 8/14, 57%; automatically guided condition: 4/8, 50%; CAU: 8/13, 62%) at the midway assessment, 29% (10/35; individually guided condition: 0%; automatically guided condition: 3/8, 38%; CAU: 7/13, 54%) at the posttest assessment, and 43% (15/35) at both follow-up assessments (6 months: 5/14, 36% in the individually guided condition, 3/8, 38% in the automatically guided condition, and 7/13, 54% in CAU; 12 months: 5/14, 36% in the individually guided condition, 4/8, 50% in the automatically guided condition, and 6/13, 46% in CAU). It is noteworthy that the relatively higher dropout rate at the midassessment was likely related to the fact that it only consisted of a self-report questionnaire and no telephone interview.

Table 2. Descriptive statistics of clinical variables per condition and assessment point.

Time point	QIDS-CR[a], mean (SD)	Participants, n (%)	PHQ-9[b], mean (SD)	Participants, n (%)	SIGH-A[c], mean (SD)	Participants, n (%)	GAD-7[d], mean (SD)	Participants, n (%)
Total sample (N=35)[e]								
Baseline	5.51 (2.84)	35 (100)	7.54 (3.79)	35 (100)	5.77 (4.51)	35 (100)	6.23 (3.72)	35 (100)
Midway[f]	—[g]	—	5.13 (2.5)	15 (43)	—	—	4.47 (3.7)	15 (43)
Posttest assessment[h]	4.92 (2.77)	25 (71)	6.0 (5.26)	7 (20)	5.32 (4.95)	25 (71)	4.43 (4.24)	7 (20)
6-month follow-up	4.92 (2.53)	13 (37)	5.58 (4.56)	19 (54)	3.69 (4.29)	13 (37)	4.63 (3.72)	19 (54)
12-month follow-up	5.13 (3.59)	16 (46)	5.25 (3.64)	20 (57)	1.06 (1.44)	16 (46)	4.2 (3.82)	20 (57)
Individually guided condition (n=14)[i]								
Baseline	4.79 (2.61)	14 (100)	6.14 (2.03)	14 (100)	4.71 (3.36)	14 (100)	4.93 (3.08)	14 (100)
Midway	—	—	4.0 (2.9)	6 (43)	—	—	3.17 (2.71)	6 (43)
Posttest assessment	4.86 (3.23)	14 (100)	3.0 (1.83)	4 (29)	5.57 (5.92)	14 (100)	2.0 (1.83)	4 (29)
6-month follow-up	5.0 (3.37)	4 (29)	4.22 (2.64)	9 (64)	1.75 (2.06)	4 (29)	3.22 (3.11)	9 (64)
12-month follow-up	4.14 (2.73)	7 (50)	4.78 (3.56)	9 (64)	1.0 (1.53)	7 (50)	4.11 (3.79)	9 (64)
Automatically guided condition (n=8)[j]								
Baseline	5.88 (2.64)	8 (100)	8.0 (3.82)	8 (100)	6.0 (5.61)	8 (100)	7.5 (4.93)	8 (100)
Midway	—	—	6.25 (3.2)	4 (50)	—	—	7.0 (5.89)	4 (50)
Posttest assessment	4.2 (2.05)	5 (62)	N/A[k]	0 (0)	5.0 (4.36)	5 (62)	N/A	0 (0)
6-month follow-up	3.8 (1.92)	5 (62)	2.75 (3.59)	4 (50)	2.8 (4.38)	5 (62)	3.5 (5.2)	4 (50)
12-month follow-up	2.0 (0.0)	2 (25)	2.25 (1.71)	4 (50)	0.0 (0.0)	2 (25)	2.25 (4.5)	4 (50)
CAU[l] condition (n=13)[m]								
Baseline	6.1 (3.23)	13 (100)	8.77 (4.87)	13 (100)	6.77 (4.97)	13 (100)	6.85 (3.36)	13 (100)
Midway	—	—	5.6 (0.55)	5 (38)	—	—	4.0 (1.87)	5 (38)
Posttest assessment	5.68 (2.25)	6 (46)	10.0 (6.0)	3 (23)	5.0 (3.29)	6 (46)	7.67 (4.62)	3 (23)
6-month follow-up	6.25 (2.22)	4 (31)	9.5 (5.28)	6 (46)	6.75 (4.99)	4 (31)	7.5 (1.87)	6 (46)
12-month follow-up	7.0 (4.04)	7 (54)	7.57 (3.31)	7 (54)	1.43 (1.51)	7 (54)	5.43 (3.55)	7 (54)

[a]QIDS-CR: Quick Inventory of Depressive Symptomatology-Clinician Rated [48].

[b]PHQ-9: Patient Health Questionnaire [59].

[c]SIGH-A: Structured Interview Guide for the Hamilton Anxiety Rating Scale [53,54].

[d]GAD-7: Generalized Anxiety Disorder Scale [42].

[e]Mean age 25.86 (SD 4.75) years; 54% (19/35) women; 46% (16/35) men.

[f]Midway: 5 weeks after randomization; only self-report assessed.

[g]Not available.

[h]Posttest assessment: 8 weeks after randomization.

[i]Mean age 27.86 (SD 6.67) years; 64% (9/14) women; 36% (5/14) men.

[j]Mean age 25.5 (SD 1.41) years; 50% (4/8) women; 50% (4/8) men.

[k]N/A: not applicable.

[l]CAU: care as usual.

[m]Mean age 23.92 (SD 2.4) years; 46% (6/13) women; 54% (7/13) men.

Intervention and Health Care Use

The 63% (22/35) of participants in the intervention conditions completed an average of 3 sessions (SD 2.9; individually guided condition: mean 4, SD 2.9; automatically guided condition: mean 2, SD 2.4). A total of 7 participants (individually guided condition: n=3, 43%; automatically guided condition: n=4, 57%) started but did not complete the first session, and only 7% (1/14) of the participants in the individually guided condition completed the booster module. In most sessions that offered the selection of an elective module (ie, sessions 2 to 7; 19/49, 39%), no elective module was chosen. Across all these sessions, the most frequently selected module dealt with improving self-esteem (9/49, 18%), followed by gratitude (6/49, 12%), improving sleep, and dealing with perfectionism (5/49, 10% each). The elective module teaching acceptance techniques was chosen in 6% (3/49) of all completed sessions. The relaxation, rumination, and reducing alcohol consumption modules were chosen the least frequently, namely, in 2% (1/49) of all completed sessions each. Only 14% (2/14) of the participants in the individually guided condition used the diary function, with a total of 8 entries in the activity diary. On average, the coaches and participants in this condition exchanged 5 (SD 4.79; median 3) messages.

At posttest assessment, 4 students (individually guided condition: n=2, 50%; CAU: n=2, 50%) indicated that they had consulted one or more health care professionals in the previous 2 months (GP: n=3, 75%; student psychologist or psychotherapist: n=2, 50%; study advisor or counselor: n=2, 50%; other: n=1, 25%). Frequencies ranged from 1 to 3 visits. This increased to 11 students (individually guided condition: n=4, 36%; automatically guided condition: n=3, 27%; CAU: n=4, 36%) at the 6-month follow-up (GP: n=4, 36%; study advisor or counselor: n=4, 36%; student psychologist or psychotherapist: n=4, 36%; self-help group: n=1, 9%; other: n=4, 36%), with frequencies ranging from single to weekly visits (self-help group). At the 12-month follow-up, 10 students (individually guided condition: n=4, 40%; automatically guided condition: n=2, 20%; CAU: n=4, 40%) had visited one or more health care professionals (study advisor or counselor: n=1, 10%; student psychologist or psychotherapist: n=5, 50%; other: n=3, 30%). The visits ranged from 1 to 7 times.

Symptom Change and Diagnoses

The results on symptom change (Table 3) suggest moderate to strong support for the null hypothesis (ie, the absence of an intervention effect) on almost all mental health outcomes at all assessment points (BF_{10} range=0.03-2.7; BF_{01} range=0.37-35.36) [68]. Values of >1 for BF_{10}, which indicates support for the hypothesis that an intervention effect is present, were found exclusively in measures of anxiety. Specifically, for the GAD-7, the interaction between condition and time (β=−0.39, 95% CI −3.78 to 2.91; BF_{01}=0.7) was 1.43 times more likely to emerge under the alternative hypothesis than under the null hypothesis at the midway assessment compared with the baseline. For the SIGH-A, this interaction was found to be 1.64 times more likely at posttest assessment versus baseline (β=−0.82, 95% CI −6.18 to 4.52; BF_{01}=0.61) and 2.7 times more likely at the 12-month follow-up versus baseline (β=−1.36, 95% CI −5.86 to 3.13; BF_{01}=0.37) under the alternative hypothesis.

Although these effects indicate a superior effect of CAU over the intervention condition, as CAU shows stronger reductions in symptoms compared with baseline at these assessment points, the strength of the support is only anecdotal [68] and is associated with significant uncertainty, as indicated by both the CIs and high-density intervals, reflecting the small sample size (see also Figures 2-5). Thus, there is no reliable evidence of an intervention effect (Table 3).

At the 6-month follow-up assessment, 13 students completed the telephone interviews, of whom 1 (8%) in the individually guided intervention condition presented with a GAD according to the MINI. In total, 16 students completed the 12-month follow-up interviews, of whom 2 (12%) in the CAU condition met diagnostic criteria: 1 (50%) for dysthymia and another (50%) for GAD. Another 14% (2/14) of the students in the individually guided condition presented with mixed anxiety and depression.

Table 3. Estimates and Bayes factors (BFs) for the interaction with condition per assessment point and instrument.

Time point	Estimate, β	95% CI[a]	95% HDI[b]	BF_{10}[c]	BF_{01}[d]
QIDS-CR[e]					
Baseline vs posttest assessment	1.25	−1.63 to 4.09	−1.62 to 4.10	0.23	4.36
Baseline vs 6-month follow-up	2.29	−1.19 to 5.79	−1.00 to 5.96	0.10	10.18
Baseline vs 12-month follow-up	2.21	−0.65 to 5.04	−0.58 to 5.08	0.06	15.67
PHQ-9[f]					
Baseline vs midway	.34	−4.47 to 5.13	−4.65 to 4.93	0.78	1.28
Baseline vs posttest assessment	3.36	−5.12 to 11.6	−5.42 to 11.27	0.22	4.54
Baseline vs 6-month follow-up	4.10	−1.24 to 9.44	−1.43 to 9.21	0.06	15.78
Baseline vs 12-month follow-up	1.70	−2.21 to 5.60	−2.10 to 5.70	0.23	4.30
SIGH-A[g]					
Baseline vs posttest assessment	−.82	−6.18 to 4.52	−6.13 to 4.54	1.64	0.61
Baseline vs 6-month follow-up	3.94	−2.15 to 10.18	−2.39 to 9.90	0.10	9.88
Baseline vs 12-month follow-up	−1.36	−5.86 to 3.13	−5.92 to 3.07	2.70	0.37
GAD-7[h]					
Baseline vs midway	−.39	−3.78 to 2.91	−3.78 to 2.91	1.43	0.70
Baseline vs posttest assessment	3.07	−3.53 to 9.64	−3.79 to 9.28	0.16	6.43
Baseline vs 6-month follow-up	3.23	−0.09 to 6.53	0.09 to 6.69	0.03	35.36
Baseline vs 12-month follow-up	.68	−3.07 to 4.60	−3.09 to 4.57	0.56	1.79

[a]CI: credibility interval.

[b]HDI: high-density interval.

[c]BF_{10}: BF indicating probability of alternative over null hypothesis.

[d]BF_{01}: BF indicating probability of null over alternative hypothesis.

[e]QIDS-CR: Quick Inventory of Depressive Symptomatology-Clinician Rated [48].

[f]PHQ-9: Patient Health Questionnaire [59].

[g]SIGH-A: Structured Interview Guide for the Hamilton Anxiety Rating Scale.

[h]GAD-7: Generalized Anxiety Disorder Scale [42].

Figure 2. Depression scores (Quick Inventory of Depressive Symptomatology-Clinician Rated [QIDS-CR]) per assessment point and pooled condition with 95% credibility intervals.

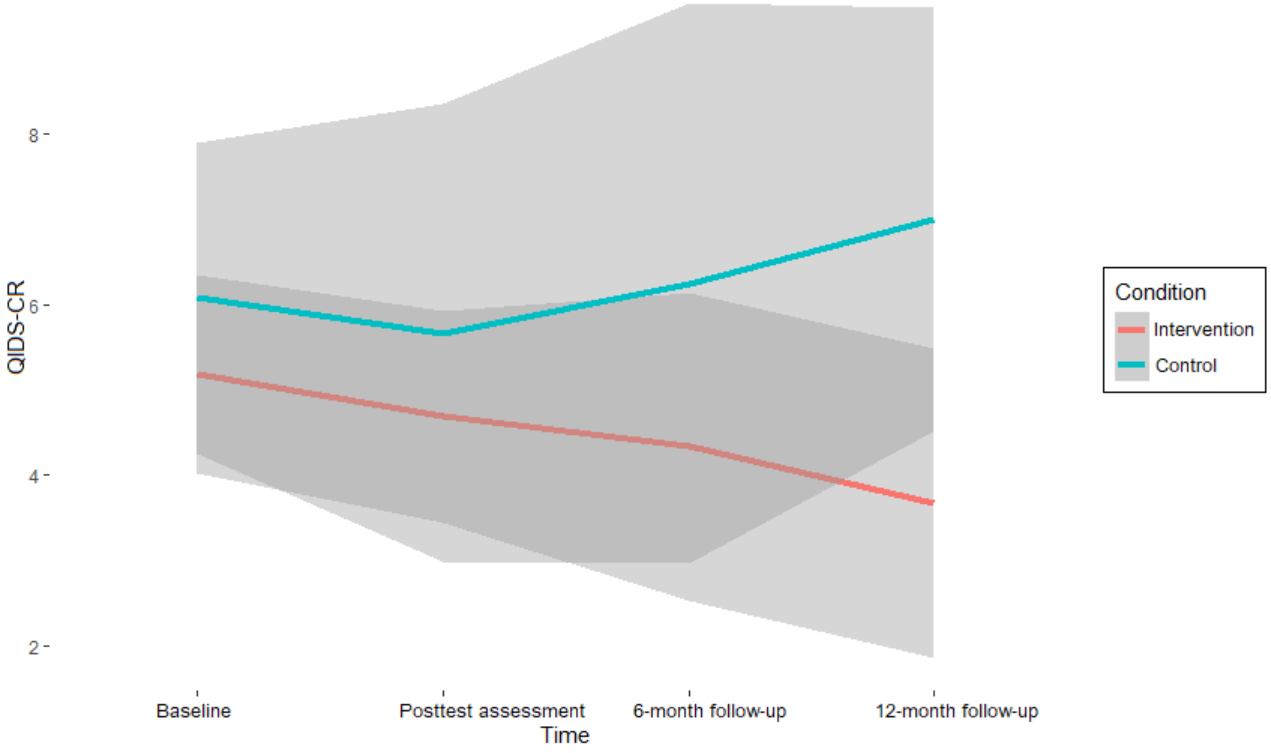


Figure 3. Self-reported depression scores (Patient Health Questionnaire [PHQ-9]) per assessment point and pooled condition with 95% credibility intervals.

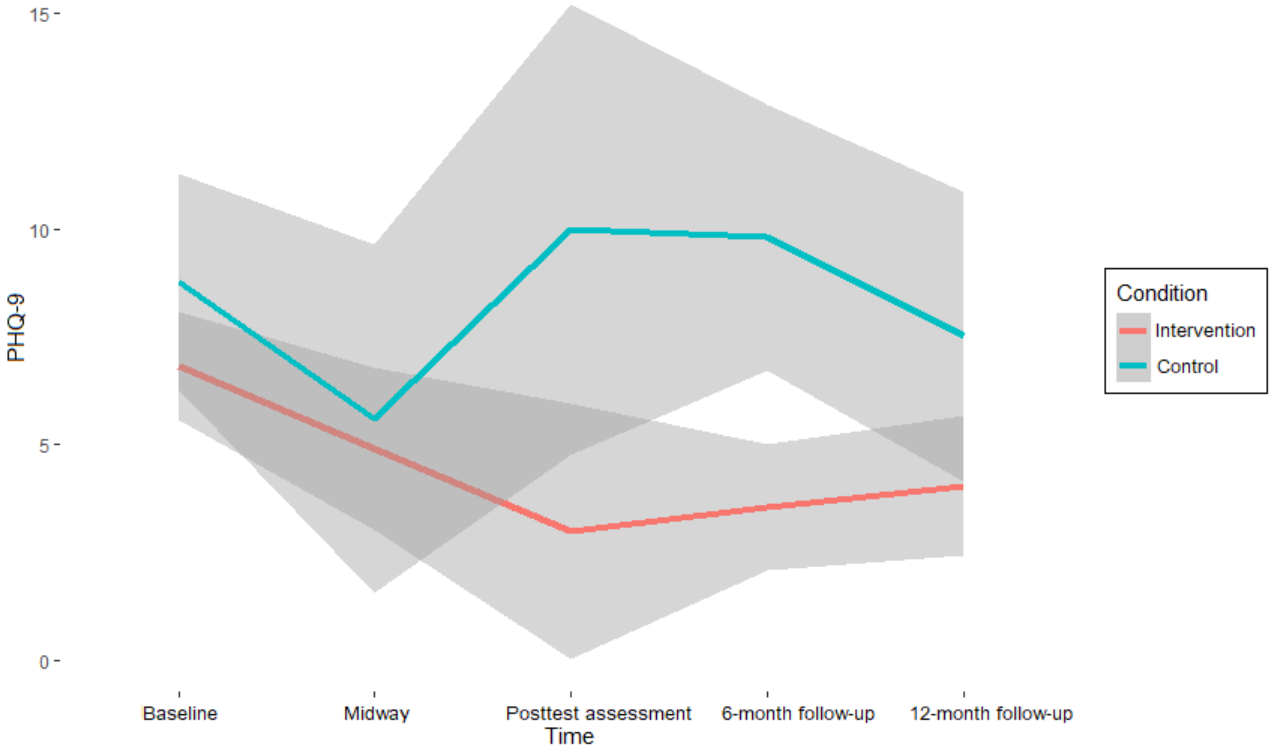


Figure 4. Anxiety scores (Structured Interview Guide for the Hamilton Anxiety Rating Scale [SIGH-A]) per assessment point and pooled condition with 95% credibility intervals.

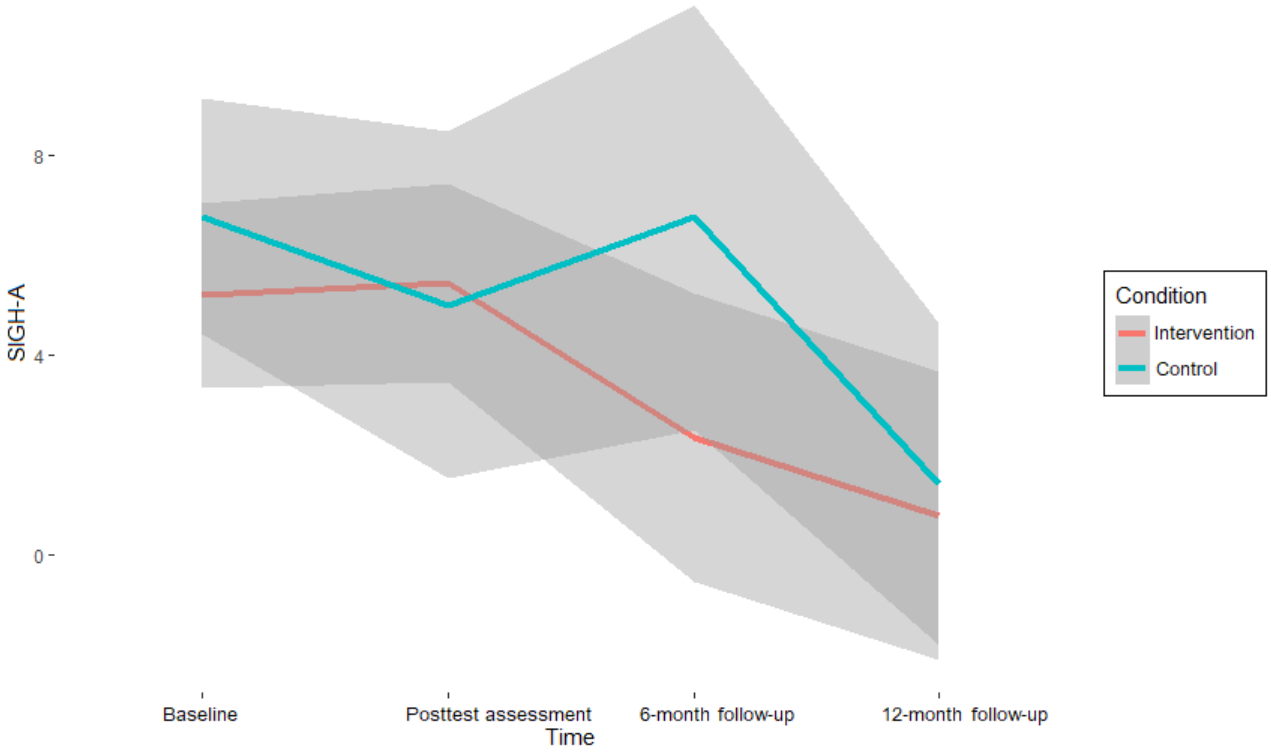
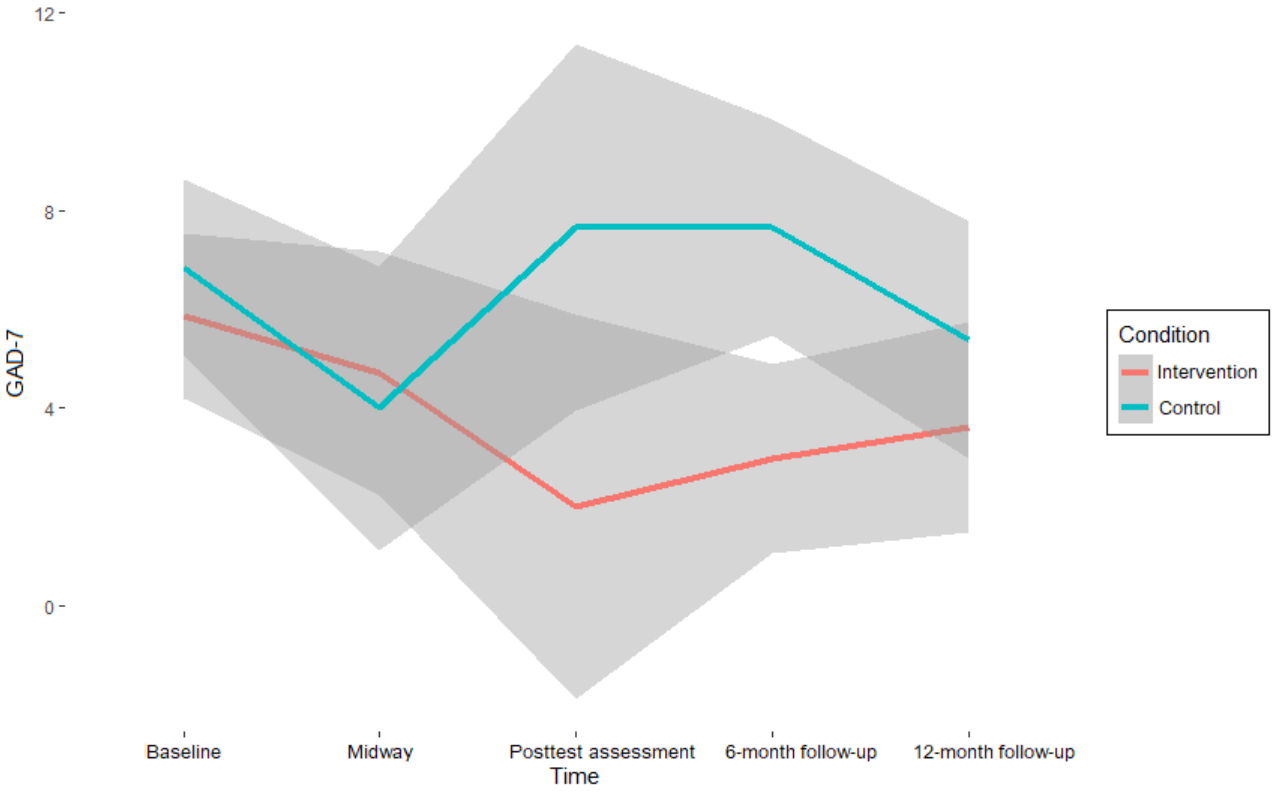


Figure 5. Self-reported anxiety scores (Generalized Anxiety Disorder Scale [GAD-7]) per assessment point and pooled condition with 95% credibility intervals.



Discussion

Despite the availability of effective low-threshold mental health interventions, such as those provided via the internet [28], students seem to be difficult to recruit for interventions on this topic. Therefore, the aims of this study were to (1) describe the recruitment process for an RCT of a web-based intervention for the prevention of depression and anxiety in students, (2) describe how participants used the intervention, and (3) conduct a basic analysis of intervention effects.

Recruitment Process

We used an extensive and multifaceted recruitment strategy, ranging from traditional print and social media advertisements to more unconventional means such as participation in a sports event. Ultimately, direct contact through students' email addresses was the most effective recruitment channel, which was a unique opportunity in the UvAcare project [44]. A larger number of students could be reached efficiently at the same time, although this method focused largely on screening students at the beginning of a semester, when exam anxiety and other stressors might not have set in yet. It is possible that more students would have fulfilled the inclusion criteria toward the end of the semester. However, this strategy potentially reduced the feeling of being individually targeted for mental health complaints. Although, to the best of our knowledge, this has not been tested yet, it is conceivable that—compared with a universal recruitment strategy—an email from the central student administration may have led to more compliant behavior in completing the screener and providing informed consent for the subsequent RCT. Moreover, the use of digital informed consent reduced the barrier to participation in the UvAcare project. Printing and posting a form, as required in the initial trial, likely negated the benefits of a digital intervention (eg, its ease of access). A recently conducted RCT supports this notion [70]. The authors assessed the effectiveness of the ICare Prevent intervention on symptom reduction rather than its preventive potential. Although their inclusion and exclusion criteria were less restrictive than those reported in this study and the target sample size was eventually reached, ≤3% of all recruited students could be randomized, largely because students did not provide written informed consent. It needs to be stated that the COVID-19 pandemic has led the evaluating Medical Ethics Committee to reconsider its reluctance to allow digital informed consent. In addition, the possibility that students genuinely did not want to consent to the research procedure cannot be excluded.

In line with the aforementioned potential explanations, the individual approach through the participant platform generated by far the largest number of study sign-ups. As the platform was geared toward participation in clinical studies, students who had registered were likely already aware of health complaints, either physical or mental, both of which are often comorbid [71]. Consequently, almost all students recruited through this channel exceeded the clinical cutoff for mental health problems. It is noteworthy that the screening instrument we used (MINI) has been shown to be overly sensitive and, therefore, likely overestimated the prevalence of major depression in our sample [58]. However, the high degree of psychopathology also sheds light on another potential underlying issue, namely, students' attitudes toward help seeking in general [31]. This problem might also explain why the most efficient approach to recruitment (ie, direct contact through email addresses) resulted in the randomization of only 35 students.

Although, in recent mental health surveys among students in the Netherlands, a response rate of <12% was found [72], detailed accounts of recruitment challenges are generally limited to treatment studies in nonstudent populations. An early systematic review of 78 studies summarized a decade of barriers to trial recruitment [73]. In line with our findings, the most common concerns were related to the information and consent procedure and the inconvenience caused by the research context (eg, strict procedures, additional appointments, and travel time, the latter being less applicable as there is no travel time and limited in-person contact with physicians or scientific staff in trials on digital interventions). Subsequent reviews have confirmed the relevance of these concerns to recruitment and participation [74,75]. In an attempt to quantify the problem, a study systematically reviewed 1017 study protocols of RCTs, most from the field of oncology and cardiovascular medicine, and compared their available recruitment and publication status [76]. Approximately one-quarter of the studies were terminated prematurely, most commonly because of recruitment challenges and, of these, only 40% were published [76]. These challenges are likely more pronounced in prevention studies [29] considering that individuals at prodromal stages of, for example, depression might not feel the urge to seek help and might prefer self-management of complaints. On the basis of these findings, a set of general recommendations [30,77] should be emphasized.

First, we encourage complementing RCTs of new interventions or in new target groups with well-designed feasibility trials [77]. A recent example is provided by a study assessing the feasibility and acceptability of the ICare Prevent intervention in Indonesian students, with positive results [78]. Importantly, such web-based interventions are novel in Indonesia, and treatment for mental health complaints is limited in the country, potentially explaining why recruitment was less problematic. Conversely, another study reported substantial challenges in recruiting participants, with the informed consent procedure again being a major barrier [79]. In this feasibility study, the ICare Prevent intervention was adapted for patients with coronary artery disease. Although such feasibility or pilot trials are regularly conducted, it is advisable to routinely extend the assessment of the feasibility, acceptability, and cost-effectiveness to the potential recruitment measures and report these in outcome papers. Therefore, a structured and preplanned evaluation of recruitment strategies is important. For example, a study indicated that recruitment via posters and websites was most cost-effective in the context of an RCT on depression relapse prevention while noting that a multifaceted recruitment strategy was crucial [80]. In our study, such a diverse set of strategies proved unsuccessful. In line with this, a review of effective approaches to recruiting participants for RCTs has shown that the quality of available evidence is low and that those effective approaches that emerged from higher-quality studies, such as unblinding trials and calling potential participants, are not applicable in many research

contexts [81]. On a brighter note, the dearth of information has advanced initiatives such as the Prioritising Recruitment in Randomised Trials project [82], which entails a web-based compilation of important recruitment-related questions and answers provided by experts. It offers an open-access database of expert-based suggestions on ways to improve recruitment for clinical trials.

Second, cocreation with end users should be considered beyond intervention development. Notably, a systematic review suggested that involving participants in the creation of study information material was not efficient [81]. However, active participation of students in the identification of effective recruitment channels beyond development (eg, flyers and posters) would allow for a targeted recruitment strategy. In line with this idea, although a broad recruitment strategy was deemed most effective initially for the ICare Prevent trial, the results outlined previously show that saving efforts for specific channels is more efficient.

Finally, it has been suggested to overcome the relatively low uptake of preventive mental health interventions by focusing on indirect prevention—interventions that do not directly address mental health problems but rather target related issues (eg, insomnia) with the aim of positively affecting mental health conditions [30]. More research is needed to investigate this possibility, but examples of digital interventions that use such an indirect prevention approach exist, for example, a set of Complaint-Directed Mini-Interventions for depression, evaluated in the general population [83]. These interventions consist of short web-based modules for improving sleep, reducing stress, and tackling excessive worrying. The authors found not only a significant reduction in depressive symptoms at the 3-month follow-up compared with a waitlist control (d=−0.7) but also a promising sign-up rate and provision of informed consent. In the context of our study, the prominent focus on depression and anxiety might have deterred students at subclinical stages. As noted previously, these students might not have identified as having depressive or anxiety complaints, thereby reducing the need for seeking help. Moreover, in the absence of an urge to seek help, the research context (eg, informed consent, interviews, and questionnaires) might have been experienced as too burdensome. This serves as a potential explanation for why mostly those students with clinical presentations of mental health conditions signed up for the RCT.

Intervention Use and Effects

The use of the intervention was low, with an average of 3 sessions completed by the participants. Some of the main intervention components were presented in these first sessions, such as psychoeducation and behavioral activation. However, the transdiagnostic components, which were mainly contained in the elective modules, were only provided from the second session onward, and other core elements such as cognitive restructuring followed in later sessions. Moreover, the analyses did not suggest an intervention effect on depression or anxiety outcomes. Only anecdotal evidence for the added benefit of individual guidance was obtained as the only participant who completed the booster session received support from a coach. Whether the low use of the intervention also affected the absence of intervention effects on mental health outcomes cannot be established because of the small sample size. However, these findings complement studies on the same intervention conducted in different contexts. Although one study reported positive within-group effects immediately after the intervention (depression: d=0.42; anxiety: d=1.19) [84], another study found no difference between the ICare Prevent intervention and CAU in an RCT among students [70]. Moreover, a meta-analysis on digital mental health interventions in students has suggested that the effect sizes of such interventions are small at best in this population (depression: d=0.18; anxiety: d=0.27), and subsequent trials will likely include null findings (prediction interval for depression: −0.26 to 0.62; prediction interval for anxiety: −0.36 to 0.90) [28]. Therefore, it is conceivable that the ICare Prevent intervention requires further adaptation before being suitable for students.

Limitations and Strengths

This study has a number of limitations. The encountered recruitment challenges resulted in a small sample size, which had considerable implications. First, it meant that we could not follow the analytical approach outlined in the study protocol [40]. Additional analyses (eg, on the relationship between dropout and outcome) were not possible. Moreover, pooling the individually and automatically guided intervention conditions was a compromise to reach stable statistical models that would allow for a description of the data distribution (see also Multimedia Appendix 4). Previous research has consistently indicated the superiority of individually guided interventions, specifically those involving human rather than technical support [85], although this applies less so for subclinical stages of depression [86]. Therefore, any suggestions and findings need to be considered exploratory and incidental and require replication. However, we chose the alternative Bayesian framework to overcome the sample size limitations inherent in frequentist statistics and provide a basic account of the intervention's effectiveness. Although we used noninformative priors, a strength of this study is that it provides a starting point for future trials. Essentially, subsequent RCTs on the topic of transdiagnostic preventive digital health interventions for students can build on our data and use them as prior information [64,87]. This would allow for a continuous research circle from feasibility trial, which generates the prior information, to full RCT, which provides up-to-date information.

Moreover, we did not use a structured evaluation of the recruitment challenges or a broader process evaluation as this was not part of the originally planned RCT. This includes a systematic appraisal of recruitment costs, which we have not reported on because of a lack of reliable data. Although this is a lesson learned for future RCTs, reporting both unforeseen recruitment challenges and data on participants and their use of the intervention is important nevertheless. For example, evidence of the effectiveness of interventions might suffer from publication bias. In this regard, a strength of this study is the publication of data that can be beneficial not only in guiding future studies on developing recruitment strategies but also in data synthesis efforts such as meta-analyses. The high costs involved in clinical trials and the intensive involvement of both staff and participants mandate such reporting of the collected

data, which hopefully can aid in preventing similar situations in the future and, therefore, save scarce resources [76].

Conclusions and Future Research

The recruitment of students for digital mental health interventions that focus on the prevention of depression and anxiety is difficult. Although targeted approaches such as direct email contact seem the most efficient, more research is needed on factors that can improve recruitment, and subsequent improved strategies need to be developed. Moreover, evidence on whether direct rather than indirect prevention efforts are suitable for this target group is mixed and requires further investigation. We provided an account of recruitment challenges as well as basic information on intervention effects that can aid future studies in the development and evaluation of similar interventions.

Acknowledgments

The ICare Prevent trial was funded by the European Commission Horizon 2020 Research and Innovation action framework under grant agreement 634757. The authors would like to thank Marleen Swenne for her work on adapting the intervention to the student context; Angelika Stefan for the initial brainstorming about Bayesian statistics; Jaël Sierra van Bentum for implementing the randomization sequence; Nine Wolters, Ruben Atteveld, and Anke Klein for their work on the UvAcare trial; Vincent Palte and Fleur Lock for conducting the clinical interviews; and Joost Gladpootjes, Hinna Mirza, Nurdane Aslan, and Khadicha Amarti for recruitment.

Data Availability

The statistical packages in RStudio (RStudio, Inc) used were as follows: bayestestR (version 0.12.1 [88]), foreign (version 0.8-82 [89]), ggplot2 (version 3.3.5 [90]), Hmisc (version 4.7-0. [91]), and Psych (version 2.2.3. [92]). Full analytical code is available [93].

Authors' Contributions

FB contributed to conceptualization, methodology, formal analysis, investigation, data curation, writing (original draft), visualization, and project administration. AK contributed to conceptualization, investigation, writing (review and editing), supervision, and funding acquisition. KN contributed to methodology, data curation, and writing (review and editing). EK contributed to supervision and writing (review and editing). RW contributed to supervision and writing (review and editing). LdK contributed to data curation and writing (review and editing). CJ contributed to conceptualization, writing (review and editing), and funding acquisition. ACZ contributed to software and writing (review and editing). KKW contributed to software and writing (review and editing). PC contributed to supervision and writing (review and editing). HR contributed to conceptualization, investigation, writing (review and editing), supervision, and funding acquisition. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Recruitment poster.
[PNG File , 2826 KB - jmir_v24i12e40892_app1.png]

Multimedia Appendix 2
Recruitment sticker.
[PNG File , 864 KB - jmir_v24i12e40892_app2.png]

Multimedia Appendix 3
Intervention description.
[DOCX File , 13 KB - jmir_v24i12e40892_app3.docx]

Multimedia Appendix 4
Sensitivity analysis on all conditions.
[DOCX File , 48 KB - jmir_v24i12e40892_app4.docx]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1185 KB - jmir_v24i12e40892_app5.pdf]

References

1. Kalin NH. The critical relationship between anxiety and depression. Am J Psychiatry 2020 May 01;177(5):365-367. [doi: 10.1176/appi.ajp.2020.20030305] [Medline: 32354270]
2. Lamers F, van Oppen P, Comijs HC, Smit JH, Spinhoven P, van Balkom AJ, et al. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). J Clin Psychiatry 2011 Mar;72(3):341-348. [doi: 10.4088/JCP.10m06176blu] [Medline: 21294994]
3. Lijster JM, Dierckx B, Utens EM, Verhulst FC, Zieldorff C, Dieleman GC, et al. The age of onset of anxiety disorders. Can J Psychiatry 2017 Apr;62(4):237-246 [FREE Full text] [doi: 10.1177/0706743716640757] [Medline: 27310233]
4. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022 Jan;27(1):281-295 [FREE Full text] [doi: 10.1038/s41380-021-01161-7] [Medline: 34079068]
5. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, WHO WMH-ICS Collaborators. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]
6. de Girolamo G, McGorry PD, Sartorius N. Age of Onset of Mental Disorders: Etiopathogenetic and Treatment Implications. Berlin, Germany: Springer; 2019.
7. Newcomb-Anjo SE, Villemaire-Krajden R, Takefman K, Barker ET. The unique associations of academic experiences with depressive symptoms in emerging adulthood. Emerg Adulthood 2016 Jul 07;5(1):75-80. [doi: 10.1177/2167696816657233]
8. Barker ET, Howard AL, Villemaire-Krajden R, Galambos NL. The rise and fall of depressive symptoms and academic stress in two samples of university students. J Youth Adolesc 2018 Jun;47(6):1252-1266. [doi: 10.1007/s10964-018-0822-9] [Medline: 29470761]
9. Karyotaki E, Cuijpers P, Albor Y, Alonso J, Auerbach RP, Bantjes J, et al. Sources of stress and their associations with mental disorders among college students: results of the World Health Organization world mental health surveys international college student initiative. Front Psychol 2020 Jul 30;11:1759 [FREE Full text] [doi: 10.3389/fpsyg.2020.01759] [Medline: 32849042]
10. Beiter R, Nash R, McCrady M, Rhoades D, Linscomb M, Clarahan M, et al. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. J Affect Disord 2015 Mar 01;173:90-96. [doi: 10.1016/j.jad.2014.10.054] [Medline: 25462401]
11. Savage MJ, James R, Magistro D, Donaldson J, Healy LC, Nevill M, et al. Mental health and movement behaviour during the COVID-19 pandemic in UK university students: prospective cohort study. Ment Health Phys Act 2020 Oct;19:100357 [FREE Full text] [doi: 10.1016/j.mhpa.2020.100357]
12. Odriozola-González P, Planchuelo-Gómez Á, Irurtia MJ, de Luis-García R. Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. Psychiatry Res 2020 Aug;290:113108 [FREE Full text] [doi: 10.1016/j.psychres.2020.113108] [Medline: 32450409]
13. Robb CA. College student financial stress: are the kids alright? J Fam Econ Iss 2017 Apr 19;38(4):514-527. [doi: 10.1007/s10834-017-9527-6]
14. Richardson T, Elliott P, Roberts R, Jansen M. A longitudinal study of financial difficulties and mental health in a national sample of British undergraduate students. Community Ment Health J 2017 Apr;53(3):344-352 [FREE Full text] [doi: 10.1007/s10597-016-0052-0] [Medline: 27473685]
15. Cuijpers P, Auerbach RP, Benjet C, Bruffaerts R, Ebert D, Karyotaki E, et al. The World Health Organization world mental health international college student initiative: an overview. Int J Methods Psychiatr Res 2019 Jun;28(2):e1761 [FREE Full text] [doi: 10.1002/mpr.1761] [Medline: 30614123]
16. Jenkins PE, Ducker I, Gooding R, James M, Rutter-Eley E. Anxiety and depression in a sample of UK college students: a study of prevalence, comorbidity, and quality of life. J Am Coll Health 2021;69(8):813-819. [doi: 10.1080/07448481.2019.1709474] [Medline: 31995452]
17. Alonso J, Mortier P, Auerbach RP, Bruffaerts R, Vilagut G, Cuijpers P, WHO WMH-ICS Collaborators. Severe role impairment associated with mental disorders: results of the WHO World Mental Health Surveys International College Student Project. Depress Anxiety 2018 Sep;35(9):802-814 [FREE Full text] [doi: 10.1002/da.22778] [Medline: 29847006]
18. Mortier P, Demyttenaere K, Auerbach RP, Green JG, Kessler RC, Kiekens G, et al. The impact of lifetime suicidality on academic performance in college freshmen. J Affect Disord 2015 Nov 01;186:254-260 [FREE Full text] [doi: 10.1016/j.jad.2015.07.030] [Medline: 26254617]
19. Bruffaerts R, Mortier P, Kiekens G, Auerbach RP, Cuijpers P, Demyttenaere K, et al. Mental health problems in college freshmen: prevalence and academic functioning. J Affect Disord 2018 Jan 01;225:97-103 [FREE Full text] [doi: 10.1016/j.jad.2017.07.044] [Medline: 28802728]
20. Bodden DH, van den Heuvel MW, Engels RC, Dirksen CD. Societal costs of subclinical depressive symptoms in Dutch adolescents: a cost-of-illness study. J Child Psychol Psychiatry 2022 Jul;63(7):771-780 [FREE Full text] [doi: 10.1111/jcpp.13517] [Medline: 34496447]

21. Agasisti T, Bertoletti A. Higher education and economic growth: a longitudinal study of European regions 2000–2017. Socio-Econ Plan Sci 2022 Jun;81:100940 [FREE Full text] [doi: 10.1016/j.seps.2020.100940]
22. Institute of Medicine (US) Committee on Prevention of Mental Disorders. In: Mrazek PJ, Haggerty RJ, editors. Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research. Washington, DC, USA: National Academies Press (US); 1994.
23. Buntrock C, Ebert DD, Lehr D, Smit F, Riper H, Berking M, et al. Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression: a randomized clinical trial. JAMA 2016 May 03;315(17):1854-1863. [doi: 10.1001/jama.2016.4326] [Medline: 27139058]
24. McDaid D, Park AL, Wahlbeck K. The economic case for the prevention of mental illness. Annu Rev Public Health 2019 Apr 01;40:373-389. [doi: 10.1146/annurev-publhealth-040617-013629] [Medline: 30601725]
25. Cuijpers P, Pineda BS, Quero S, Karyotaki E, Struijs SY, Figueroa CA, et al. Psychological interventions to prevent the onset of depressive disorders: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2021 Mar;83:101955 [FREE Full text] [doi: 10.1016/j.cpr.2020.101955] [Medline: 33333441]
26. Moreno-Peral P, Conejo-Cerón S, Rubio-Valera M, Fernández A, Navas-Campaña D, Rodríguez-Morejón A, et al. Effectiveness of psychological and/or educational interventions in the prevention of anxiety: a systematic review, meta-analysis, and meta-regression. JAMA Psychiatry 2017 Oct 01;74(10):1021-1029 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.2509] [Medline: 28877316]
27. Barnett P, Arundell LL, Saunders R, Matthews H, Pilling S. The efficacy of psychological interventions for the prevention and treatment of mental health disorders in university students: a systematic review and meta-analysis. J Affect Disord 2021 Mar 01;280(Pt A):381-406. [doi: 10.1016/j.jad.2020.10.060] [Medline: 33227669]
28. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
29. Cuijpers P, van Straten A, Warmerdam L, van Rooy MJ. Recruiting participants for interventions to prevent the onset of depressive disorders: possible ways to increase participation rates. BMC Health Serv Res 2010 Jun 25;10:181 [FREE Full text] [doi: 10.1186/1472-6963-10-181] [Medline: 20579332]
30. Cuijpers P. Indirect prevention and treatment of depression: an emerging paradigm? Clin Psychol Eur 2021 Dec;3(4):e6847 [FREE Full text] [doi: 10.32872/cpe.6847] [Medline: 36398290]
31. Ebert DD, Mortier P, Kaehlke F, Bruffaerts R, Baumeister H, Auerbach RP, WHO World Mental Health-International College Student Initiative collaborators. Barriers of mental health treatment utilization among first-year college students: first cross-national results from the WHO World Mental Health International College Student Initiative. Int J Methods Psychiatr Res 2019 Jun;28(2):e1782 [FREE Full text] [doi: 10.1002/mpr.1782] [Medline: 31069905]
32. Hirsch JK, Rabon JK, Reynolds EE, Barton AL, Chang EC. Perceived stress and suicidal behaviors in college students: conditional indirect effects of depressive symptoms and mental health stigma. Stigma Health 2019 Feb;4(1):98-106. [doi: 10.1037/sah0000125]
33. Wada M, Suto MJ, Lee M, Sanders D, Sun C, Le TN, et al. University students' perspectives on mental illness stigma. Ment Health Prev 2019 Jun;14:200159 [FREE Full text] [doi: 10.1016/j.mph.2019.200159]
34. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017 Nov;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
35. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
36. Ennis E, McLafferty M, Murray E, Lapsley C, Bjourson T, Armour C, et al. Readiness to change and barriers to treatment seeking in college students with a mental disorder. J Affect Disord 2019 Jun 01;252:428-434. [doi: 10.1016/j.jad.2019.04.062] [Medline: 31003112]
37. Küchler AM, Schultchen D, Pollatos O, Moshagen M, Ebert DD, Baumeister H. StudiCare mindfulness-study protocol of a randomized controlled trial evaluating an internet- and mobile-based intervention for college students with no and "on demand" guidance. Trials 2020 Nov 26;21(1):975 [FREE Full text] [doi: 10.1186/s13063-020-04868-0] [Medline: 33243300]
38. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
39. Cuijpers P. Why primary prevention often is no prevention at all. Eur Neuropsychopharmacol 2022 May;58:1-3. [doi: 10.1016/j.euroneuro.2022.01.004] [Medline: 35077931]
40. Bolinski F, Kleiboer A, Karyotaki E, Bosmans JE, Zarski AC, Weisel KK, et al. Effectiveness of a transdiagnostic individually tailored Internet-based and mobile-supported intervention for the indicated prevention of depression and anxiety (ICare Prevent) in Dutch college students: study protocol for a randomised controlled trial. Trials 2018 Feb 20;19(1):118 [FREE Full text] [doi: 10.1186/s13063-018-2477-y] [Medline: 29458407]

41. Radloff L. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977 Jun;1(3):385-401. [doi: 10.1177/014662167700100306]
42. 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]
43. 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 20:22-57. [Medline: 9881538]
44. Klein A, Wolters NE, Bol EJM, Koelen J, de Koning L, Roetink SS, et al. Online computer or therapist-guided cognitive behavioral therapy in university students with anxiety and/or depression: study protocol of a randomised controlled trial. BMJ Open 2021 Nov 26;11(11):e049554 [FREE Full text] [doi: 10.1136/bmjopen-2021-049554] [Medline: 34836897]
45. Link2trials. 2017. URL: https://www.link2trials.com/ [accessed 2022-12-06]
46. Results of the 2018 Dam tot Damloop. Le Champion. 2018. URL: https://results.sporthive.com/events/6447830197266687232 [accessed 2022-12-06]
47. Weisel KK, Zarski AC, Berger T, Krieger T, Schaub MP, Moser CT, et al. Efficacy and cost-effectiveness of guided and unguided internet- and mobile-based indicated transdiagnostic prevention of depression and anxiety (ICare Prevent): a three-armed randomized controlled trial in four European countries. Internet Interv 2019 Apr;16:52-64 [FREE Full text] [doi: 10.1016/j.invent.2018.04.002] [Medline: 30775265]
48. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 01;54(5):573-583. [doi: 10.1016/s0006-3223(02)01866-8] [Medline: 12946886]
49. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). 4th edition. Washington, DC, USA: American Psychiatric Association; 1994.
50. Trivedi MH, Rush AJ, Ibrahim HM, Carmody TJ, Biggs MM, Suppes T, et al. The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. Psychol Med 2004 Jan;34(1):73-82. [doi: 10.1017/s0033291703001107] [Medline: 14971628]
51. Bernstein IH, Rush AJ, Suppes T, Trivedi MH, Woo A, Kyutoku Y, et al. A psychometric evaluation of the clinician-rated Quick Inventory of Depressive Symptomatology (QIDS-C16) in patients with bipolar disorder. Int J Methods Psychiatr Res 2009 Jun;18(2):138-146 [FREE Full text] [doi: 10.1002/mpr.285] [Medline: 19507161]
52. Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, et al. An evaluation of the quick inventory of depressive symptomatology and the hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. Biol Psychiatry 2006 Mar 15;59(6):493-501 [FREE Full text] [doi: 10.1016/j.biopsych.2005.08.022] [Medline: 16199008]
53. Shear MK, Vander Bilt J, Rucci P, Endicott J, Lydiard B, Otto MW, et al. Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). Depress Anxiety 2001;13(4):166-178. [Medline: 11413563]
54. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [doi: 10.1111/j.2044-8341.1959.tb00467.x] [Medline: 13638508]
55. Clark DB, Donovan JE. Reliability and validity of the Hamilton Anxiety Rating Scale in an adolescent sample. J Am Acad Child Adolesc Psychiatry 1994;33(3):354-360. [doi: 10.1097/00004583-199403000-00009] [Medline: 8169180]
56. Williams JB, Kobak KA, Giller E, Reasner DS, Curry L, Detke MJ. Comparison of site-based versus central ratings in a study of generalized anxiety disorder. J Clin Psychopharmacol 2015 Dec;35(6):654-660. [doi: 10.1097/JCP.0000000000000422] [Medline: 26488677]
57. Sheehan DV, Sheehan KH, Shytle RD, Janavs J, Bannon Y, Rogers JE, et al. Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID). J Clin Psychiatry 2010 Mar;71(3):313-326. [doi: 10.4088/JCP.09m05305whi] [Medline: 20331933]
58. Levis B, Benedetti A, Riehm KE, Saadat N, Levis AW, Azar M, et al. Probability of major depression diagnostic classification using semi-structured versus fully structured diagnostic interviews. Br J Psychiatry 2018 Jun;212(6):377-385 [FREE Full text] [doi: 10.1192/bjp.2018.54] [Medline: 29717691]
59. 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]
60. Donker T, van Straten A, Marks I, Cuijpers P. Quick and easy self-rating of Generalized Anxiety Disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. Psychiatry Res 2011 Jun 30;188(1):58-64. [doi: 10.1016/j.psychres.2011.01.016] [Medline: 21339006]
61. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.r-project.org/ [accessed 2022-12-06]
62. Bürkner PC, Gabry J, Weber S, Johnson A, Modrak M, Badr HS, et al. brms: Bayesian Regression Models using 'Stan'. V. 2.17.0. The Comprehensive R Archive Network. 2022. URL: https://rdrr.io/cran/brms/ [accessed 2022-12-06]

63. Guo J, Gabry J, Goodrich B, Weber S. rstan: R Interface to Stan. v. 2.21.7. The Comprehensive R Archive Network. 2022. URL: https://cran.r-project.org/web/packages/rstan/vignettes/rstan.html [accessed 2022-12-06]
64. Wagenmakers EJ, Marsman M, Jamil T, Ly A, Verhagen J, Love J, et al. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. Psychon Bull Rev 2018 Mar;25(1):35-57 [FREE Full text] [doi: 10.3758/s13423-017-1343-3] [Medline: 28779455]
65. Verhagen J, Wagenmakers EJ. Bayesian tests to quantify the result of a replication attempt. J Exp Psychol Gen 2014 Aug;143(4):1457-1475. [doi: 10.1037/a0036731] [Medline: 24867486]
66. He J, Wang W, Huang M, Wang S, Guan X. Bayesian inference under small sample sizes using general noninformative priors. Mathematics 2021 Nov 05;9(21):2810. [doi: 10.3390/math9212810]
67. van Doorn J, van den Bergh D, Böhm U, Dablander F, Derks K, Draws T, et al. The JASP guidelines for conducting and reporting a Bayesian analysis. Psychon Bull Rev 2021 Jun;28(3):813-826 [FREE Full text] [doi: 10.3758/s13423-020-01798-5] [Medline: 33037582]
68. Lee MD, Wagenmakers EJ. Bayesian Cognitive Modeling: A Practical Course. Cambridge, UK: Cambridge University Press; 2014.
69. 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]
70. Karyotaki E, Klein AM, Ciharova M, Bolinski F, Krijnen L, de Koning L, et al. Guided internet-based transdiagnostic individually tailored Cognitive Behavioral Therapy for symptoms of depression and/or anxiety in college students: a randomized controlled trial. Behav Res Ther 2022 Mar;150:104028 [FREE Full text] [doi: 10.1016/j.brat.2021.104028] [Medline: 35066365]
71. Merikangas KR, Calkins ME, Burstein M, He JP, Chiavacci R, Lateef T, et al. Comorbidity of physical and mental disorders in the neurodevelopmental genomics cohort study. Pediatrics 2015 Apr;135(4):e927-e938 [FREE Full text] [doi: 10.1542/peds.2014-1444] [Medline: 25755242]
72. Dopmeijer JM, Nuijen J, Busch MC, Tak N. Monitor Mentale gezondheid en Middelengebruik Studenten hoger onderwijs. Deelrapport I. Mentale gezondheid van studenten in het hoger onderwijs [Screening of mental health and substance use of students in higher education. Subreport 1. Mental health of students in higher education]. Rijksinstituut voor Volksgezondheid en Milieu RIVM. 2021. URL: https://rivm.openrepository.com/handle/10029/625362 [accessed 2022-12-06]
73. Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. J Clin Epidemiol 1999 Dec;52(12):1143-1156 [FREE Full text] [doi: 10.1016/s0895-4356(99)00141-9]
74. Rosbach M, Andersen JS. Patient-experienced burden of treatment in patients with multimorbidity - a systematic review of qualitative data. PLoS One 2017 Jun 23;12(6):e0179916 [FREE Full text] [doi: 10.1371/journal.pone.0179916] [Medline: 28644877]
75. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemp Clin Trials Commun 2018 Sep;11:156-164 [FREE Full text] [doi: 10.1016/j.conctc.2018.08.001] [Medline: 30112460]
76. Kasenda B, von Elm E, You J, Blümle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA 2014 Mar 12;311(10):1045-1051 [FREE Full text] [doi: 10.1001/jama.2014.1361] [Medline: 24618966]
77. 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. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
78. Rahmadiana M, Karyotaki E, Schulte M, Ebert DD, Passchier J, Cuijpers P, et al. Transdiagnostic internet intervention for Indonesian university students with depression and anxiety: evaluation of feasibility and acceptability. JMIR Ment Health 2021 Mar 05;8(3):e20036 [FREE Full text] [doi: 10.2196/20036] [Medline: 33666553]
79. Bendig E, Bauereiß N, Buntrock C, Habibović M, Ebert DD, Baumeister H. Lessons learned from an attempted randomized-controlled feasibility trial on "WIDeCAD" - an Internet-based depression treatment for people living with coronary artery disease (CAD). Internet Interv 2021 Apr;24:100375 [FREE Full text] [doi: 10.1016/j.invent.2021.100375] [Medline: 33732627]
80. Krusche A, Rudolf von Rohr I, Muse K, Duggan D, Crane C, Williams JM. An evaluation of the effectiveness of recruitment methods: the staying well after depression randomized controlled trial. Clin Trials 2014 Apr;11(2):141-149 [FREE Full text] [doi: 10.1177/1740774514521905] [Medline: 24686105]
81. Treweek S, Pitkethly M, Cook J, Fraser C, Mitchell E, Sullivan F, et al. Strategies to improve recruitment to randomised trials. Cochrane Database Syst Rev 2018 Mar 22;2(2):MR000013 [FREE Full text] [doi: 10.1002/14651858.MR000013.pub6] [Medline: 29468635]
82. Healy P, Galvin S, Williamson PR, Treweek S, Whiting C, Maeso B, et al. Identifying trial recruitment uncertainties using a James Lind Alliance Priority Setting Partnership - the PRioRiTy (Prioritising Recruitment in Randomised Trials) study. Trials 2018 Mar 01;19(1):147 [FREE Full text] [doi: 10.1186/s13063-018-2544-4] [Medline: 29490702]

83. Lokman S, Leone SS, Sommers-Spijkerman M, van der Poel A, Smit F, Boon B. Complaint-directed mini-interventions for depressive complaints: a randomized controlled trial of unguided web-based self-help interventions. J Med Internet Res 2017 Jan 04;19(1):e4 [FREE Full text] [doi: 10.2196/jmir.6581] [Medline: 28052840]
84. Weisel KK, Zarski AC, Berger T, Krieger T, Moser CT, Schaub MP, et al. User experience and effects of an individually tailored transdiagnostic internet-based and mobile-supported intervention for anxiety disorders: mixed-methods study. J Med Internet Res 2020 Sep 16;22(9):e16450 [FREE Full text] [doi: 10.2196/16450] [Medline: 32936085]
85. Koelen JA, Vonk A, Klein A, de Koning L, Vonk P, de Vet S, et al. Man vs. machine: a meta-analysis on the added value of human support in text-based internet treatments ("e-therapy") for mental disorders. Clin Psychol Rev 2022 Aug;96:102179 [FREE Full text] [doi: 10.1016/j.cpr.2022.102179] [Medline: 35763975]
86. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, 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 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
87. Wagenmakers EJ, Love J, Marsman M, Jamil T, Ly A, Verhagen J, et al. Bayesian inference for psychology. Part II: example applications with JASP. Psychon Bull Rev 2018 Mar;25(1):58-76 [FREE Full text] [doi: 10.3758/s13423-017-1323-7] [Medline: 28685272]
88. Makowski D, Ben-Shachar MS, Lüdecke D. bayestestR: describing effects and their uncertainty, existence and significance within the Bayesian framework. J Open Source Softw 2019 Aug 13;4(40):1541 [FREE Full text] [doi: 10.21105/joss.01541]
89. R Core Team, Bivand R, Carey VJ, DebRoy S, Eglen S, Guha R. foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ... v. 0.8-83. The Comprehensive R Archive Network. 2022. URL: https://cran.r-project.org/web/packages/foreign/index.html [accessed 2022-12-06]
90. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Cham, Switzerland: Springer; 2016.
91. Harrell Jr FE. Hmisc: Harrell Miscellaneous. The Comprehensive R Archive Network. 2022 Nov 18. URL: https://cran.r-project.org/web/packages/Hmisc/index.html [accessed 2022-12-06]
92. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. R package version 2.2.3. The Comprehensive R Archive Network. 2022 Sep 29. URL: https://cran.r-project.org/web/packages/psych/index.html [accessed 2022-12-06]
93. Bolinski F. ICare Prevent NL - lessons learned. Open Science Framework. 2022 Jun 10. URL: https://osf.io/tn9yf/?view_only=4d89221c73af432a8b36a1a9c97e43f5 [accessed 2022-10-06]

Abbreviations

BF: Bayes factor
CAU: care as usual
CI: credibility interval
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GAD: generalized anxiety disorder
GAD-7: 7-item Generalized Anxiety Disorder Scale
GP: general practitioner
MDD: major depressive disorder
MINI: Mini International Neuropsychiatric Interview
PHQ-9: 9-item Patient Health Questionnaire
QIDS-CR: Quick Inventory of Depressive Symptomatology-Clinician Rated
RCT: randomized controlled trial
SIGH-A: Structured Interview Guide for the Hamilton Anxiety Rating Scale

Edited by A Mavragani; submitted 26.07.22; peer-reviewed by C Benjet, S Adam; comments to author 18.08.22; revised version received 02.09.22; accepted 13.09.22; published 14.12.22.

Please cite as:
Bolinski F, Kleiboer A, Neijenhuijs K, Karyotaki E, Wiers R, de Koning L, Jacobi C, Zarski AC, Weisel KK, Cuijpers P, Riper H Challenges in Recruiting University Students for Web-Based Indicated Prevention of Depression and Anxiety: Results From a Randomized Controlled Trial (ICare Prevent)
J Med Internet Res 2022;24(12):e40892
URL: https://www.jmir.org/2022/12/e40892
doi:10.2196/40892
PMID:36515986

Original Paper

What Patients Find on the Internet When Looking for Information About Percutaneous Coronary Intervention: Multilanguage Cross-sectional Assessment

Cristina M ulea[1], MD; Valentin Nădă an[1], MD, PhD; Tatiana Ursachi[1], MD; Paul-Cătălin Toboltoc[1,2], MD; Theodora Benedek[1], MD, PhD

[1]George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Targu Mures, Romania
[2]Lucian Blaga University of Sibiu, Sibiu, Romania

Corresponding Author:
Valentin Nădă an, MD, PhD
George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures
38 Gheorghe Marinescu Street
Targu Mures, 540142
Romania
Phone: 40 265215551
Email: valentin.nadasan@umfst.ro

Abstract

Background: The internet provides general users with wide access to medical information. However, regulating and controlling the quality and reliability of the considerable volume of available data is challenging, thus generating concerns about the consequences of inaccurate health care–related documentation. Several tools have been proposed to increase the transparency and overall trustworthiness of medical information present on the web.

Objective: We aimed to analyze and compare the quality and reliability of information about percutaneous coronary intervention on English, German, Hungarian, Romanian, and Russian language websites.

Methods: Following a rigorous protocol, 125 websites were selected, 25 for each language sub-sample. The websites were assessed concerning their general characteristics, compliance with a set of eEurope 2002 credibility criteria, and quality of the informational content (namely completeness and accuracy), based on a topic-specific benchmark. Completeness and accuracy were graded independently by 2 evaluators. Scores were reported on a scale from 0 to 10. The 5 language subsamples were compared regarding credibility, completeness, and accuracy. Correlations between credibility scores on the one hand, and completeness and accuracy scores, on the other hand, were tested within each language subsample.

Results: The websites' compliance with credibility criteria was average at best with scores between 3.0 and 6.0. In terms of completeness and accuracy, the website subsets qualified as poor or average, with scores ranging from 2.4 to 4.6 and 3.6 to 5.3, respectively. English language websites scored significantly higher in all 3 aspects, followed by German and Hungarian language websites. Only German language websites showed a significant correlation between credibility and information quality.

Conclusions: The quality of websites in English, German, Hungarian, Romanian, and Russian languages about percutaneous coronary intervention was rather inadequate and may raise concerns regarding their impact on informed decision-making. Using credibility criteria as indicators of information quality may not be warranted, as credibility scores were only exceptionally correlated with content quality. The study brings valuable descriptive data on the quality of web-based information regarding percutaneous coronary intervention in multiple languages and raises awareness about the need for responsible use of health-related web resources.

(J Med Internet Res 2022;24(12):e41219) doi:10.2196/41219

KEYWORDS

percutaneous coronary intervention; consumer health informatics; internet; health education; health information; quality; reliability; informed decision-making; credibility; content quality; medical information

Introduction

On account of its accessibility and interactivity, the internet has become a popular and widely used tool for independent medical documentation among the general public. The proportion of people who turn to the web-based environment in search of health-related information has been steadily increasing [1]. This practice, although regarded as convenient from the consumers' point of view, has raised concerns among physicians, as the quality of the web-based medical information and the patients' or caregivers' ability to select relevant information are often seen as questionable [2]. Therefore, the negligent use of the internet may impact the physician-patient relationship and consumers' medical decision-making, leading to unjustified fears (also known as 'cyberchondria'), defiance of medical advice, or inclination toward self-diagnosis and self-treatment [3-6].

The interventional treatment of coronary artery disease is one of the topics of high interest among patients and caregivers, as the condition is one of the leading causes of morbidity and mortality globally [7]. The quality and reliability of the information available online on the subject of percutaneous coronary intervention (PCI) may have a considerable impact on the general public's understanding of the technique, compliance, and outcomes of their therapeutic decisions. Hence, this study aimed to assess and compare the quality of information about PCI on a sample of English, German, Hungarian, Romanian, and Russian websites and to evaluate the reliability of credibility criteria as indicators of information quality. These aspects are hoped to provide valuable insight into the quality of web-based health information, not only on behalf of internet users and medical practitioners but also website owners and policy makers, to serve as foundation for the effective education of the general population on the topic of internet health-related documentation.

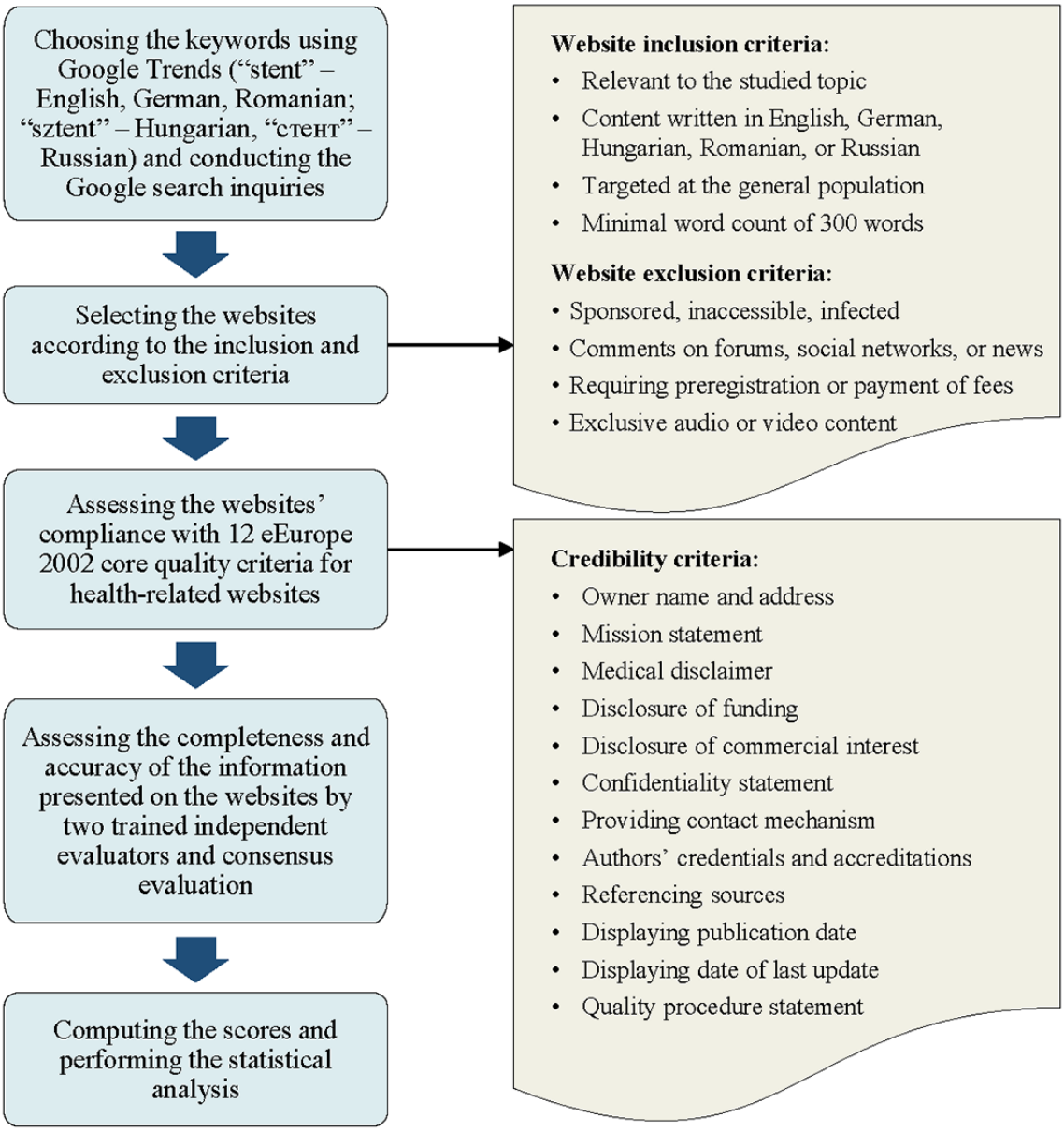
Methods

Sample Selection

The research was designed as an observational cross-sectional study. Its sample consisted of 125 PCI-related websites intended for the general population in 5 languages (25 for each included language)—English, German, Hungarian, Romanian, and Russian. The Google Search engine was used to identify eligible websites, using "stent" as query term (used as such in English, German, and Romanian; "sztent" in Hungarian; and "стент" in Russian). The query terms were selected based on their popularity as shown by Google Trends, a tool that analyzes the frequency of top search queries in Google Search across various regions and languages. The links returned by the Google Search engine were screened according to a set of preestablished inclusion and exclusion criteria. To be included, a website had to address the subject of coronary stenting, presenting the information in the desired language and a minimum of 300 words. The information had to be targeted at internet users without medical education. Pages addressing subjects other than PCI, sponsored pages appearing in the top hierarchical positions in the results list, and infected or inaccessible pages were excluded. Websites consisting exclusively of audio or video content and websites allowing access only after registration or payment of a fee were also excluded. Similarly, web pages that presented the topic of interest in the form of news or comments on forums and social networks—in other words, pieces of information not meant to thoroughly present the subject of PCI—were not included in the sample. Websites deemed fit for inclusion were consecutively analyzed following a rigorous protocol, briefly illustrated in Figure 1.

The Google searches were performed in April 2019 for English, Hungarian, and Romanian language websites, while the Russian language inquiry took place in June 2019. German language websites were subsequently added to the study in November 2020. A total number of 83, 39, 121, 167, and 94 websites in English, German, Hungarian, Romanian, and Russian language, respectively, were screened until the acquisition of the 25 eligible links for each of the subsamples.

Figure 1. Flowchart representing the main steps of the study.



Data Collection

Firstly, the examination was aimed at the websites' general characteristics and their compliance with 12 general credibility criteria derived from the eEurope 2002 core quality criteria for health-related websites supported by the Commission of the European Communities [8] (Figure 1). Next, the selected pages underwent an exhaustive evaluation of their informational content based on a topic-specific benchmark (Multimedia Appendix 1) [9]. The benchmark was developed using published literature and evidence-based guidelines on the subject of interest as sources of information, in such a way that it covered the topic of PCI to an extent considered sufficient and comprehensible for nonprofessionals. It included information on the following aspects: definitions and introductory notions about PCI, types of coronary stents, indications for the procedure, preprocedural preparation of the patient, description of the procedure, the postprocedural period, what to know or do at home, risks, benefits, costs, other treatment options, general prevention and prophylaxis methods, as well as general warnings regarding alternative treatments. To ensure a practical grading manner, the benchmark was divided logically into 50 items. Their presence on the studied websites was evaluated regarding completeness (ie, the presence of the item on the studied website, evaluated in a binary fashion) and accuracy (ie, the extent to which the item was correctly presented on the studied website, graded on a 3-point scale). The benchmark was reviewed by medical professionals, specialists in the fields of cardiology and interventional radiology, from both Romania and the United States.

The data on the websites' general characteristics and compliance with credibility criteria were collected by one operator, while the assessments regarding the websites' informational content

were performed independently by 2 evaluators for all websites. The websites' compliance with the 12 selected eEurope 2002 credibility criteria was assessed in a binary fashion, with 1 point given to every criterion that was met. Based on the obtained sum, the relative credibility score of the given web page was calculated as previously described by Nădăşan et al [9]. Similarly, based on the points awarded for completeness and accuracy for each website, their relative completeness and accuracy scores were computed. All relative scores were reported on a 0-10 scale. The resulting scores were categorized as very poor (0-2), poor (2.1-4), average (4.1-6), good (6.1-8), or very good (8.1-10). The analyzed data are available in Multimedia Appendix 2.

Statistical Analysis

For each included website, the degree of agreement between the 2 evaluators was assessed using the Cohen kappa statistic, a test that measures interrater reliability and is regarded as more robust than simply computing the percentage of agreement, as it adjusts for agreement occurring by chance. Kappa coefficients may range from −1 to 1. A kappa value of 1 indicates perfect agreement, while a value of 0 corresponds to the rate of agreement expected by chance alone. In our study, a coefficient of less than 0.8 prompted a reevaluation to reach a consensus.

The Kolmogorov-Smirnov test was used to analyze the normality of the data, based on which the comparisons of data with normal and nonnormal distributions were performed using the 2-tailed Student *t* test and the 2-tailed Mann-Whitney test, respectively. The correlations between credibility scores on the one hand and completeness and accuracy scores on the other hand were analyzed using the Spearman rank correlation test.

The statistical analyses were conducted using IBM SPSS Statistics for Windows (version 22.0; IBM Corp). The threshold value for statistical significance was set at a value of α=.05. The obtained scores are presented as mean (SD).

Results

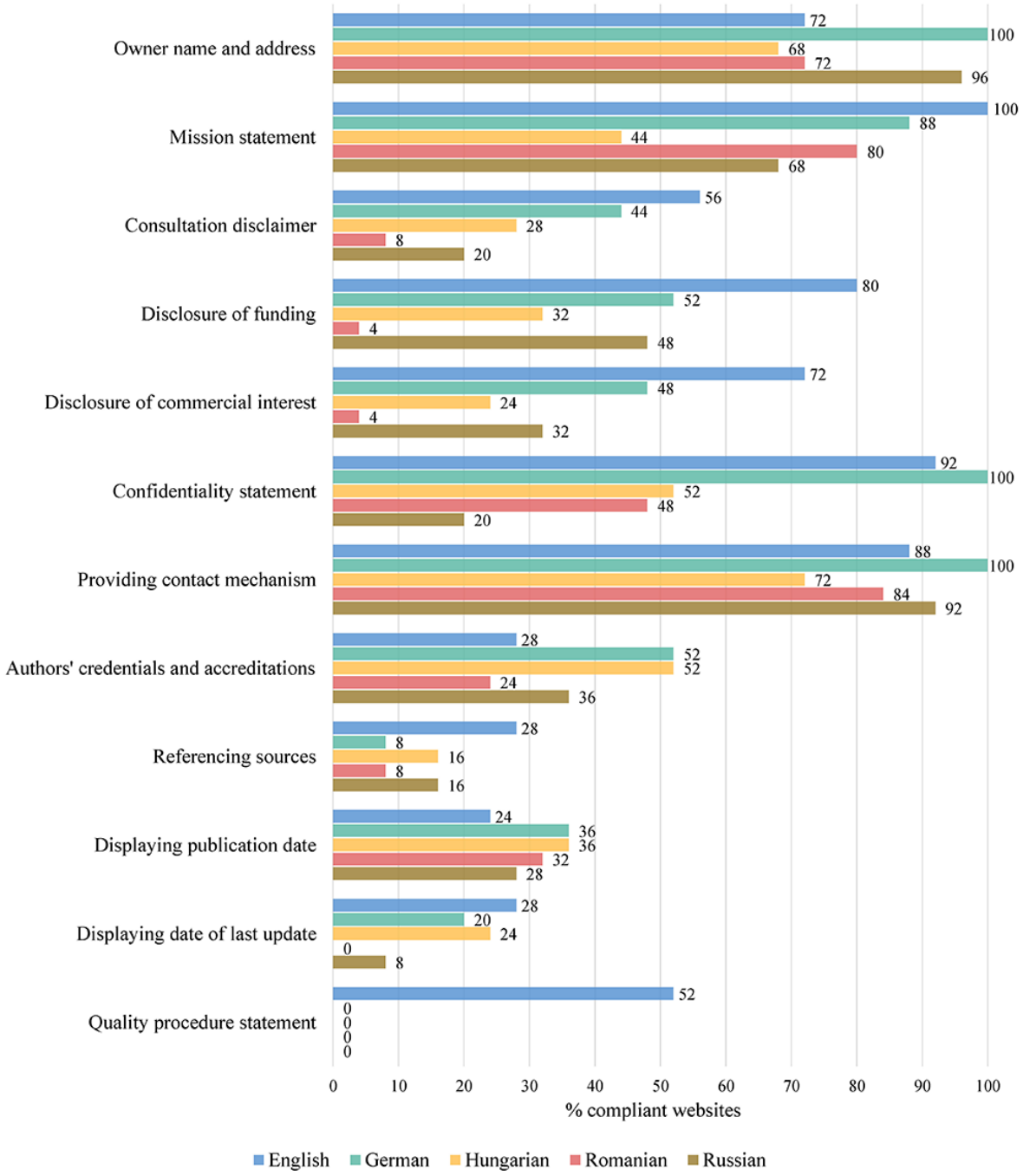
Of the 125 included websites, nearly two-thirds had a general medical approach, comprising information belonging to multiple medical specialties. Most of the web pages were owned by private or state medical service providers. In terms of purpose, the pages were predominantly educational. As far as their format was concerned, the most often identified were company presentation pages. Most websites were characterized by a conventional medicine approach. The detailed distribution of the studied websites according to their general characteristics is shown in Table 1.

Table 1. The absolute (n) and relative (%) frequencies of the websites based on their general characteristics.

General characteristics	Values, n (%)
Specialization	
Single medical specialty	40 (32)
Multiple medical specialties	85 (68)
Website ownership	
Foundation or nongovernmental organization	15 (12)
Private or state health care provider	43 (34.4)
Commercial company	23 (18.4)
Manufacturer or distributor of medical supplies and equipment	6 (4.8)
Private person	3 (2.4)
Educational or research institution	9 (7.2)
Unidentifiable	26 (20.8)
Main purpose	
Educational	68 (54.4)
Commercial	48 (38.4)
Socialization or support	9 (7.2)
Website format	
Thematic	8 (6.4)
Medical or general portal	37 (29.6)
Electronic publication	21 (16.8)
Company presentation page or web-based shop	52 (41.6)
Blog or personal page	3 (2.4)
Other	4 (3.2)
Medical paradigm	
Conventional medicine	108 (86.4)
Mixed (ie, alternative and conventional) approach	5 (4)
Unidentifiable	12 (9.6)

The websites' overall compliance with the selected eEurope 2002 credibility criteria was highly variable, with some criteria being fulfilled to a greater extent (providing a direct contact mechanism: 87.2%; including the owner's name and address: 81.6%; and providing a mission statement: 76%), while others were identified on few of the included web pages (providing a quality procedure statement: 10.4%; including referencing sources: 15.2%; and displaying the date of last update: 16%). The remaining credibility criteria were identified on approximately one to two-thirds of the studied pages (displaying the publication date of the articles and a consultation disclaimer: 31.2% each; including disclosure of commercial interest: 36%; including the authors' credentials and accreditations: 38.4%; providing a declaration of funding: 43.2%; and offering a confidentiality statement: 62.4%). Figure 2 illustrates the compliance of each of the 5 groups of websites with the selected credibility criteria.

Figure 2. The websites' level of compliance with the credibility criteria by language subsample.



The overall mean relative credibility, completeness, and accuracy scores were 4.4 (SD 2.2), 3.2 (SD 1.6), and 4.3 (SD 1.6), respectively. The mean scores of the 5 language subsamples are summarized in Figure 3. The results of the comparison and correlation tests are presented in Table 2 and Table 3, respectively.

Figure 3. The websites' relative credibility, completeness, and accuracy scores by language subsample.

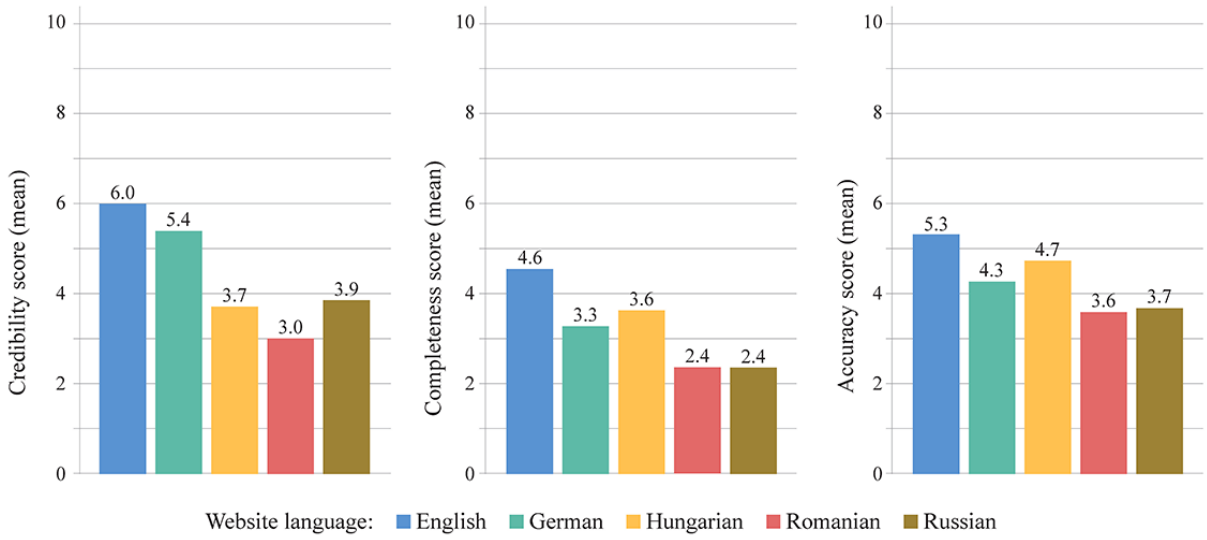


Table 2. The results of the comparison tests between language subsamples regarding the relative credibility score (RQS), relative completeness score (RCS), and relative accuracy score (RAS). *P* values with statistical significance are emphasized in italics.

Compared language subsamples	*P* values		
	RQS	RCS	RAS
English vs German	.19[a]	*.002* [a]	*.004* [a]
English vs Hungarian	*.001* [a]	*.04* [a]	.11[a]
English vs Romanian	*<.001* [b]	*<.001* [b]	*<.001* [a]
English vs Russian	*<.001* [b]	*<.001* [a]	*.001* [a]
German vs Hungarian	*.01* [a]	.33[a]	.18[a]
German vs Romanian	*<.001* [b]	*.01* [b]	.06[a]
German vs Russian	*<.001* [b]	*.01* [a]	.18[a]
Hungarian vs Romanian	.37[b]	*.002* [b]	*.004* [a]
Hungarian vs Russian	.94[b]	*.002* [a]	*.02* [a]
Romanian vs Russian	.16[b]	.94[b]	.88[a]

[a]Student *t* test.

[b]Mann-Whitney test.

Table 3. The results of the Spearman rank correlation tests between the credibility scores and informational content quality scores for all the language subsamples (*P* values with statistical significance are emphasized in italics).

Language subsamples	Tested variables	Correlation coefficient	*P* values
English			
	Credibility vs completeness	0.3439	.09
	Credibility vs accuracy	0.0578	.78
German			
	Credibility vs completeness	0.4672	*.01*
	Credibility vs accuracy	0.3994	*.04*
Hungarian			
	Credibility vs completeness	–0.3507	.08
	Credibility vs accuracy	–0.1940	.35
Romanian			
	Credibility vs completeness	–0.2592	.21
	Credibility vs accuracy	–0.3169	.12
Russian			
	Credibility vs completeness	0.1280	.54
	Credibility vs accuracy	–0.0645	.75

Discussion

Principal Findings

As far as the websites' credibility is concerned, the obtained scores were average at best, with English and German language websites acquiring the highest results (mean 6.0, SD 1.8 and mean 5.4, SD 1.3, respectively), significantly higher than those of Hungarian, Romanian, and Russian language websites, which were graded as poor. Although certain criteria (eg, displaying the owner's name and contact information as well as providing a direct contact modality) were largely met by web pages in all 5 languages, other criteria (eg, providing authors' credentials and accreditations, bibliographic references, articles' dates of publication and last update, or offering a quality procedure statement) were scarcely included. This may raise a red flag since authorship and providing references are perceived as important indicators of medical information reliability. Apparently, Hungarian, Romanian, and Russian language websites' owners may not pay enough attention to credibility or are unaware of this aspect. These findings are consistent with previously published literature investigating the credibility of web-based information about different medical topics [10-12]. Moreover, compliance with the credibility criteria for health-related websites as measured by the Health on the Net Code of Conduct has been shown to vary largely depending on the type of organization and health conditions [13].

Regarding the evaluation of the websites' informational content, the completeness and accuracy of the data about PCI in the 5 studied languages were found to be rather unsatisfactory, with the obtained scores only getting average and poor labels. In terms of completeness, English language websites acquired the highest scores (mean 4.6, SD 1.6), significantly higher than those of the websites in the other 4 language subsamples. German and Hungarian language websites also performed significantly better than the Romanian and Russian websites. In terms of accuracy of data, English language websites had significantly higher scores than German, Romanian, and Russian but not Hungarian websites, which had significantly higher accuracy scores than Romanian and Russian websites. A relative superiority of English language health-related websites compared to Spanish ones has been observed as early as 2001 in a study covering multiple medical conditions (ie, breast cancer, depression, obesity, and childhood asthma), and it was more recently compared to Turkish websites focusing on an orthopaedic intervention [14,15]. Leaving aside methodological differences, the results of these studies call attention to possible language-mediated inequities and suggest that a multilingual approach to web-based documentation may provide more complete coverage of the topic. Apparently, in some countries such as Romania, the low quality of web-based information about PCI seems to be in line with the low number of PCIs per million individuals, as shown by the latest statistics published by the European Society of Cardiology [16]. Efforts to increase the quality of web-based information about PCI would be a reasonable step in countries where access to these interventions is wanting.

The correlation assessments did not find statistically significant relationships between the credibility of the PCI-related websites and the quality of their informational content, with one exception. In the case of German language websites, the compliance with credibility criteria exhibited statistically significant, moderate strength correlations with the websites' coverage of the topic (ie, completeness and accuracy). The lack of consistent correlations between credibility and content quality has been previously reported in investigations focusing on various medical conditions, such as stroke and depression or procedures such as first aid instructions in case of choking

[17-19]. The results may suggest that the selected credibility criteria are not reliable indicators of information quality on PCI-related websites in the studied languages, and therefore, cannot be recommended to nonprofessionals as marks of trustworthiness.

Inferences

Despite the growing demand for web-based medical information, the recognition of the importance of patient participation in medical decision-making, and the impact of health-related web content on consumer health [20,21], the credibility and quality of websites about PCI—a procedure globally used to mitigate the consequences of the most common type of heart disease—has not yet been rigorously analyzed. To the best of our knowledge, this is the first study analyzing and comparing the quality of information about PCI on English, German, Hungarian, Romanian, and Russian websites aimed at the general population. The results of this study may be used to raise awareness among internet users about the limitations and potential hazards of using the web as a source of information about PCI. Engaging in safe internet browsing is crucial, as it may prevent poor decision-making and potential complications caused by delayed intervention as well as a deterioration of the physician-patient relationship [20,22]. Although the web-based environment is easily accessible and convenient, it is highly advisable that consumers engage in web-based medical documentation with precaution and always turn to medical professionals for advice. Furthermore, medical practitioners should fully acknowledge the reality of e-patients and handle it appropriately [23,24]. In this regard, the involvement of health professionals in the development of plain language and accurate web-based health resources and their involvement in providing guidance to patients with inadequate health literacy in accessing proper information on the internet could prove beneficial [25].

Strengths and Limitations

It is worth noting the strengths of this study. First, the inclusion of multiple languages, of which at least three are spoken by vast numbers of individuals worldwide, allows for the extrapolation of the results and recommendations to a large population. For instance, according to Ethnologue [26], English is the most widely used language around the globe, being spoken by approximately 1.5 billion people across more than 140 countries. Moreover, both Russian and German are among the top 15 most widely spoken languages, with nearly 260 and 135 million speakers, respectively.

Second, most of the previously published studies focus on assessing health-related web-based sources based on credibility (reliability), readability, or design criteria (eg, the Health on the Net Code, JAMA score, DISCERN score, Flesch-Kincaid readability test, and SMOG Readability Index) [27]. As acknowledged by the authors of the DISCERN instrument [28], not even this tool was designed to actually measure the scientific fidelity of the information. Our study addresses not only the credibility or reliability dimension but also the quality of the content by evaluating the completeness and accuracy of information based on an evidence-based, topic-specific quality benchmark.

Third, to minimize subjectivity and the human error factor, the content quality assessments were conducted by 2 independent evaluators.

The main limitations of the study are related to some inherent traits of web-based research. Internet users may turn to various search engines or use different keywords, consecutively obtaining different search results [29]. Moreover, the continuously changing dynamics of the web-based environment make it virtually impossible for the results of this study to be precisely replicated. Additionally, the sample size may be argued as small. However, most internet users limit their inquiries to the first Google Search results page (on average, the first 10 search results) [30]. Therefore, by simulating a popular search strategy among lay internet users, we are confident that our results are likely to reflect common experiences. The study brings valuable descriptive data on the quality of web-based PCI-related information in multiple languages and has the potential to raise awareness about the need for responsible use of health-related web resources.

Conclusions

The quality and reliability of the web-based information about PCI on English, German, Hungarian, Romanian, and Russian websites are rather unsatisfactory, and there are significant differences in the quality of information across the studied languages. It is safe to say that the internet does not provide the general public with good-quality medical information on the aforementioned topic. Moreover, the selected credibility criteria cannot be recommended as consistent indicators of information quality. Further efforts ought to be made by website developers to improve the trustworthiness of web-based health information. Since medical websites have become one of the most trusted sources of health-related documentation, it is crucial to both parties involved in the medical act (ie, lay users and medical practitioners) to develop awareness of the potential dangers internet documentation may pose. Our results could contribute to advances in the fields of preventive medicine or public health, supporting the importance of internet education among the general population.

Acknowledgments

The authors wish to thank Steven S Maxood, MD, PhD, cardiologist at Providence Saint Mary Heart and Vascular Center, Walla Walla, WA, United States; and László Hadadi, MD, PhD, cardiologist at Targu Mures County Emergency Clinical Hospital, Romania, for their input in the elaboration of the benchmark. We would also like to thank Márta Szabó, MD; László Szabó, MD; and Teodora Sibinovska, MD, at the time medical students at the George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, for their contribution to data acquisition.

This research received no specific grant from any funding agency in the public, commercial, or nonprofit sectors.

Authors' Contributions

CM contributed to the conception of the study; the elaboration of the benchmark; the acquisition, analysis, and interpretation of data; and drafted the manuscript. VN designed and coordinated the study, contributed to the analysis and interpretation of data, and critically revised the manuscript. TU and PCT were involved in the acquisition of data and contributed to the drafting of the manuscript. TB contributed to the elaboration of the benchmark and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topic-specific benchmark for assessing the quality of information about percutaneous coronary intervention.
[PDF File (Adobe PDF File), 499 KB - jmir_v24i12e41219_app1.pdf]

Multimedia Appendix 2
Quality assessment data and the results of statistical analyses.
[XLSX File (Microsoft Excel File), 154 KB - jmir_v24i12e41219_app2.xlsx]

References

1. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
2. Kim J, Kim S. Physicians' perception of the effects of Internet health information on the doctor-patient relationship. Inform Health Soc Care 2009 Sep 13;34(3):136-148. [doi: 10.1080/17538150903102422] [Medline: 19670004]
3. Fox S, Duggan M. Health Online 2013. Pew Research Center. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2022-11-18]
4. Ahmad F, Hudak PL, Bercovitz K, Hollenberg E, Levinson W. Are physicians ready for patients with Internet-based health information? J Med Internet Res 2006 Sep 29;8(3):e22 [FREE Full text] [doi: 10.2196/jmir.8.3.e22] [Medline: 17032638]
5. White RW, Horvitz E. Cyberchondria. ACM Trans Inf Syst 2009 Nov 01;27(4):1-37. [doi: 10.1145/1629096.1629101]
6. Starcevic V, Berle D. Cyberchondria: towards a better understanding of excessive health-related Internet use. Expert Rev Neurother 2013 Feb;13(2):205-213. [doi: 10.1586/ern.12.162] [Medline: 23368807]
7. Khan MA, Hashim MJ, Mustafa H, Baniyas MY, Al Suwaidi SKBM, AlKatheeri R, et al. Global epidemiology of Ischemic heart disease: results from the global burden of disease study. Cureus 2020 Jul 23;12(7):e9349 [FREE Full text] [doi: 10.7759/cureus.9349] [Medline: 32742886]
8. Commission OTECB. eEurope 2002: quality criteria for health related websites. J Med Internet Res 2002 Dec;4(3):E15 [FREE Full text] [doi: 10.2196/jmir.4.3.e15] [Medline: 12554546]
9. Nădăşan V, Roşca AN, Tarcea M, Ábrám Z, Măruşteri M. The quality of Romanian breast cancer websites: a five-year longitudinal assessment. J Cancer Educ 2018 Jun;33(3):703-707. [doi: 10.1007/s13187-016-1145-8] [Medline: 27888472]
10. Lovett J, Gordon C, Patton S, Chen CX. Online information on dysmenorrhoea: an evaluation of readability, credibility, quality and usability. J Clin Nurs 2019 Oct 20;28(19-20):3590-3598 [FREE Full text] [doi: 10.1111/jocn.14954] [Medline: 31162870]
11. Weber AS, Verjee M, Rahman ZH, Ameerudeen F, Al-Baz N. Typology and credibility of Internet health websites originating from Gulf Cooperation Council countries. East Mediterr Health J 2015 Feb 02;20(12):804-811 [FREE Full text] [Medline: 25664519]
12. Lawrentschuk N, Sasges D, Tasevski R, Abouassaly R, Scott AM, Davis ID. Oncology health information quality on the Internet: a multilingual evaluation. Ann Surg Oncol 2012 Mar;19(3):706-713. [doi: 10.1245/s10434-011-2137-x] [Medline: 22146882]
13. 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 21;34(9):1884-1891 [FREE Full text] [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
14. 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. JAMA 2001 May 23;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]
15. Küçükdurmaz F, Mutlu S, Mutlu H, Parvizi J. A comparison of the quality of online information about total knee arthroplasty available in turkish and english:a cross-sectional study. Acta Orthop Traumatol Turc 2015;49(4):370-374. [doi: 10.3944/aott.2015.14.0291] [Medline: 26312462]
16. Timmis A, Townsend N, Gale C, Torbica A, Lettino M, Petersen SE, European Society of Cardiology. European society of cardiology: cardiovascular disease statistics 2019. Eur Heart J 2020 Jan 01;41(1):12-85. [doi: 10.1093/eurheartj/ehz859] [Medline: 31820000]

17. Nadasan V, Vancea G, Georgescu PA, Tarcea M, Abram Z. The credibility, completeness and accuracy of information about first aid in case of choking on the Romanian websites. JAQM 2011;6(3):18-26 [FREE Full text]
18. Griffiths KM, Christensen H. Quality of web based information on treatment of depression: cross sectional survey. BMJ 2000 Dec 16;321(7275):1511-1515 [FREE Full text] [doi: 10.1136/bmj.321.7275.1511] [Medline: 11118181]
19. Popescu S, Popescu A, Maior D, Dănilă M, Dobria M, Nădă an V. Widely accepted credibility criteria for online health-related information are not correlated with content quality of stroke webpages in two languages of Central and Eastern European countries. JIM 2018;3(3):196-201 [FREE Full text] [doi: 10.2478/jim-2018-0031]
20. Couper MP, Singer E, Levin CA, Fowler FJ, Fagerlin A, Zikmund-Fisher BJ. Use of the Internet and ratings of information sources for medical decisions: results from the DECISIONS survey. Med Decis Making 2010 Sep;30(5 Suppl):106S-114S. [doi: 10.1177/0272989X10377661] [Medline: 20881159]
21. Smith R, Devine P, Jones H, DeNittis A, Whittington R, Metz J. Internet use by patients with prostate cancer undergoing radiotherapy. Urology 2003 Aug;62(2):273-277. [doi: 10.1016/s0090-4295(03)00251-6] [Medline: 12893334]
22. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the Internet on health care and the physician-patient relationship: national U.S. survey among 1.050 U.S. physicians. J Med Internet Res 2003;5(3):e17 [FREE Full text] [doi: 10.2196/jmir.5.3.e17] [Medline: 14517108]
23. Ferguson T, Frydman G. The first generation of e-patients. BMJ 2004 May 15;328(7449):1148-1149 [FREE Full text] [doi: 10.1136/bmj.328.7449.1148] [Medline: 15142894]
24. 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]
25. Arif N, Ghezzi P. Quality of online information on breast cancer treatment options. Breast 2018 Feb;37:6-12. [doi: 10.1016/j.breast.2017.10.004] [Medline: 29040893]
26. Eberhard D, Simons G, Fennig CD. Ethnologue: languages of the world. Ethnologue. 2022. URL: http://www.ethnologue.com [accessed 2022-10-27]
27. Battineni G, Baldoni S, Chintalapudi N, Sagaro GG, Pallotta G, Nittari G, et al. Factors affecting the quality and reliability of online health information. Digit Health 2020 Aug 30;6:2055207620948996 [FREE Full text] [doi: 10.1177/2055207620948996] [Medline: 32944269]
28. Background to discern. DISCERN. URL: http://www.discern.org.uk/background_to_discern.php [accessed 2022-11-06]
29. Dy CJ, Taylor SA, Patel RM, McCarthy MM, Roberts TR, Daluiski A. Does the quality, accuracy, and readability of information about lateral epicondylitis on the internet vary with the search term used? Hand (N Y) 2012 Dec 30;7(4):420-425 [FREE Full text] [doi: 10.1007/s11552-012-9443-z] [Medline: 24294163]
30. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]

Abbreviations

PCI: percutaneous coronary intervention

Edited by G Eysenbach; submitted 19.07.22; peer-reviewed by I Iqbal, E Nichele; comments to author 27.09.22; revised version received 07.11.22; accepted 17.11.22; published 06.12.22.

Please cite as:
ulea CM, Nădă an V, Ursachi T, Toboltoc PC, Benedek T
What Patients Find on the Internet When Looking for Information About Percutaneous Coronary Intervention: Multilanguage Cross-sectional Assessment
J Med Internet Res 2022;24(12):e41219
URL: https://www.jmir.org/2022/12/e41219
doi:10.2196/41219
PMID:36472906

Original Paper

Building Digital Literacy in Older Adults of Low Socioeconomic Status in Singapore (Project Wire Up): Nonrandomized Controlled Trial

Nerice Heng Wen Ngiam[1,2,3*], MBBS; Wan Qi Yee[2*], BSSc (Hons); Nigel Teo[2], MPH; Ka Shing Yow[1,4], MBBS; Amrish Soundararajan[1,5], MBBS; Jie Xin Lim[1,6], PhD; Haikel A Lim[1,7,8], BSSc (Hons), MSc, MD; Angeline Tey[1,2,9], MBBS, MMed; Kai Wen Aaron Tang[1,2,7], MBBS; Celine Yi Xin Tham[1,10], BSSc (Hons); Jamaica Pei Ying Tan[2], BSSc, MPH; Si Yinn Lu[11], MPH; Sungwon Yoon[12,13], MPH, PhD; Kennedy Yao Yi Ng[1,2,14*], MBBS, MMed; Lian Leng Low[2,13,15,16*], MBBS, MMed, MCI

[1]TriGen — Trigenerational Homecare, Singapore, Singapore
[2]Population Health and Integrated Care Office, Singapore General Hospital, Singapore, Singapore
[3]Department of Internal Medicine, Singapore General Hospital, Singapore, Singapore
[4]Department of Internal Medicine, National University Health System, Singapore, Singapore
[5]Department of Family Medicine, National University Health System, Singapore, Singapore
[6]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore
[7]Department of Psychiatry, National Healthcare Group, Singapore, Singapore
[8]Medical Education Office, Duke-NUS Medical School, Singapore, Singapore
[9]Department of Respiratory and Critical Care Medicine, Tan Tock Seng Hospital, Singapore, Singapore
[10]Medical Social Services, Ng Teng Fong General Hospital, Singapore, Singapore
[11]Research and Translational Innovation Office, SingHealth Community Hospitals, Singapore, Singapore
[12]Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore
[13]Centre for Population Health Research and Implementation, SingHealth Regional Health System, Singapore, Singapore
[14]Division of Medical Oncology, National Cancer Centre Singapore, Singapore, Singapore
[15]SingHealth Duke-NUS Department of Family Medicine, Singapore, Singapore
[16]Outram Community Hospital, SingHealth Community Hospitals, Singapore, Singapore
[*]these authors contributed equally

Corresponding Author:
Lian Leng Low, MBBS, MMed, MCI
Population Health and Integrated Care Office
Singapore General Hospital
10 Hospital Boulevard, Level 7
Singapore, 168582
Singapore
Phone: 65 69703018
Email: low.lian.leng@singhealth.com.sg

Abstract

Background: In a rapidly digitalizing world, the inability of older adults to leverage digital technology has been associated with weaker social connections and poorer health outcomes. Despite the widespread digital adoption in Singapore, older adults, especially those of lower socioeconomic status (SES), still face difficulties in adopting information and communications technology and are typically digitally excluded.

Objective: We aimed to examine the impact of the volunteer-led, one-on-one, and home-based digital literacy program on digital literacy and health-related outcomes such as self-reported loneliness, social connectedness, quality of life, and well-being for older adults of low SES.

Methods: A nonrandomized controlled study was carried out in Singapore between July 2020 and November 2021 involving 138 digitally excluded community-dwelling older adults aged ≥55 years and of lower SES. Older adults awaiting participation

in the program served as controls. Older adults under the intervention were equipped with a smartphone and cellular data, underwent fortnightly to monthly digital literacy training with volunteers to learn digital skills, and digitally connected to their existing social networks. Primary outcome was the improvement in self-reported digital literacy. Secondary outcomes included improvements in University of California, Los Angeles 3-item loneliness scale, Lubben Social Network Scale-6, EQ-5D-3L and EQ visual analogue scale scores, and Personal Wellbeing Score.

Results: There were significant improvements in digital literacy scores in the intervention group as compared to controls (mean difference 2.28, 95% CI 1.37-3.20; *P*<.001). Through multiple linear regression analyses, this difference in digital literacy scores remained independently associated with group membership after adjusting for differences in baseline scores, age, gender, education, living arrangement, housing type, and baseline social connectivity and loneliness status. There was no statistically significant difference in University of California, Los Angeles 3-item loneliness scale, Lubben Social Network Scale-6, Personal Wellbeing Score, or EQ-5D Utility and visual analogue scale score.

Conclusions: This study adds to the growing research on digital inclusion by showing that a volunteer-led, one-on-one, and home-based digital literacy program contributed to increase digital literacy in older adults of low SES. Future studies should look into developing more older adult–friendly digital spaces and technology design to encourage continued digital adoption in older adults and, eventually, impact health-related outcomes.

(J Med Internet Res 2022;24(12):e40341) doi:10.2196/40341

KEYWORDS

digital literacy; health determinants; COVID-19 pandemic; social distancing; older adults; loneliness; social connectedness; quality of life; well-being; digital inclusivity; web-based; information and communication technology

Introduction

Background

In today's rapidly digitalizing world, more than one-third of the population remain digitally unconnected [1], of which older adults have the least presence on the internet despite the biopsychosocial benefits brought about by digital technology [2,3]. A review of information and communication technology (ICT) interventions on reducing social isolation in older adults has demonstrated a positive impact on social support, social connectedness, and social isolation [4]. Moreover, ICT use also has a positive impact on health and well-being by contributing to fewer depressive symptoms and higher self-rated health and subjective well-being in older adults [5,6], with these relationships mediated by reduced loneliness [7]. Conversely, the lack of digital literacy can affect older adults' ability to access health resources and is associated with social isolation and poorer health outcomes [8-11]. At the same time, results from some studies imply that ICT use might not always be related to improved mood, quality of social relationships, and well-being [12,13].

Notwithstanding, digitally exclusion in older adults is often correlated with lower socioeconomic status (SES) [14]. This is further exacerbated by the COVID-19 pandemic, where physical distancing measures and pandemic control policies have contributed to increased social isolation and loneliness [15-17] and were associated with adverse outcomes such as depression, social anxiety, cognitive impairment, and early mortality [18,19]. Given the promising positive impact of ICT use and the negative impact of digital exclusion on loneliness, social connectedness, health, and well-being for older adults, it is of pertinent interest to better understand how digital literacy can be effectively improved among older adults in Singapore and investigate its impact on health-related outcomes, especially in the midst of the COVID-19 pandemic.

Digital Adoption in Singapore

Despite the widespread digital adoption in Singapore, older adults still face difficulties in adopting ICT [20] due to psychosocial or socioeconomic reasons. As an effort to improve digital literacy among older adults in Singapore, the Infocomm Media Development Authority in Singapore launched the Seniors Go Digital Program [21]. The program was designed to address digital access by providing subsidized smartphones and mobile data subscriptions to older adults of low SES. However, the program had a lower-than-expected impact on their target group as the program was put together rapidly to meet the urgent needs during the pandemic, but there were still a substantial number of older adults not reached [22]. Insights from older adults' learning have suggested personalized approaches in a home environment to best encourage disadvantaged older adults to participate in learning [23].

Objective

Cognizant that a more deliberate approach was needed to reach out to older adults of lower SES who are digitally excluded, TriGen, a voluntary organization, and the Singapore General Hospital collaborated with Infocomm Media Development Authority and senior activity centers in Singapore on a volunteer-led, one-on-one, and home-based digital literacy program, Project Wire Up. The pilot study aimed to contribute to the international literature on digital literacy and learning in older adults by examining the impact of the home-based digital literacy building program on (1) digital literacy and (2) self-reported loneliness, social connectedness, as well as other health-related outcomes such as quality of life and subjective well-being for older adults. We hypothesized that the program would result in (1) improved digital literacy and (2) reduced perceived loneliness and improved social connectivity, quality of life, and well-being in older adults.

Methods

Intervention: Project Wire Up

Project Wire Up was a volunteer-led, one-on-one, goal-directed, and home-based digital literacy program. The program adopted a three-pronged approach: older adults were (1) equipped with smartphones and internet connection; (2) trained by volunteers for 6 sessions (1 to 2 hours per session) over 3 months that were held in the older adults' homes; and (3) digitally connected to existing social networks. Working with national agencies, under the program, older adults of lower SES who were not digitally equipped could purchase a smartphone at a one-off price of US $15 and a 1-year mobile data plan at US $4 per month. Digital skills training was conducted during the home visits by trained volunteers, who guided older adults through a tiered curriculum of increasing difficulty that could be tailored to the needs of older adults. At the base level, older adults were taught the basic use of the phone, such as making calls and sending messages, before progressing to other social telecommunication platforms (eg, *WhatsApp*) or entertainment platforms (eg, *YouTube*). More digitally savvy older adults were taught advanced smartphone functions such as accessing government websites, making purchases, or paying bills on the web [24]. At the end of the program, older adults would be connected to existing formal and informal networks through platforms such as mobile communication apps. Supplementary materials have been provided in Multimedia Appendix 1.

Participant and Recruitment

A nonrandomized, waitlist-controlled design was carried out between July 2020 and November 2021 to evaluate the effects of a volunteer-led, one-on-one, and home-based digital literacy program among older adults of lower SES residing in Singapore. The inclusion criteria for the study were as follows: residents in the southeast region of Singapore; aged >55 years; belonging to lower SES (as indicated by residency in public rental housing or recipient of Public Assistance Scheme [25], which usually requires a per capita monthly household income of US $477 or less); and agreeable to partake in the digital literacy program for at least 2 visits or more. These older adults were generally digitally excluded. Our study intentionally reached out to these digitally excluded individuals by working with local older adult care service providers and community partners. The recruitment of participants involved phone calls and door-to-door outreach.

Upon agreement to join the program, participants were assigned to either the intervention or control arm using convenience sampling based on the referral timing to the program. For intervention group participants, baseline data were collected before exposure to the intervention, and follow-up data were collected after the completion of the intervention, which typically lasted for 3 months. As the study used a waitlist design, older adults in both the intervention group and control group were enrolled in the program, but for the control group, the baseline data were collected at the time of referral, and follow-up data were collected before exposure to the intervention, approximately 4 weeks after baseline data collection weeks (median 27, IQR 22-43 days; Figure 1).

Participants recruited from July 2020 to November 2020 were assigned to the intervention arm, whereas participants recruited from November 2020 to July 2021 were assigned to the control arm (Figure 2).

Figure 1. Overview of the participant's journey with relation to data collection.

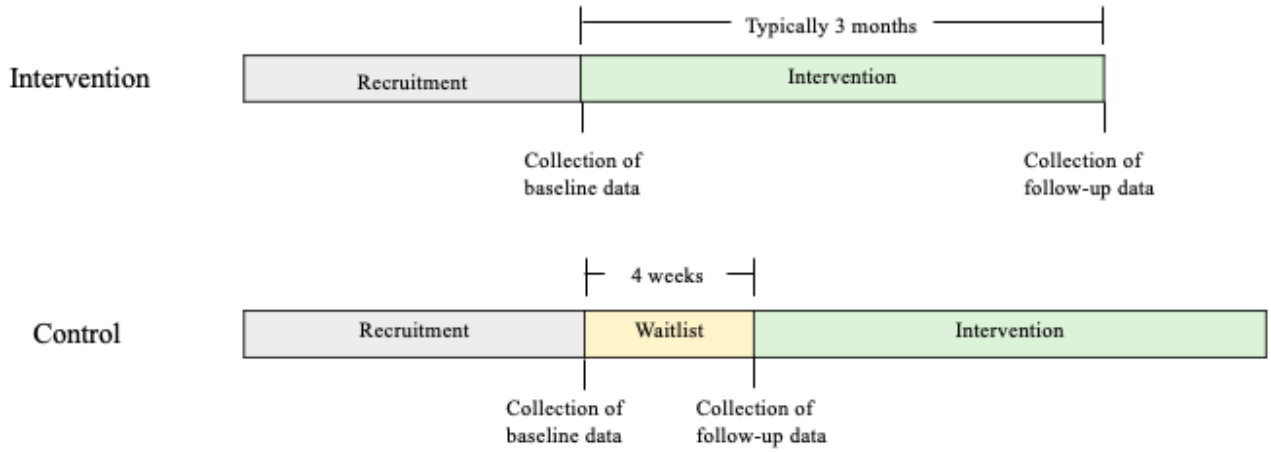
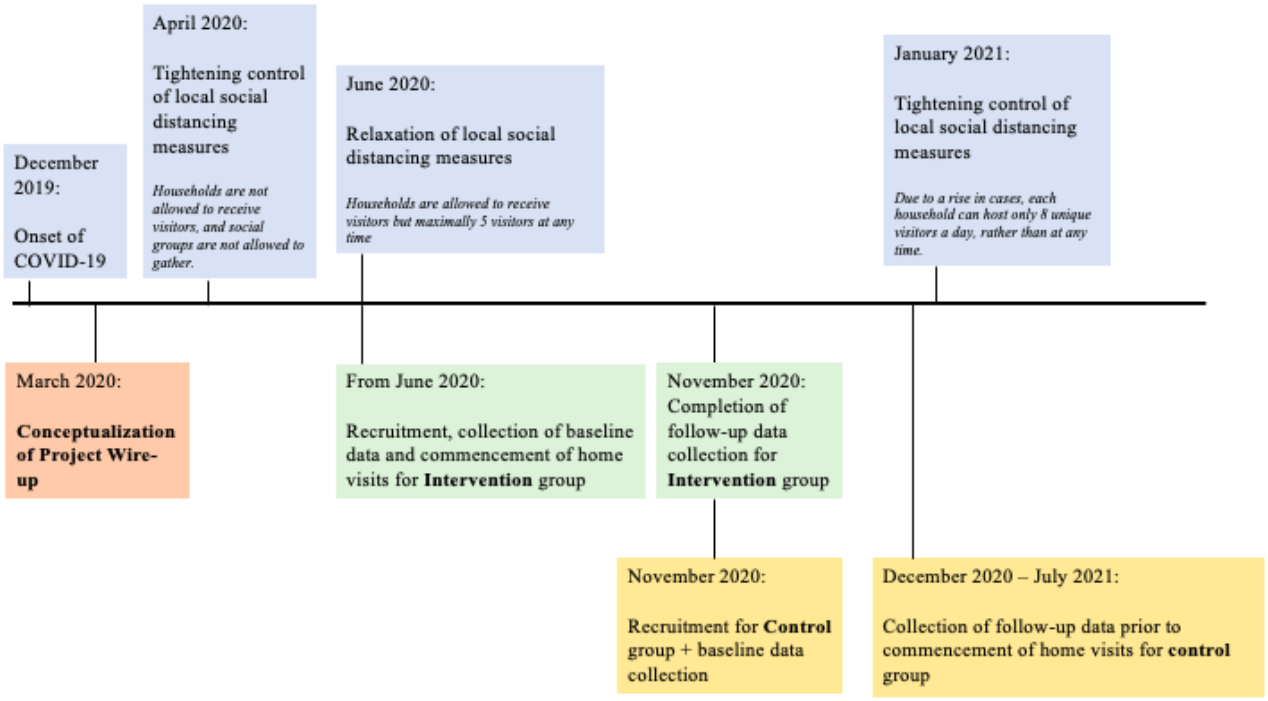


Figure 2. Timeline of the program with relation to policy changes in local social distancing measures.



Data Collection

Data were collected from participants either in person or via telephone through standardized self-reported questionnaires in participants' preferred language. Standardized training was provided for surveyors prior to household visits for recruitment and survey administration. If there was no response at the first instance, visits or phone calls were conducted on at least 3 separate occasions, with at least 1 scheduled on a weekend, to maximize participation. Data collection was completed in November 2021.

Measures

The primary outcome was digital literacy score, and secondary outcomes were Lubben Social Network Scale-6 (LSNS-6), University of California, Los Angeles 3-item loneliness scale (UCLA-3), EQ-5D, and Personal Wellbeing Score (PWS).

There is no universally accepted definition of digital literacy [26]. Hence, in this study, digital literacy is defined as the knowledge of the functional use of smartphones. To measure digital literacy, a 13-item self-reported digital literacy scale was constructed based on 4 domains of smartphone usage relevant and applicable to older adults [27]: Social (staying connected with social networks); Pass Time (using phones for relaxation or entertainment); Reassurance (feeling safe in an emergency); and Instrumental (obtaining news and information and accessing health, government, and banking services). An overall digital literacy score was computed by binarizing the scores (0=do not know how to use, 1=know how to use) and summed, with scores ranging from 0 to 13. The scale has been locally validated [28].

Social connectivity was measured using the locally validated LSNS-6 [29], where a higher numerical score indicates greater social connectedness [30]. Perceived loneliness was assessed using the UCLA-3, where participants scoring from 3-5 were classified as "not lonely," whereas those scoring from 6-9 were classified as "lonely" [31]. Subjective well-being was assessed using PWS [32]. Quality of life was assessed using locally validated EQ-5D-3L and EQ visual analogue scale [33-35].

Data and Statistical Analysis

Power analysis for sampling size was not calculated prior to the study as this was a pilot study. The aim of the study was to recruit at least 100 older adults in total.

Analysis was by intention to treat. Participant characteristics in both intervention and control groups were described by frequencies and their proportions for categorical variables and by means and 95% CI for numerical data. Independent sample 2-tailed *t* test, Wilcoxon sign-rank test, and χ^2 tests were used to compare differences in baseline characteristics between participants of the different groups. Paired sample 2-tailed *t* test and Wilcoxon sign-ranked test were conducted to assess differences in participants characteristics and outcomes between baseline and follow-up for continuous variables within each group, dependent on the nature of data distribution within variables. Differences in loneliness statuses among participants between groups were explored by conducting a logistic regression analysis, adjusting for the baseline loneliness statuses of participants in the model.

Regression coefficients (β) and odds ratios of the association between group membership (control vs intervention) with the various outcome measures over time were estimated using a series of hierarchical linear or logistic regression models, dependent on the nature of the outcome variable in question. In these longitudinal analyses, the first model (Model 1) adjusted for baseline outcome scores/statuses. The second model (Model 2) adjusted for age, gender, education, housing type, and living arrangement at baseline, along with predictors in Model 1. In the third and final model (Model 3), social isolation and

loneliness statuses at baseline were included as covariates alongside predictors indicated in Model 2. Statistical significance was set at $P<.05$ and tests were 2-tailed. Complete-case analysis was used for missing data. All analyses were conducted using STATA software (version 14; StataCorp LLC) [36].

Ethics Approval

Ethical approval was obtained from SingHealth Centralized Institutional Review Board (2020/2722). Eligible participants provided written informed consent. The study follows the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) reporting guidelines [37].

Results

Demographics

From July 2020 to November 2021, 150 older adults were invited to participate in the study. Of the 91 participants assigned to the intervention arm, 84 were included for analysis, with 7 participants excluded from analysis. Of the 59 participants assigned to the control arm, 5 were excluded from analysis, leaving 54 included for analysis (Figure 3).

Participants in both intervention and control groups were similar in age, gender, marital status, race, living arrangement, smartphone ownership, social connectivity, loneliness status, quality of life, and subjective well-being as seen in Table 1. Control group participants were found to have a significantly higher digital literacy score at baseline when compared to those in the intervention group (mean difference 2.28, 95% CI 1.37-3.20; $P<.001$). Participants in the intervention arm had a median of 3.5 (IQR 2-5) visits across the study duration.

Figure 3. Consolidated Standards of Reporting Trials (CONSORT) flow diagram.

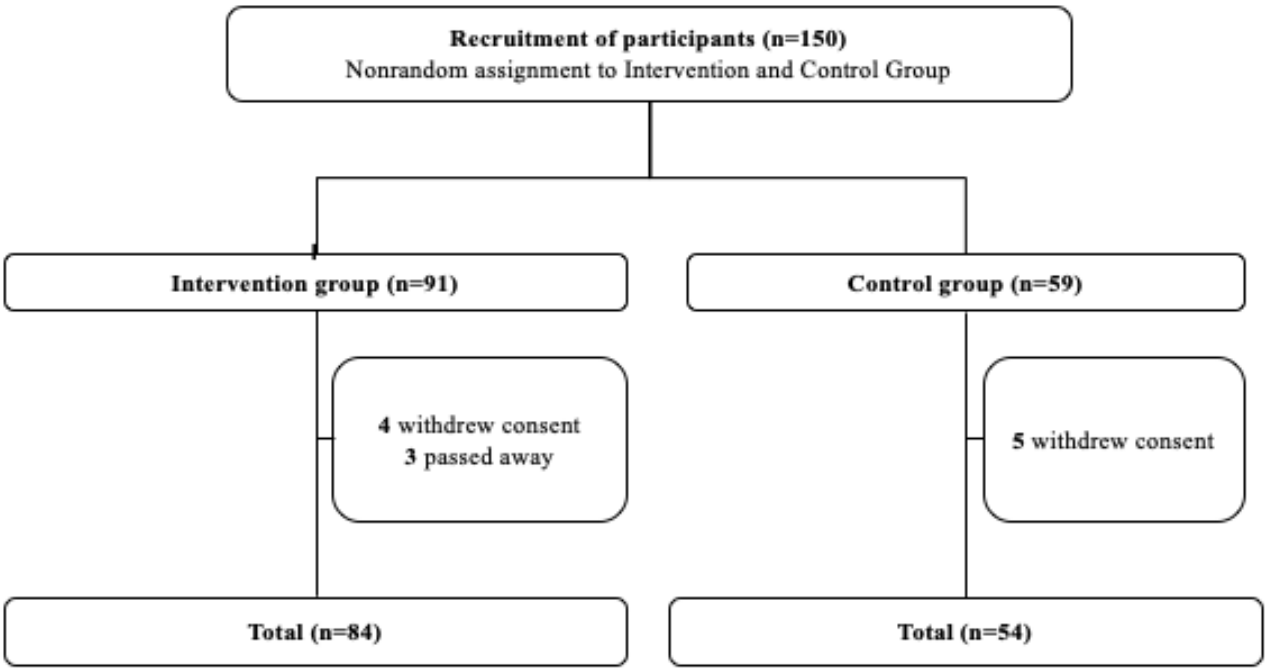


Table 1. Participants' baseline characteristics.

Characteristics	Intervention (n=84)	Control (n=54)	*P* value
Age (years), mean (95% CI)	72.31 (70.38-74.24)	71.85 (69.86-73.85)	.74
Gender, n (%)			.70
Male	37 (44)	22 (41)	
Female	47 (56)	32 (59)	
Ethnicity, n (%)			.57
Chinese	67 (80)	40 (74)	
Indian	8 (10)	5 (9)	
Malay	9 (11)	9 (17)	
Education, n (%)			.53
No formal education	23 (27)	12 (22)	
Primary	34 (40)	23 (43)	
Secondary	22 (26)	18 (34)	
Diploma and higher	5 (6)	1 (2)	
Not married (single, separated, divorced, or widowed), n (%)	24 (29)	19 (35)	.41
Living alone, n (%)	50 (60)	26 (48)	.19
Housing type A, n (%)			.21
Rental	66 (79)	47 (87)	
Self-owned	18 (21)	7 (13)	
Housing type B, n (%)			.25
1-room	52 (62)	30 (56)	
2-room	21 (25)	20 (37)	
3-room and above	11 (13)	4 (7)	
Have a mobile phone at baseline, n (%)	72 (86)	47 (87)	.83
Have a smartphone at baseline, n (%)	51 (61)	39 (74)	.12
Digital literacy score, mean (95% CI)	*3.77 (2.94-4.93)*	*5.09 (4.15-6.04)*	*.04*
LSNS-6[a] score, mean (95% CI)	11.07 (9.83-12.31)	12.36 (10.31-14.41)	.29
Loneliness (UCLA-3[b] score=6-9), n (%)	23 (27)	10 (19)	.23
Personal Wellbeing Score (n=120), mean (95% CI)	8.42 (7.78-9.06)	7.47 (6.51-8.44)	.10
EQ-5D Utility (n=135), mean (95% CI)	0.80 (0.75-0.85)	0.77 (0.69-0.85)	.74
EQ-5D visual analogue scale (n=135), mean (95% CI)	66.79 (62.74-70.84)	70.54 (64.59-76.48)	.30

[a]LSNS-6: Lubben Social Network Scale-6.

[b]UCLA-3: University of California, Los Angeles 3-item loneliness scale.

Primary Outcome

The intervention group observed a statistically significant difference in the change in their mean digital literacy score before and after program, as compared to those in the control group (mean difference in change: 2.28, 95% CI 1.37-3.20; *P*<.001; Table 2). Statistical control for this difference in baseline digital literacy scores was implemented in the analyses pertaining to the digital literacy score. Through multiple linear regression analyses, this change in digital literacy scores remained independently associated with group membership after adjusting for baseline digital literacy scores and differences in age, gender, education, living arrangement, housing type, and baseline social connectivity and loneliness status (Model 2, β=1.91, 95% CI 0.93-2.89; *P*<.001 and Model 3, β=1.90, 95% CI 0.91-2.90, *P*<.001), as seen in Table 3.

The domain-level analyses showed that the greatest gain was in the Instrumental domain (obtaining news and information and accessing health, government, and banking services), where the participants in the intervention arm learned, on average, approximately 1 more new function than the control arm, followed by the Reassurance, Social, and Pastime domains. The before and after program difference in all domains except for the Pastime domain remained statistically significant after controlling for covariates (Table 3).

Table 2. Intervention and control group differences.

Variable	Intervention, mean (95% CI)	Control, mean (95% CI)	Mean difference (95% CI)	*P* value
Primary outcome analysis				
Digital literacy score	2.42 (1.73 to 3.11)	0.13 (–0.48 to 0.75)	2.28 (1.37 to 3.20)	<.001
Domain-level analyses				
Social domain	0.64 (0.43 to 0.85)	0.20 (0.00 to 0.40)	0.44 (0.15 to 0.73)	.003
Instrumental domain	0.77 (0.42 to 1.12)	0.04 (–0.28 to 0.36)	0.74 (0.27 to 1.20)	.002
Reassurance domain	0.46 (0.24 to 0.07)	–0.15 (–0.40 to 0.07)	0.61 (0.30 to 0.92)	<.001
Pastime domain	0.54 (0.29 to 0.78)	0.13 (–0.15 to 0.39)	0.41 (0.05 to 0.76)	.03

Table 3. Multiple linear regression analyses.

Variable	Model 1[a]		Model 2[b]		Model 3[c]	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Primary outcome analysis						
Digital literacy score	1.99 (1.02 to 2.95)	<.001	1.91 (0.93 to 2.89)	<.001	1.90 (0.91 to 2.90)	<.001
Domain-level analyses						
Social domain	0.34 (0.09 to 0.61)	.009	0.35 (0.09 to 0.61)	.009	0.35 (0.08 to 0.61)	.01
Instrumental domain	0.66 (0.23 to 1.10)	.003	0.71 (0.26 to 1.15)	.002	0.70 (0.24 to 1.15)	.003
Reassurance domain	0.32 (0.06 to 0.59)	.02	0.30 (0.03 to 0.57)	.03	0.32 (0.05 to 0.60)	.02
Pastime domain	0.30 (–0.02 to 0.62)	.06	0.29 (–0.05 to 0.64)	.10	0.28 (–0.06 to 0.63)	.11

[a]Model 1: group (control [reference group] / intervention), baseline domain score.

[b]Model 2: predictors in Model 1 and age, gender, education, housing type and living arrangement at baseline.

[c]Model 3: predictors in Model 2 and social isolation (Lubben Social Network Scale-6) and loneliness (University of California, Los Angeles 3-item loneliness scale) at baseline.

Secondary outcomes

There was no statistically significant difference in LSNS-6 (mean difference –1.47, 95% CI –3.42 to 0.49; *P*=.14), EQ-5D Utility score (mean difference 0.09, 95% CI –0.02 to 0.2; *P*=.11) and visual analogue scale score (mean difference 1.20, 95% CI –6.11 to 8.52; *P*=.45), or PWS (mean difference –1.28, 95% CI –2.45 to –0.12; *P*=.69). Loneliness status, as measured by UCLA-3, showed no significant changes between the 2 groups before and after the intervention period (odds ratio 1.35, 95% CI 0.42-4.35, *P*=.62; Table 4).

Table 4. Secondary outcome analysis.

Variable	Intervention, mean (95% CI)	Control, mean (95% CI)	Mean difference (95% CI)	Odds ratio (95% CI)	*P* value
LSNS-6[a]	–1.26 (–2.58 to 0.06)	0.20 (–1.26 to 1.67)	–1.47 (–3.42 to 0.49)	N/A[b]	.14
EQ-5D Utility score	–0.07 (–0.15 to –0.002)	0.02 (–0.07 to 0.11)	0.09 (–0.02 to 0.20)	N/A	.11
EQ-5D VAS[c] score	1.85 (–2.63 to 6.33)	3.06 (–2.81 to 8.92)	1.20 (–6.11 to 8.52)	N/A	.45
PWS[d] total	–0.73 (–1.50 to 0.04)	0.56 (–0.34 to 1.45)	–1.28 (–2.45 to –0.12)	N/A	.69
Loneliness	N/A	N/A	N/A	1.35 (0.42 to 4.35)	.62

[a]LSNS-6: Lubben Social Network Scale-6.

[b]N/A: not applicable.

[c]VAS: visual analogue scale.

[d]PWS: Personal Wellbeing Score.

Discussion

Principal Findings

Our study revealed that a volunteer-led, one-on-one, and home-based digital literacy program undertaking a goal-directed approach contributed to a significant increase in digital literacy among community-dwelling older adults of low SES strata in Singapore who are digitally excluded. However, the program did not result in expected improvements in loneliness, social connectedness, quality of life, and personal well-being.

The increase in the knowledge of smartphone use in older adults of lower SES who are digitally excluded seen from our study could be attributed to key elements of our program [38]. The home-based, one-to-one approach to digital learning allowed us to better contextualize the digital training to each older adult while being able to provide close mentoring and support—factors found in adult learning literature to encourage ICT learning [23]. By breaking down the digital training into different tiers and matching each older adult's capability, the program was able to build older adults' confidence and sustain their motivation for smartphone learning, which has been shown to be important for technology adoption in older adults [39].

This increase in digital literacy did not translate into expected changes observed for loneliness, social connectedness, quality of life, and subjective well-being in our study. A possible explanation for this finding is that the older adults did not have any existing social networks to be tapped into and were at risk of social isolation as suggested by their LSNS-6 score being less than 12 [30]. Although deliberate efforts were made to digitally connect participants to their existing social networks, the participants' limited social connections, a lack of corresponding digital adoption among peers in their social networks, and a dearth of social activities for older adults available on the digital space presented as challenges to the program. As such, the increase in digital literacy might not have translated to sustained use of new technology in older adults' lives and or an increment in social activities or connections, resulting in a lack of observed changes for loneliness, social isolation, quality of life, and well-being. This finding is supported by studies in the literature that postulate that ICT use results in improvement in health-related outcomes in older adults by connecting them to their social networks, gaining social support, and engaging in activities of interest [4,40]

An implication from our study results is the need for digital literacy programs to move toward encouraging long-term digital adoption in older adults to truly impact health outcomes [22]. Future studies should look at the design of current web-based spaces and digital technologies, to develop more digitally inclusive spaces for older adults. This can increase the confidence and compatibility of digital technology with older adults, resulting in greater interest or motivations in older adults to take up digital technology [41] and the sustainability of digital literacy programs through continued use beyond programs [22].

Strengths and Limitations

This study has several strengths. To the best of our knowledge, this is one of the first few studies in the world assessing the impact of a home-based digital literacy program on improving digital literacy among community-dwelling older adults of a low SES amid the COVID-19 pandemic. This provides vital empirical information required in the planning of future digital literacy programs for this vulnerable group in view of possible future pandemics.

Furthermore, data collection was conducted in person or via telephone interviews, unlike prior studies using web-based surveys to explore the effect of COVID-19 on older adults in other countries. This methodology allowed us to include participants who were digitally excluded and might not have been included in other web-based studies due to these older adults' limitations or unwillingness to access the internet [42,43]. Through our study methodology, we were able to have a more representative picture of the impacts of our digital literacy program on older adults with little (or no) smartphone use during a pandemic.

At the same time, the design and implementation of our study was constrained by the practical limitations in implementing the intervention during the COVID-19 pandemic. The intervention group and control group participants were followed up at different time frames, as it would not be feasible to keep these control participants waiting beyond 4 weeks before being digitally equipped during the peak of the COVID-19 pandemic given the risk of potential social isolation. At the same time, using a waitlist design ensured that both groups were made up of individuals with the same inclination to participate in a digital literacy program, with the only difference between groups being the timing of the intervention. The delay for control group participants allowed the team to see the changes in digital knowledge and behavior across time when they are not participating in the digital literacy program. Follow-up data for the control group was specifically collected before they started the program to reduce confounding effects the training had on control participant's digital literacy skills.

A nonrandom assignment of participants to groups was used due to the waitlist approach. A difference in baseline digital literacy score between groups was observed, where the control group had a higher baseline digital literacy score. To mediate this, statistical control for this difference was implemented in the analyses pertaining to the digital literacy score.

Finally, outcomes were self-reported by older adults, which may have impacted the accuracy in measuring changes in key outcomes such as digital literacy. Moving forward, future studies should use blinding of assessors and include objective assessments of the older adult's digital literacy where practically feasible.

Conclusion

Our study has provided preliminary empirical evidence to support the effectiveness of a volunteer-led, one-on-one, and home-based digital literacy program for older adults of lower SES in Singapore. Although the current intervention has a limited impact on secondary outcomes such as loneliness, social connectedness, quality of life, or subjective well-being, our findings is a step toward ensuring digital inclusivity in a world where there is rising inequity due to rapid digitalization. In the postpandemic world, digital use will no longer be a choice but an essential part of daily living [22]. For those who lack digital resources and know-how, their ability to access services and resources that impact the various determinants of health can be impeded, which can lead to adverse health outcomes [44]. Future studies should look into developing more age-friendly web-based spaces and technology design, which can encourage continued digital adoption in older adults and, eventually, impact health-related outcomes.

Acknowledgments

This research is supported by the Singapore Ministry of Health's National Medical Research Council under the Fellowship Programme by SingHealth Regional Health System, Population-based, Unified, Learning System for Enhanced and Sustainable (PULSES) Health Centre Grant (NMRC/CG/C027/2017_SHS), the Healthy, Empowered and Active Living (HEAL) fund and the Infocomm Media Development Authority Digital for Life Fund.

Authors' Contributions

LLL, KYYN, and NHWN conceived the project and led the research team in designing and implementing the study. KSY, AS, JXL, AT, KWAT, CYXT, NT, WQY, KYYN, and NHWN implemented the study and collected the data. NT and JXL analyzed the data. WQY and NHWN cowrote the first draft of the main text, supplementary materials, and constructed the figures and tables. All authors reviewed and revised the main text and supplementary materials.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Compilation of all supplementary material inclusive of study protocol, questionnaires, training material, and Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) checklist.
[DOCX File , 714 KB - jmir_v24i12e40341_app1.docx]

References

1. World internet usage statistics. Internet World Stats: Usage and Population Statistics. 2021 Dec 31. URL: https://www.internetworldstats.com/stats.htm [accessed 2022-04-11]
2. van Volkom M, Stapley JC, Amaturo V. Revisiting the digital divide: generational differences in technology use in everyday life. N Am J Psychol 2014;16(3):557-574 [FREE Full text]
3. Zhang F, Kaufman D. The impacts of social interactions in MMORPGs on older adults' social capital. Comput Human Behav 2015 Oct;51:495-503 [FREE Full text] [doi: 10.1016/j.chb.2015.05.034]
4. Chen YRR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
5. Cotten S, Ford G, Ford S, Hale T. Internet use and depression among older adults. Comput Human Behav 2012 Mar;28(2):496-499 [FREE Full text] [doi: 10.1016/j.chb.2011.10.021]
6. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol Behav Soc Netw 2015 May;18(5):268-272 [FREE Full text] [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]
7. Chopik WJ. The benefits of social technology use among older adults are mediated by reduced loneliness. Cyberpsychol Behav Soc Netw 2016 Sep;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
8. Gibbons C. Digital access disparities: policy and practice overview. In: Sheon, A. Conference Report: Digital Skills: A Hidden 'Super' Social Determinant of Health. 2018 Oct 15 Presented at: Digital Skills and Connectivity as Social Determinants of Health; October 3-5, 2018; Washington, DC URL: https://iaphs.org/conference-report-digital-skills-hidden-super-social-determinant-health/
9. Susło R, Paplicki M, Dopierała K, Drobnik J. Fostering digital literacy in the elderly as a means to secure their health needs and human rights in the reality of the twenty-first century. Family Medicine & Primary Care Review 2018;20(3):271-275. [doi: 10.5114/fmpcr.2018.78273]
10. Norval C, Arnott JL, Hanson VL. What's on your mind?: investigating recommendations for inclusive social networking and older adults. 2014 Apr 26 Presented at: CHI '14: the SIGCHI Conference on Human Factors in Computing Systems; April 26 to May 1, 2014; Toronto, ON p. 3923-3932. [doi: 10.1145/2556288.2556992]
11. Ang S, Lim E, Malhotra R. Health-related difficulty in internet use among older adults: correlates and mediation of its association with quality of life through social support networks. Gerontologist 2021 Jul 13;61(5):693-702. [doi: 10.1093/geront/gnaa096] [Medline: 32744310]
12. Elliot AJ, Mooney CJ, Douthit KZ, Lynch MF. Predictors of older adults' technology use and its relationship to depressive symptoms and well-being. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):667-677 [FREE Full text] [doi: 10.1093/geronb/gbt109] [Medline: 24304556]
13. Hayes M, van Stolk-Cooke K, Muench F. Understanding Facebook use and the psychological affects of use across generations. Comput Human Behav 2015 Aug;49:507-511 [FREE Full text] [doi: 10.1016/j.chb.2015.03.040]

14. Tirado-Morueta R, Hernando-Gómez Á, Aguaded-Gomez JI. The capacity of elderly citizens to access digital media in Andalusia (Spain). Information, Communication & Society 2015 Nov 19;19(10):1427-1444. [doi: 10.1080/1369118x.2015.1111401]
15. Clair R, Gordon M, Kroon M, Reilly C. The effects of social isolation on well-being and life satisfaction during pandemic. Humanit Soc Sci Commun 2021 Jan 27;8(1):1-6. [doi: 10.1057/s41599-021-00710-3]
16. Lim WS, Liang CK, Assantachai P, Auyeung TW, Kang L, Lee W, et al. COVID-19 and older people in Asia: Asian Working Group for Sarcopenia calls to actions. Geriatr Gerontol Int 2020 Jun 14;20(6):547-558 [FREE Full text] [doi: 10.1111/ggi.13939] [Medline: 32365259]
17. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Division of Behavioral and Social Sciences and Education, Board on Health Sciences Policy, Board on Behavioral, Cognitive, and Sensory Sciences, Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults. Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System. Washington, DC: National Academies Press; 2020.
18. Beaunoyer E, Dupéré S, Guitton MJ. COVID-19 and digital inequalities: reciprocal impacts and mitigation strategies. Comput Human Behav 2020 Oct;111:106424 [FREE Full text] [doi: 10.1016/j.chb.2020.106424] [Medline: 32398890]
19. Chan A, Raman P, Ma S, Malhotra R. Loneliness and all-cause mortality in community-dwelling elderly Singaporeans. Demogr Res 2015 Jun 03;32:1361-1382. [doi: 10.4054/demres.2015.32.49]
20. Straughan PT, Tov W, Kim S, Cheng T, Hoskins S, Tan M. Attitudes, behaviours, and the well-being of older Singaporeans in the time of COVID-19: perspectives from the Singapore Life Panel. ROSA Research Briefs 2020 Jul;1:1-21 [FREE Full text]
21. Seniors Go Digital. Infocomm Media Development Authority. URL: http://www.imda.gov.sg/en/seniorsgodigital [accessed 2022-04-11]
22. Ng IYH, Sun LS, Pang N, Kim D, Soh G, Pakianathan PV, et al. From digital exclusion to universal digital access in Singapore. National University of Singapore 2021 Jan:1-28 [FREE Full text]
23. Pont B, Sweet R. Adult learning and ICT: how to respond to the diversity of needs. In: , editor. ICT and Learning: Supporting Out-of-School Youth and Adults. Paris, France: OECD Publishing; May 12, 2006.
24. Lim HA, Lee JSW, Lim MH, Teo LPZ, Sin NSW, Lim RW, et al. Bridging connectivity issues in digital access and literacy: reflections on empowering vulnerable older adults in Singapore. JMIR Aging 2022 May 03;5(2):e34764 [FREE Full text] [doi: 10.2196/34764] [Medline: 35503520]
25. Comcare long term assistance (public assistance scheme). Centre for Seniors. URL: https://www.silverschemes.sg/post_single_age/comcare-long-term-assistance-public-assistance-scheme/ [accessed 2022-04-11]
26. Schreurs K, Quan-Haase A, Martin K. Problematizing the digital literacy paradox in the context of older adults' ICT use: aging, media discourse, and self-determination. Canadian Journal of Communication 2017 May 23;42(2):359-377. [doi: 10.22230/cjc.2017v42n2a3130]
27. Wang Y, Matz-Costa C, Miller J, Carr DC, Kohlbacher F. Uses and gratifications sought from mobile phones and loneliness among Japanese midlife and older adults: a mediation analysis. Innov Aging 2018 Sep;2(3):igy027 [FREE Full text] [doi: 10.1093/geroni/igy027] [Medline: 30480146]
28. Soundararajan A, Lim JX, Ngiam NHW, Tey AJY, Tang AKW, Lim HA, et al. Smartphone ownership, digital literacy, and the mediating role of connectedness and loneliness in improving the wellbeing among community-dwelling Singaporean older adults of low socio-economic status. PsyArXiv. Preprint published online on February 9, 2022. [FREE Full text] [doi: 10.31234/osf.io/jx8b7]
29. Ge L, Yap CW, Ong R, Heng BH. Social isolation, loneliness and their relationships with depressive symptoms: a population-based study. PLoS One 2017 Aug 23;12(8):e0182145 [FREE Full text] [doi: 10.1371/journal.pone.0182145] [Medline: 28832594]
30. 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]
31. 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 09;110(15):5797-5801 [FREE Full text] [doi: 10.1073/pnas.1219686110] [Medline: 23530191]
32. Benson T, Sladen J, Liles A, Potts HWW. Personal Wellbeing Score (PWS)-a short version of ONS4: development and validation in social prescribing. BMJ Open Qual 2019 Apr 11;8(2):e000394 [FREE Full text] [doi: 10.1136/bmjoq-2018-000394] [Medline: 31206049]
33. Merchant RA, Liu SG, Lim JY, Fu X, Chan YH. Factors associated with social isolation in community-dwelling older adults: a cross-sectional study. Qual Life Res 2020 Sep 06;29(9):2375-2381. [doi: 10.1007/s11136-020-02493-7] [Medline: 32253669]
34. Luo N, Wang P, Thumboo J, Lim Y, Vrijhoef HJM. Valuation of EQ-5D-3L health states in Singapore: modeling of time trade-off values for 80 empirically observed health states. Pharmacoeconomics 2014 May 12;32(5):495-507. [doi: 10.1007/s40273-014-0142-1] [Medline: 24519603]

35. Rabin R, Gudex C, Selai C, Herdman M. From translation to version management: a history and review of methods for the cultural adaptation of the EuroQol five-dimensional questionnaire. Value Health 2014;17(1):70-76 [FREE Full text] [doi: 10.1016/j.jval.2013.10.006] [Medline: 24438719]
36. StataCorp. Stata Statistical Software: Release 14. StataCorp LLC. 2015. URL: https://www.stata.com/ [accessed 2022-11-18]
37. Haynes AB, Haukoos JS, Dimick JB. TREND reporting guidelines for nonrandomized/quasi-experimental study designs. JAMA Surg 2021 Sep 01;156(9):879-880. [doi: 10.1001/jamasurg.2021.0552] [Medline: 33825826]
38. Lim HA, Lee JSW, Lim MH, Teo LPZ, Sin NSW, Lim RW, et al. Bridging connectivity issues in digital access and literacy: reflections on empowering vulnerable older adults in Singapore. JMIR Aging 2022 May 03;5(2):e34764. [doi: 10.2196/34764] [Medline: 35503520]
39. Wlodkowski RJ, Ginsberg MB. Enhancing Adult Motivation to Learn: A Comprehensive Guide for Teaching All Adults, 4th ed. San Francisco, CA: John Wiley & Sons; 2017.
40. Han M, Tan XY, Lee R, Lee JK, Mahendran R. Impact of social media on health-related outcomes among older adults in Singapore: qualitative study. JMIR Aging 2021 Feb 17;4(1):e23826 [FREE Full text] [doi: 10.2196/23826] [Medline: 33595437]
41. den Haan M, Brankaert R, Kenning G, Lu Y. Creating a social learning environment for and by older adults in the use and adoption of smartphone technology to age in place. Front Public Health 2021;9:568822 [FREE Full text] [doi: 10.3389/fpubh.2021.568822] [Medline: 34222160]
42. Parlapani E, Holeva V, Nikopoulou VA, Kaprinis S, Nouskas I, Diakogiannis I. A review on the COVID-19-related psychological impact on older adults: vulnerable or not? Aging Clin Exp Res 2021 Jun 17;33(6):1729-1743 [FREE Full text] [doi: 10.1007/s40520-021-01873-4] [Medline: 33999378]
43. Malhotra R, Visaria A. Social isolation in the COVID-19 pandemic? is maintaining social connections online a viable option for all older persons: insights from Singapore. SSRN Electronic Journal 2020 Jan:1-17. [doi: 10.2139/ssrn.3609081]
44. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021 Mar 17;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]

Abbreviations

ICT: information and communications technology
LSNS-6: Lubben Social Network Scale-6
PWS: Personal Wellbeing Score
SES: socioeconomic status
TREND: Transparent Reporting of Evaluations with Nonrandomized Designs
UCLA-3: University of California, Los Angeles 3-item loneliness scale

Edited by R Kukafka; submitted 16.06.22; peer-reviewed by S Kim, W Wei Chin; comments to author 12.07.22; revised version received 05.09.22; accepted 13.11.22; published 02.12.22.

Please cite as:
Ngiam NHW, Yee WQ, Teo N, Yow KS, Soundararajan A, Lim JX, Lim HA, Tey A, Tang KWA, Tham CYX, Tan JPY, Lu SY, Yoon S, Ng KYY, Low LL
Building Digital Literacy in Older Adults of Low Socioeconomic Status in Singapore (Project Wire Up): Nonrandomized Controlled Trial
J Med Internet Res 2022;24(12):e40341
URL: https://www.jmir.org/2022/12/e40341
doi:10.2196/40341
PMID:36459398

Original Paper

A Novel Approach to Characterize State-level Food Environment and Predict Obesity Rate Using Social Media Data: Correlational Study

Chuqin Li[1], PhD; Alexis Jordan[1], BSc; Jun Song[2], PhD; Yaorong Ge[1], PhD; Albert Park[1], PhD

[1]Department of Software and Information Systems, College of Computing and Informatics, University of North Carolina at Charlotte, Charlotte, NC, United States

[2]Department of Statistics, Korea University, Seoul, Republic of Korea

Corresponding Author:
Albert Park, PhD
Department of Software and Information Systems, College of Computing and Informatics
University of North Carolina at Charlotte
9201 University Blvd
Charlotte, NC, 28223
United States
Phone: 1 704 687 8668
Email: al.park@uncc.edu

Abstract

Background: Community obesity outcomes can reflect the food environment to which the community belongs. Recent studies have suggested that the local food environment can be measured by the degree of food accessibility, and survey data are normally used to calculate food accessibility. However, compared with survey data, social media data are organic, continuously updated, and cheaper to collect.

Objective: The objective of our study was to use publicly available social media data to learn the relationship between food environment and obesity rates at the state level.

Methods: To characterize the caloric information of the local food environment, we used food categories from Yelp and collected caloric information from MyFitnessPal for each category based on their popular dishes. We then calculated the average calories for each category and created a weighted score for each state. We also calculated 2 other dimensions from the concept of access, acceptability and affordability, to build obesity prediction models.

Results: The local food environment characterized using only publicly available social media data had a statistically significant correlation with the state obesity rate. We achieved a Pearson correlation of 0.796 between the predicted obesity rate and the reported obesity rate from the Behavioral Risk Factor Surveillance System across US states and the District of Columbia. The model with 3 generated feature sets achieved the best performance.

Conclusions: Our study proposed a method for characterizing state-level food environments only using continuously updated social media data. State-level food environments were accurately described using social media data, and the model also showed a disparity in the available food between states with different obesity rates. The proposed method should elastically apply to local food environments of different sizes and predict obesity rates effectively.

(J Med Internet Res 2022;24(12):e39340) doi:10.2196/39340

KEYWORDS

obesity; social media; machine learning; lifestyle; environment; food; correlation; modeling; predict; rates; outcome; category; dishes; popular; mobile phone

Introduction

Background

The current obesity epidemic poses critical public health challenges. Obesity is a major risk factor for other chronic diseases, such as cardiovascular disease, cancer, diabetes, and respiratory disorders, which account for 60% of the deaths worldwide [1]. Excessive body weight has resulted in a medical expenditure of US $100 billion per year [2,3]. From 2017 to 2018, the prevalence of obesity among adults in the United States was 42.4% [4]. This number has more than tripled since the 1960s. From 1960 to 1962, the obesity rate was 13.4% [5].

Environmental factors, including the types of available food, have been identified as one of the main drivers of obesity [3,6,7]. It was reported that American adults have developed a preference for dining out with friends as opposed to cooking at home [8]. This preference could potentially impact health outcomes. A market research survey conducted in 2017 found that those who frequent fast-food restaurants are more concerned about the value of money spent and service speed than the actual healthiness of the food offered [8]. This indication that the perceived food availability tends to affect dietary outcomes has been furthered only in a literature review conducted by Caspi et al [9]. Those who live in areas highly saturated with high-fat food items tend to have health issues. In addition, those who live in lower-income areas are more likely to have at least one diet-related health issue [9]. In the United States, people tend to eat what is affordable and available to them. Environments littered with low-cost, high-fat foods tend to be obesogenic. With food expenditures for dining out increasing in recent years [3,10], understanding the food environmental factors is critical in counteracting the obesity epidemic and understanding related human behavior.

Recent studies have suggested that the local food environment can be measured by the degree of food accessibility [6,11]. These studies measured food accessibility using survey data [12], yellow pages phone books [13,14], and local business directories [15]. A limited number of samples and a significant delay between the collection and reporting of data are major limitations of these traditional methods [9]. With the proliferation of social media, the data from social media are organic, continuously updated, and generally free for large-scale collection. Several studies have used social media data to learn food environments by estimating the calorie density of the foods mentioned in tweets [16] or using the linguistic variables from tweets [17-19] to predict the local obesity rate.

In this study, we leveraged large-scale social media data sets to measure food environments at the state level and predict state-level obesity rates. It remained unclear whether we could characterize state-level food environments from the perspective of *concept of access* and predict obesity rate according to the perspective using publicly available social media data. Obesity rate was obtained from the Behavioral Risk Factor Surveillance System (BRFSS), the nation's premier system for collecting data to improve public health.

The primary aim of this descriptive study was to understand the impact of food environment on obesity with three specific research questions (RQs):

1. RQ1: Is there a difference between the available food categories in low and high obesity prevalent states?
2. RQ2: How can we use calorie information to quantify state-level food environments?
3. RQ3: Can we predict state-level obesity rate using publicly available social media data?

We reported our novel approaches and findings. To date, to our knowledge, our study is the first to combine information from Yelp and MyFitnessPal (MFP) to learn about the local food environment and then to predict the state-level obesity rate.

Related Work

Calorie With Obesity

An increase in daily calorie consumption is a major cause of the obesity epidemic [7]. The daily calorie intake rose by >500 calories in adults and >150 calories in children between 1977 and 2006 [20,21], as did the portion size in restaurants [22]. Exposure to a larger portion size increases the risk of increasing calorie intake and, therefore, weight gain [23]. Similarly, calorie intake is also affected by a higher number of local dining options. For example, the prevalence of obesity is lower in areas with supermarkets and higher in areas with higher numbers of fast-food restaurants [12].

Analysis of the data on environmental changes has identified the changes on food environment as a potential cause for the increase in caloric intake. The enormous growth in dining out, particularly at "fast-food" outlets, is a trend that has received a lot of attention. Fast-food outlets increased from approximately 30,000 in 1970 to >233,000 locations in 2004 in the United States [3]. Fast food can contribute to increasing obesity rate because it generally provides food that is poor in micronutrients, low in fiber, high in glycemic load, and excessive in portion size and calorie [24,25].

How to Characterize or Quantify Local Food Environment

Food access dimensions can be conceptualized using the *concept of access* proposed by Penchansky and Thomas [26]. The concept of access uses 5 dimensions to conceptualize the local food environment, namely availability, accessibility, affordability, acceptability, and accommodation [9,26]. Availability refers to the relationship between the number and type of food suppliers available to customers. Accessibility refers to the relationship between the location of food suppliers and the location of customers, which is more geographically inherent than availability. Accessibility could be measured by the travel time and distance between food suppliers and customers. Affordability refers to the price customers need to pay for the food. Acceptability refers to customers' attitudes toward a business. Accommodation is another dimension of access, which assesses whether local businesses accept and adapt to local customers' needs.

A variety of approaches have been used to learn about local food environments by measuring the degree of food access.

These approaches typically fall into 2 categories. The first category consists of methods that capture food environment by relying on respondent-based data. The accessibility of food stores was asked about in surveys or questionnaires. The methods in the second category used the geographic information system (GIS) technology. GIS measures the buffer distance to food stores or the density of food stores in an area [12-15]. By 2007, the GIS-based measures of food environment outnumbered the respondent-based measures, and the trend of using GIS measures continued [9,27,28]. The GIS data used in previous studies primarily used publicly available data sets, such as the United States yellow pages phone book [13,14], published data from the local Departments of Environmental Health and state Departments of Agriculture [12], and local business directories [15]. A major limitation of these traditional data collections is that they are cost-ineffective and labor intensive; moreover, these methods can only gather a limited number of samples, and there is a significant delay between the collection and reporting of data [9]. In the following section, we will illustrate quantifying the environment using social media data.

Using Social Media Data to Learn Obesity-Related Factors or Predict the Obesity Rate

Social media is used to characterize social factors [29] and food environment in relation to obesity. Nguyen et al [16] characterized food environment by calculating the calorie density of the foods mentioned in tweets and the percentage of each food theme out of all food-related Yelp entries from that state. They found that Twitter and Yelp posts that were indicative of higher caloric foods were related to higher mortality, higher prevalence of chronic conditions, and worse self-rated health [16]. Researchers also tried to understand healthy and unhealthy food images shared on social media in relation to obesity [30]. They created an image classifier and tested it out to classify Twitter images into definitively healthy, healthy, unhealthy, and definitively unhealthy categories. Social media was also used to understand obesity-preventive factors, such as physical activity [31]. The authors described how individuals organically use social media to encourage and sustain physical activity for obesity prevention.

Social media can also be used to predict obesity rate. Fried et al [17] presented "the predictive power behind the language of food on social media." They collected the food-related tweets that contained meal-related hashtags: dinner, breakfast, lunch, brunch, snack, meal, and supper. Then, they used the lexical feature from the bag-of-words model and topic features obtained from latent Dirichlet allocation to predict whether a state's obesity rate is above or below the national median. Their best model reached an accuracy of 80.39% in predicting overweight. Culotta [18] used the linguistic variables (Linguistic Inquiry and Word Count and PERMA) from tweets and demographic variables to predict health-related statistics for the 100 most populous counties in the United States. The Pearson correlation for obesity between the predicted and real rates was 0.64. Abbar et al [19] conducted a study similar to the one by Culotta [18]. Abbar et al [19] used the linguistic variables (Linguistic Inquiry and Word Count), food features, average calorie per serving for food, and demographic variables from food-related tweets to predict county-wide obesity rate, achieving a correlation of 0.775 for obesity. Public posts about food and eating behaviors may spread through social networks [32]. These studies demonstrated a successful application of Twitter data in predicting state health outcomes. Although Yelp data together with Twitter data have been used to characterize food environment by Nguyen et al [16], no previous study has been found to use Yelp and MFP data to predict state obesity.

Methods

Data Collection

Our study used 3 data sources: (1) Yelp, (2) MFP, and (3) BRFSS. The data used in this study to describe state-level food environments were collected by the research team via the Yelp application programming interface (API) [33] and the web scraping tool, BeautifulSoup.

Yelp is a leading crowd-sourced review site in the United States that allows users to search for restaurants and local businesses [34]. Users can post reviews and upload photos concerning a business's foods and services, which makes Yelp a location-based social media platform. To date, Yelp [35] ranks 52nd in the United States and 231st worldwide based on internet traffic and engagement [36].

The Yelp API allows users to search and query Yelp for more than 50 million businesses in 32 countries [33]. To obtain the data for this study, we converted 5-digit US zip codes to latitude and longitude coordinates and then queried the detailed business content via the Yelp API by searching the businesses near the provided locations. The data were collected in September 2020 and consisted of the profiles of 353,431 businesses in the United States.

An example of a restaurant's listing on Yelp [35] is shown in Figure 1. As shown in Figure 1, the profile of each business includes its name, average rating, price level, and categories and the number of reviews it has received. Each business can choose up to 3 terms (categories) to describe its services and offerings. The queried business profile returned by the Yelp API not only contains the mentioned fields but also includes other details of the business, such as the business ID, address, URL to the business's home page on Yelp [35], photos, and hours of operation. It is worth noting that chain businesses can have the same name, but each location has its unique business ID.

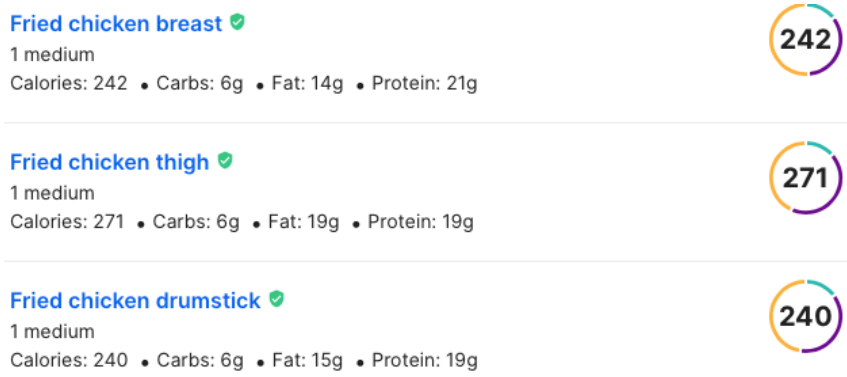
Yelp publishes reviews of many service businesses, such as restaurants, hospitals, and recreational activities. We removed businesses that were not related to the food industry in this study (eg, hardware stores). To do this, 2 independent reviewers first evaluated the relevance of each selected category to the food field independently. The 2 judgments reached 100% agreement with κ=1. A total of 226 categories were selected from 332 categories. In our collected data set, the total number of businesses is 353,431. The average rating of each business is 4.00 (SD 0.75), the average number of reviews of each business is 99.16 (SD 260.32), and the average price is US $1.60 ($ is the unit Yelp use to approximate cost per person for a meal) with an SD of 0.56.

To understand and objectively compare these categories, we further collected data on each category's most popular 100 restaurants nationwide and their most popular dishes for use as a proxy to estimate the caloric density of each category. We used BeautifulSoup [37] to collect popular dishes from each restaurant. We also used this web scraping tool to collect the nutritional information (ie, calories) of each popular dish from MFP. MFP is one of the most popular calorie-tracking smartphone apps worldwide with >10 million users [38]. MFP provides powerful tools to help users easily track their meals and physical activity. We collected food nutrition information by searching the food name in MFP's nutrition database. Figure 2 shows an example search result page, which appeared when we searched the term "Fried Chicken." We collected nutrition records for 37,295 dishes from MFP, and the total number of nutrition records is 3,110,744.

We obtained the state-level obesity rate data from the BRFSS, the nations' state-based health surveillance system that tracks the behavioral risk factors of residents in the United States. BRFSS provided the ground truth for the prevalence of obesity via self-reported obesity data among adults in the United States by state and territory in 2019. We collected the obesity rates for 49 states and the District of Columbia, excluding New Jersey, owing to insufficient data collection by the BRFSS in 2019 [39].

Figure 1. Example of the Yelp business list page.

Figure 2. Example of the MyFitnessPal nutrition fact list page.



RQ1: Is There a Difference Between the Available Food Categories in States With Low and High Obesity Prevalence?

We first characterized a local food environment based on the literature and then illustrated the quantification of the environment using social media data in RQ2. We based our characterization on food access dimensions [26]. Specifically, we focused on 3 highly distinct dimensions: availability, affordability, and acceptability [9]. Availability refers to the relationship between the number and type of food suppliers available to customers. Affordability refers to the price customers need to pay for the food. Acceptability refers to customers' attitudes toward a business.

We used the category information for each business in Yelp to calculate the availability of those food categories. Specifically, we defined the availability of a category of food as the number of available restaurants compared with the overall choices at the state level. For example, the availability of Mexican food will be equal to 1 if all the restaurants in that area sell Mexican food. Similarly, if 50% of the state's restaurants sell Mexican food, its availability will be 50%.

After calculating the availability of all food categories, we further compared the availability of food categories between states with low prevalence of obesity and those with high prevalence of obesity. We aimed to understand the impact of local food availability, a dimension that has been widely studied [3,16,40], on the state-level obesity rate. The 2 states we selected were Colorado and Mississippi. In 2019, Mississippi had the highest obesity rate (40.8%), whereas Colorado had the lowest obesity rate (23.8%) [39]. We first calculated the availability of each category in the 2 preselected locations and further analyzed what categories of restaurants are more available in locations with high or low obesity rate. The category with the

biggest availability difference was further compared by adopting dimensions from the *concept of access*.

The affordability and acceptability of the categories were then compared. Affordability refers to the food price customers need to pay. Price may affect the food choices of users. Low-income populations have a high risk of living in poor food environments and bear much of the burden of obesity and chronic diseases [14]. We estimated affordability using the price category data for each business. Here, we converted the price categories into numeric numbers for future analysis. For example, $ would have been converted to 1, and $$$$ would have been converted to 4. Acceptability refers to the client's attitude toward the service provider. We used the average customer rating and the total number of reviews of a business to measure customers' attitudes concerning a business. Studies have shown that consumers' preference increases with the number of reviews [41], and consumer-generated restaurant ratings are positively associated with the web-based popularity of restaurants [42]. The businesses with higher ratings and more reviews are considered more likely to be accepted by customers than businesses with poor ratings and a limited number of reviews.

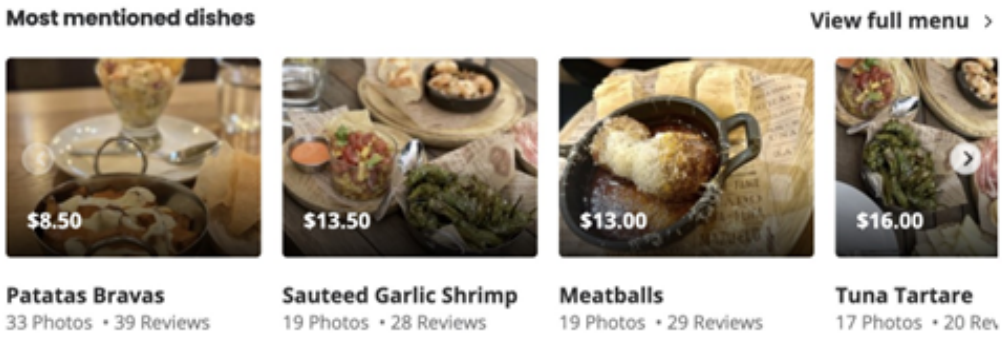
RQ2: How Can We Use Calorie Information to Quantify State-Level Food Environments?

Because calorie intake is one of the major contributors to obesity, it is critical to understand the nutritional content of food to evaluate its effect on obesity. We evaluated the state-level food environment quantitatively using the nutritional information, specifically calorie information, collected from MFP. The categories were turned into average calories per gram for popular dishes in representative restaurants. The caloric density of each food category, which was weighted by the availability of that category in a state, became the weighted score of the caloric density of the state.

To calculate the caloric density for each category, we first collected popular dishes in each category. We chose the top 100 restaurants with the highest number of reviews for each category nationwide and used the web scraping tool, BeautifulSoup, to collect the popular dishes. Yelp [35] listed the most mentioned dishes for each restaurant on the Yelp [35] pages (Figure 3). Subsequently, these popular dishes were searched in the MFP food nutrition database.

We calculated the mean calorie content of a popular dish by averaging the calories per gram of all records returned from MFP for that dish. It should be noted that the nutrition database of MFP contains a combination of foods added by MFP and foods that are added by users, and various units of measures (eg, g, gram, package, breast, oz, piece, and slices) are used. We selected gram as the unified measuring unit for comparison. We included all records that use "gram" or variations of "gram" (eg, "g," "gr," and "grams") as their measuring unit.

Figure 3. Example of the Yelp page.



RQ3: Can We Predict State-Level Obesity Rates Using Calorie Information of Different Restaurant Categories and Dimensions From the Concept of Access Using Publicly Available Social Media Data?

On the basis of the results of RQ1 and RQ2, we created features from the availability, affordability, and acceptability of food categories and state weighted score for caloric density for the state-level food environment to describe the local food environment.

We classified these features into 3 sets: (1) category availability: the degree of availability of each category at the state level; (2) category affordability and acceptability: the average price of, average rating of, and average number of reviews for each category at the state level; and (3) state weighted score for caloric density: calculated weighted score for caloric density for each state. We used the scikit-learn [43] library to build our machine learning models. We applied a combination of different feature sets and used several popular machine learning models (ie, random forest regression, support vector machine regression, and XGBoost regression) for prediction. We did not use the state-of-the-art deep learning models (eg, convolutional neural network regression) in this study because we had a limited number of samples. Deep learning models would need a large sample size to outperform traditional machine learning techniques [42]. Because we were predicting obesity rate at the state level, we used the leave-one-out cross-validation. Leave-one-out cross-validation is an extreme version of k-fold cross-validation, where k is set to N. N is the number of observations in the data set. For N times, a model is created and trained on all the data except for 1 point, and a prediction is made for that point. Thus, we used information from the District of Columbia and 49 states to predict the obesity rate for the other state. Then, we repeated this 50 times while changing the predicting location. We evaluated our approach by calculating the Pearson correlation between the real and predicted obesity rates.

Results

RQ1: Is There a Difference Between the Available Food Categories in States With Low and High Obesity Prevalence?

We extracted business profile data of the food-related businesses located in the 2 preselected areas from the collected Yelp data. A summary of the data is presented in Table 1. First, we calculated the availability of each category in the given areas. In Mississippi, the categories with high availability included "Fast Food," "Burgers," "Seafood," and "Sandwiches." In Colorado, the categories with high availability were "Mexican," "Breakfast and Brunch," "Sandwiches," and "Burgers." The "Sandwiches" and "Burgers" categories had high availability in both Mississippi and Colorado. We further explored the differences in the availability of each category to understand the state-level food environment in both state with low obesity prevalence and state with high obesity prevalence. This was also done to highlight the importance of access to different types of food. We used the net value to measure the availability differences between the 2 different locations. The net differences were used to rank the categories in descending order.

Results for the net differences are listed in Table 2. A larger net value indicated a bigger difference. The net difference for all categories is significantly different by the *z* test. We found that 42.7% (59/138) of categories showed significant differences between the 2 states.

As shown in Table 2, a total of 40% (16/40) of categories are more significantly available in Mississippi than in Colorado ($P \leq .001$), including "Fast Food," "Buffets," and "Donuts." "Diners" and "Chinese" are more significantly available in Mississippi than in Colorado ($P \leq .01$). "Ice Cream and Frozen Yogurt" is also found to be more available in Mississippi; however, the difference is not as significant as the aforementioned categories based on *P* values.

Alcohol-related businesses, including "Breweries," "Cocktail Bars," "Beer Bar," "Wine Bars," and "Pubs," were found to be significantly more available in Colorado. Moreover, "Breakfast and Brunch," "Coffee and Tea," "Mexican," "American (new)," "Pizza," "Food Truck," "Vietnamese," "Thai," "Asian Fusion," "Ramen," "Juicy Bars and Smoothies," "Indian," and "Cafes" were also found to be more available in Colorado than in Mississippi at $P \leq .001$. "Bakeries" and "Beer, Wine, and Spirits" were more available in Colorado than in Mississippi ($P \leq .01$).

"Fast Food" was found to have the biggest availability difference between Colorado and Mississippi. We further explored this category to fully understand the state-level food environment and the importance of access to different types of food. The availability of "Fast Food" in Mississippi was 13.49% (519/3845), whereas the availability of "Fast Food" in Colorado was 5.03% (358/7109). Because fast food was found to have the biggest difference in availability, we investigated the relationship between the availability of fast-food restaurants and the state-level obesity rate.

We visualized the availability of fast-food restaurants in a map (Figure 4, left) and scatter plot to show the relationship between the availability of fast-food restaurants and the prevalence of state-level obesity (Figure 4, right). We found that the availability of fast-food restaurants was positively correlated with the obesity rate at the state level, with a resulting Pearson correlation of 0.676. From the heat map, we also found that the northeast had the lowest availability of fast food, and the Midwest and south had a higher availability of fast food than the west. We further adopted dimensions from *the concept of access* to compare fast-food restaurants with other restaurants.

We compared the acceptability (rating and number of reviews; Figure 5) and affordability (price; Figure 6) between fast-food and other restaurants.

In Figures 5 and 6, the x-axis shows the state-level obesity rate, and each vertical line represents a state with its corresponding obesity rate. The blue and orange solid lines are the average rating and average number of reviews (Figure 5) and average price (Figure 6) based on restaurant type in the state, and the shadow of each line is the CI. Results showed that the acceptability of fast-food restaurants was lower than that of other restaurants, irrespective of the prevalence of obesity. We found that the average rating of fast-food restaurants showed a negative relationship with the obesity rate at the state level. The residents in areas with high obesity rate gave fast-food restaurants a lower rating than the residents in areas with low obesity rate. We also found that the range of the number of reviews showed a negative relationship with obesity rate. Results on affordability showed that the price level of fast-food restaurants was lower than that of other restaurants. In addition, the prices in fast-food restaurants and other restaurants had similar trends, which indicated that the prices in fast-food restaurants are affected by the local price indices.

Table 1. A summary of the collected data for Colorado and Mississippi.

	Region	
	Colorado	Mississippi
Business, n (%)	7109 (2.01)	3845 (1.09)
Business categories, n (%)	215 (95.1)	142 (62.8)
Rating, mean (SD)	4.02 (0.74)	3.83 (0.96)
Reviews, mean (SD)	106.59 (197.71)	22.05 (50.14)
Price (US $), mean (SD)	1.66 (0.57)	1.50 (0.55)

Table 2. The 40 categories with the highest availability difference between Colorado (low obesity rate) and Mississippi (high obesity rate).

Category	Net value
Fast food[a]	0.0844[b]
Seafood[a]	0.0824[b]
Breakfast and brunch	0.0679[b]
Burgers[a]	0.0493[b]
Southern[a]	0.0470[b]
Mexican	0.0423[b]
Bars	0.0415[b]
Chicken wings[a]	0.0364[b]
American (new)	0.0353[b]
Steakhouses[a]	0.0298[b]
Pizza	0.0278[b]
Food trucks	0.0275[b]
Breweries	0.0235[b]
Buffets[a]	0.0227[b]
Coffee and tea	0.0216[b]
Cajun or creole[a]	0.0204[b]
Cafes	0.0184[b]
Cocktail bars	0.0177[b]
Convenience stores[a]	0.0175[b]
Barbeque[a]	0.0170[b]
Soul food[a]	0.0170[b]
Vietnamese	0.0156[b]
Restaurants[a]	0.0149[b]
Italian[a]	0.0115[c]
Beer bar	0.0111[b]
Thai	0.0108[b]
Bakeries	0.0105[c]
Asian fusion	0.0103[b]
Chinese[a]	0.0098[c]
Japanese[a]	0.0097[b]
Wine bars	0.0094[b]
Ramen	0.0089[b]
Pubs	0.0085[b]
Juice bars and smoothies	0.0082[b]
Tex-Mex[a]	0.0081[b]
Donuts[a]	0.0079[b]

Category	Net value
Indian	0.0078[b]
Beer, wine, and spirits	0.0075[c]
Diners[a]	0.0075[c]
Ice cream and frozen yogurt[a]	0.0071[d]

[a]This category is more available in Mississippi, which has a higher obesity rate than Colorado.
[b]$P \leq .001$.
[c]$P \leq .01$.
[d]$P \leq .05$.

Figure 4. The relationship between the availability of fast-food restaurants and the state-level obesity rate. Left: availability of fast-food restaurants in a map; Right: scatter plot with the relationship between the availability of fast-food restaurants and the prevalence of state-level obesity.

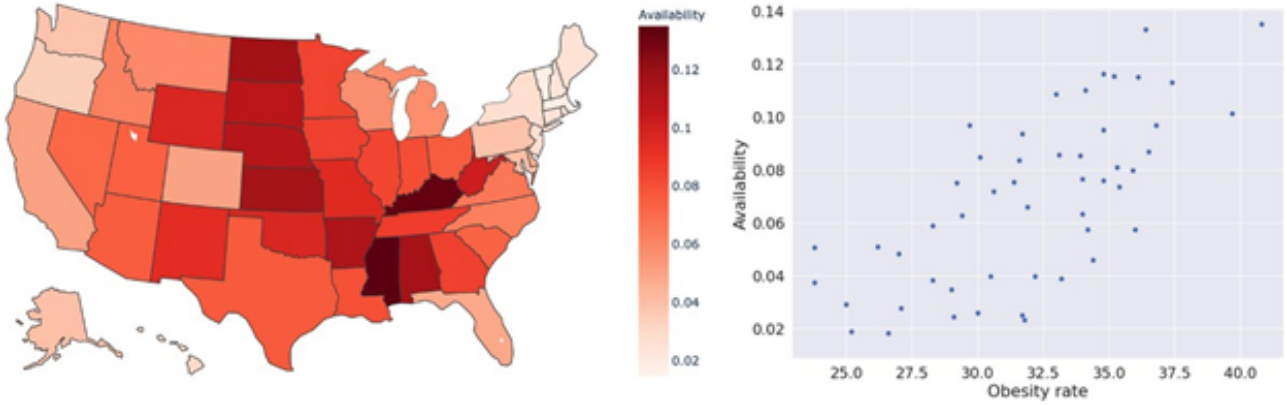


Figure 5. The relationship between the acceptability of restaurant type and the state-level obesity rate. Left: The relationship between the average rating of restaurant type and the state-level obesity rate; Right: The relationship between the average price of restaurant type and the state-level obesity rate.

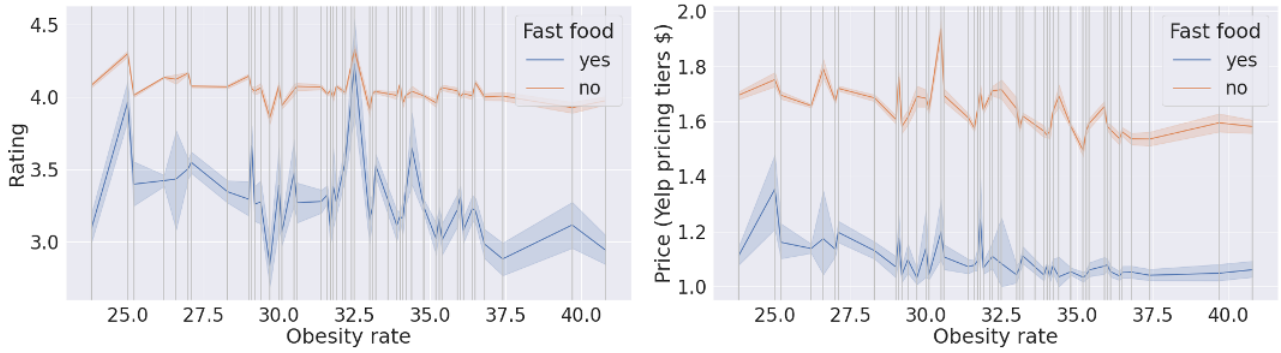
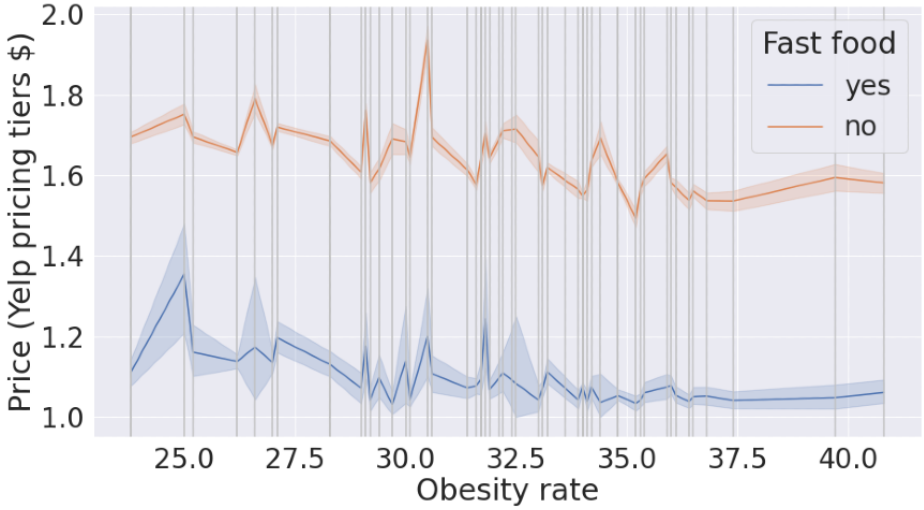


Figure 6. The relationship between the affordability of restaurant type and the state-level obesity rate.



RQ2: How Can We Use Calorie Information to Quantify State-Level Food Environments?

The first step in quantifying a food environment was to collect the popular dishes of each category. The popular dishes of the food categories gave us an idea of why some categories were more popular in areas with high obesity. We listed the most popular dishes of the categories that we found in RQ1 to be more popular in Mississippi (Table 3) and of those that we found in RQ1 to be more popular in Colorado (Table 3). Fried food in Colorado is not as popular as in Mississippi. We collected 12,316 popular dishes for the categories that were more available in Mississippi, of which 120 (1.2%) were fried chicken. In categories that were more available in Colorado, 0.44% (114/25,910) of the popular dishes were fried chicken. The statistical test showed that the difference in proportions between the fried chicken in Mississippi and the fried chicken in Colorado was significant with *a P* value less than the significant level of .001. Similarly, the percentage of other fried foods, such as fried catfish, fried shrimp, chicken, fried steak, and fried oysters, was significantly higher in Mississippi than in Colorado. This finding is consistent with literature studies showing that the intake of fried food is associated with obesity [39].

The second step was to calculate the caloric density of each category based on the calorie information of all the available popular dishes. On average, there were 166 popular dishes per category. Table 4 shows the 5 most popular dishes per category along with the caloric density of each dish and each category. We collected up to 100 most popular (ie, highest number of reviews) restaurants in each category. A table containing the caloric densities of all categories is provided in Multimedia Appendix 1.

We further calculated the caloric density of each popular dish. The caloric density of the dishes ranged from 0.556 to 62.383, with a median value of 2.399. Bakery food had a relatively high caloric density. For example, the caloric densities of almond croissant and pecan pie were >4. Fatty meat also had a high caloric density. The caloric density of Peking duck reached 8.847, which is even higher than that of fried chicken. Cooking method also affected the caloric density. For example, the caloric density for poached egg was 1.414, for scrambled egg was 1.649, and for Eggs Benedict was 2.208; likewise, the calories per gram for fried catfish was 3.283 and for fresh fish was 1.188. Salad and soup were found with low caloric densities. The calories per gram for beet salad and French onion soup were <1 based on our calculation.

Using the calorie information of these popular dishes, we calculated the caloric density of each category by averaging the caloric density of all popular dishes. The caloric density of a category varied from 1.941 to 23.452, with a median value of

5.473. The “Cheesesteaks” was the category with the highest caloric density, followed by the “Fried Chicken” with a caloric density of 17.310. “Fruits and Veggies,” “Food Tours,” “Shaved Snow,” “Gay Bars,” and “Honey” were categories with the lowest caloric density among all food categories, with caloric density <4.

Finally, we converted the caloric density for each category into a weighted score for caloric density for each state. The estimated weighted score for caloric density for the states ranged from 5.786 to 6.430. Washington had the lowest estimated weighted score for caloric density, while Georgia had the highest estimated weighted score for caloric density among all the states. Colorado’s score was 5.955, and Mississippi’s score was 6.305. We performed a 2-sample *z* test between these 2 states. The result showed a significant difference with a *z* value of 12.759 and *P*<.001. The relationship between the state estimated weighted score for caloric density and state obesity rate is shown in Figure 7. The estimated weighted score for the caloric density of states calculated using our approach showed a strong positive correlation (*r*=0.671; *P*<.001) with the state-level obesity rate. A higher estimated weighted score for the caloric density of a state indicates that the state-level food environment is more prone to obesity by serving high–calorie density food. Moreover, the estimated caloric density weighted score for southern food is higher than those for other areas in the United States, especially in Georgia, Alabama, and Mississippi.

Table 3. The most popular dishes for categories more available in Colorado and Mississippi.

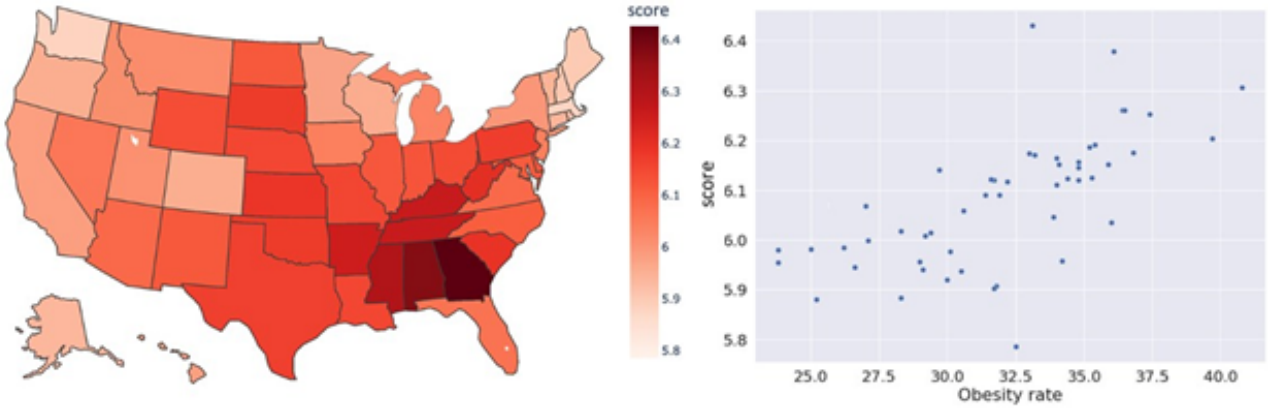
Regions	Popular dishes
Mississippi	• Fried Chicken • French Toast • Fish Tacos • Clam Chowder • Crab Cakes • Fried Catfish • Eggs Benedict • Fish and Chips • Filet Mignon • Beef Brisket
Colorado	• French Toast • Fish Tacos • Pork Belly • Eggs Benedict • Pad Thai • Fish and Chips • Fried Chicken • Spring Rolls • Caesar Salad • Avocado Toast

Table 4. The example of top 5 popular dishes and their caloric density for selected categories.

Category	Popular dish 1 (caloric density)	Popular dish 2 (caloric density)	Popular dish 3 (caloric density)	Popular dish 4 (caloric density)	Popular dish 5 (caloric density)	Caloric density for the category
Chicken wings	Fried chicken (2.240)	Boneless wings (1.836)	Buffalo wings (2.02)	Kimchi fried rice (3.271)	Chicken strips (2.108)	17.31
Diners	French toast (2.545)	Eggs Benedict (2.208)	Chicken fried steak (2.665)	Huevos rancheros (1.147)	Scrambled eggs (1.649)	7.289
Soul food	Fried chicken (2.240)	Fried catfish (3.283)	Sweet potato pie (2.525)	Red beans and rice (1.880)	Chicken breast (1.453)	6.337
Patisserie or cake shop	Almond croissant (4.102)	Chocolate croissant (3.926)	French toast (2.545)	Eggs Benedict (2.208)	Tiramisu (3.034)	6.298
Southern	Fried chicken (2.240)	Fried catfish (3.283)	Pecan pie (4.749)	Pork chop (1.590)	French toast (2.545)	5.667
Smokehouse	Pulled pork sandwich (2.452)	Baby back ribs (2.301)	Beef brisket (2.043)	Brisket sandwich (2.698)	Pulled pork (2.112)	5.51
American (new)	French toast (2.545)	Eggs Benedict (2.208)	Poached egg (1.414)	Fish tacos (1.498)	Beet salad (0.845)	5.047
Brasseries	French onion soup (0.808)	Pork chop (1.590)	Steak frites (2.465)	Duck confit (2.646)	Beef tartare (2.698)	4.78
Poke	Poke bowl (1.482)	Seaweed salad (3.510)	Spicy tuna (1.955)	Octopus (1.838)	Fresh fish (1.188)	4.716
Dim sum	Shrimp dumplings (1.620)	Peking duck (8.847)	BBQ[a] pork buns (2.505)	Har gow (1.741)	Xiao Long Bao (2.419)	4.215

[a]BBQ: barbecue.

Figure 7. The weighted score for caloric density of each state. Left: The weighted score for caloric density in a map; Right: scatter plot with the relationship between the weighted score for caloric density and the prevalence of state-level obesity.



RQ3: Can We Predict State-Level Obesity Rates Using Publicly Available Social Media Data?

We generated 3 sets of features for the prediction. The feature sets were as follows: (1) category availability, (2) category affordability and acceptability, and (3) weighted score for caloric density. Affordability and acceptability were created at the state level for the identified 226 categories. The estimated state weighted score for caloric density was calculated in RQ2. Because each state had only 1 estimated weighted score for caloric density, prediction models other than linear regression were not applicable for prediction using this set of features. For categories that did not exist in a state, we used 0 to fill in the missing values for the categories' availability, affordability, and acceptability. Approximately 24% (11,065/46,104) of the features were filled with 0. Table 5 presents the results of comparing different prediction models with different combinations of input. We used the Pearson correlation coefficient between the actual obesity rate and predicted obesity rate to evaluate it.

The random forest model with all 3 sets of features performed the best. In addition, the Pearson correlation coefficient between the predicted and real obesity rates was 0.796, which indicates that the predicted value was correlated with the real value.

Table 5. Pearson correlation coefficients for different combinations of input for prediction.

Features	Linear regression	Random forest regression	SVM[a] regression	XGBoost regression
Category availability	0.407	0.763	0.712	0.742
Category affordability and acceptability	0.402	0.776	0.593	0.743
State weighted score for caloric density	0.622	—[b]	—	—
Category availability+category affordability and acceptability	0.403	0.791	0.642	0.731
Category availability+state weighted score for caloric density	0.336	0.771	0.714	0.710
Category availability+category affordability and acceptability+state weighted score for caloric density	0.402	0.796[c]	0.643	0.708

[a]SVM: support vector machine.

[b]Not available.

[c]The best performing model.

Discussion

Principal Findings

In this study, we characterized food environments using the data from Yelp and MFP with innovative data collection and processing methods. We also predicted state-level obesity rates. In addition, our study contributed a new method to calculate food environment and data to estimate the calorie densities of different popular dishes and restaurant categories for future studies.

Our results showed a disparity in the available food categories between Colorado and Mississippi (ie, Colorado had a low obesity rate, and Mississippi had a high obesity rate). "Fast-food" restaurants were found to be more available in Mississippi than in Colorado. Fast-food consumption has been found to be strongly associated with weight gain and obesity [3]. Individual-level diet and weight outcomes are thought to improve in neighborhoods that have access to high-quality food [44]. Comparing the state-level food availability difference, we found that abundant access to fast-food options may contribute to a negative group-level health outcome. Although fast-food restaurants are notorious for serving high-calorie, low-nutritional foods [24,25] such as hamburgers, French fries, and fish and chips [45], some differences have been found. By comparing the popularity of fast-food restaurants with other restaurants in

Figure 5, we found that fast-food restaurants always have a lower number of reviews than other restaurants. However, in the District of Columbia, the average number of reviews of fast-food restaurants is higher than that of other restaurants. This may be because more alternative fast foods are available in cities, such as salad, sushi, and poke, which are considered light and healthy [46].

In addition to using the available food category to characterize the state food environment, we also used the popular dish and nutrition content of popular dishes to quantify the state food environment. To our knowledge, this is the first study to conduct a large-scale analysis of popular dishes. We compared popular dishes in Colorado and Mississippi. We found that fried foods are more popular in Mississippi. This finding is consistent with the literature showing that the intake of fried food is associated with obesity [47]. Using the collected popular dishes, we calculated the weighted score for caloric density for each state. Similar studies exist. For example, Nguyen et al [16] quantified the state food environment by calculating the caloric density of food mentions in geo-tagged tweets. They used a list of more than 1430 popular foods and beverages from the US Department of Agriculture's National Nutrient Database and calculated calories per 100 g for each food item [16]. Abbar et al [19] calculated the average calories by checking the calories per serving for the selected 500 food keywords. In contrast to these 2 studies, we used MFP, the biggest food database available [38], to obtain nutrition data. We collected nutrition data for 37,295 dishes, which allowed for an effective use of data points. In our study, Pearson correlation of weighted score for caloric density of states to state obesity rates was 0.671, which outperformed one of the aforementioned previous studies [19] in which the Pearson correlation of tweet caloric value to state obesity rates was 0.629.

To the best of our knowledge, our prediction model is the first to use Yelp and MFP data to predict state obesity rates. In contrast to previous studies that used Twitter data to predict obesity rate [17-19], our model using Yelp and MFP data had less selection bias. First, Twitter users are younger than the general public [48]; however, the user group of Yelp is more evenly distributed by age, with 33% of the users aged ≥55 years [49]. Second, the previous studies using Twitter data for prediction only used sampled data because of the massive amount of Twitter data. Although these studies used the same data source, their collection methods were different, which could have skewed the results.

Public Health Implications

Our study helped us understand the impact of the food environment and related human behavior by showing the correlation between state-level food environment and obesity rate. Because of the pervasive use of smartphones and social media apps like Yelp across the country, researchers could use social media data to gain an understanding of food environments in any part of America and other countries as well. In sum, our model has the potential to evaluate food environments.

Not only does our model map out a landscape of the local food environment but it also allows us to characterize the trajectory of public health. The copious amounts of information on social media allow public health practitioners to monitor changes in food availability and population over time and use this information to predict changes in state obesity levels. Similarly, computational methods could be used to inform dieting habits at the individual level. This allows for an early intervention in areas or individuals facing the greatest risk of increasing obesity rates or becoming obese.

Our study has reiterated a few fundamental findings related to the importance of environment [9,18,19]. Our findings suggest that those who live in areas with a considerable availability of high-calorie, fast foods are more likely to be obese. This alludes to the idea that people eat what is readily available to them. Politicians and city planners could potentially use this information to develop an infrastructure of healthy food options in areas that have been traditionally concentrated with fast-food restaurants. This sort of environmental intervention could potentially influence community behavior and lead to better health outcomes.

Limitations and Future Direction

The first limitation of our study lies in the data collection. Yelp provides substantial data for local businesses; however, the Yelp API results are restricted to 1000 results for each query. We could collect up to 1000 business data points for each zip code center up to a distance of 40 km (approximately 25 miles). In urban environments, 1 zip code may have >1000 businesses. To address this issue, we ran several rounds for each zip code and removed the duplicates. Despite this effort, missing data may skew our results, especially those about urban areas. We found a second limitation when collecting nutritional data from MFP. For each search query, MFP returned 10 pages with 10 records on each page. Some popular dishes did not have an exact match, in which case MFP returned a partially matching dish. Therefore, some caloric information may not be accurate. We averaged all the results to reduce the effects of inaccurate information. Another limitation is not capturing the actual consumption. We did not have information on the food consumed at a person's home. In this study, we calculated the caloric density of popular dishes. Nevertheless, we found that high–caloric density food is correlated with obesity rate, consistent with a previous study that was conducted at the individual level [50]. To bolster our findings, a similar analysis should be replicated at the zip code–level to better inform the local food environment. We used the state-level food environment in this study because BRFSS provides state-level obesity rate. More granular analysis will provide a better insight into how socioeconomic status and the local food environment may be correlated with obesity [14,51-53]. The information collected and calculated in this study could also be used to fuse a personalized mobile health app to help user have a better experience with obesity prevention management. For example, a specialized dashboard [54] could be added to the mobile health app when using information from GPS to measure physical activity along with a heat map showing where a person goes within their neighborhood.

Conclusions

This study used social media data to characterize state-level food environments. State-level food environments show a

disparity in the available food between states with different obesity rates, suggesting the importance of food environment. Using the availability of different categories of food along with affordability and acceptability data captured on social media, we created a state-level obesity rate prediction model with a 0.796 correlation. Using our proposed method, public health practitioners could monitor the changes in areas that face the greatest risk of increasing obesity rates to counter the obesity pandemic.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The caloric density of all food categories.
[DOCX File , 28 KB - jmir_v24i12e39340_app1.docx]

References

1. 2008-2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases: Prevent and Control Cardiovascular Diseases, Cancers, Chronic Respiratory Diseases and Diabetes. World Health Organization. 2009. URL: https://apps.who.int/iris/bitstream/handle/10665/44009/9789241597418_eng.pdf?sequence=1 [accessed 2020-10-30]
2. Thompson D, Edelsberg J, Colditz GA, Bird AP, Oster G. Lifetime health and economic consequences of obesity. Arch Intern Med 1999 Oct 11;159(18):2177-2183. [doi: 10.1001/archinte.159.18.2177] [Medline: 10527295]
3. Rosenheck R. Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk. Obes Rev 2008 Nov;9(6):535-547. [doi: 10.1111/j.1467-789X.2008.00477.x] [Medline: 18346099]
4. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017-2018. NCHS Data Brief 2020 Feb(360):1-8 [FREE Full text] [Medline: 32487284]
5. Fryar CD, Carroll MD, Ogden CL. Prevalence of Overweight, Obesity, and Extreme Obesity Among Adults Aged 20 and Over: United States, 1960–1962 Through 2013–2014. National Center for Health Statistics. 2016. URL: https://www.cdc.gov/nchs/data/hestat/obesity_adult_13_14/obesity_adult_13_14.htm [accessed 2022-11-07]
6. Hall KD. Did the food environment cause the obesity epidemic? Obesity (Silver Spring) 2018 Jan;26(1):11-13 [FREE Full text] [doi: 10.1002/oby.22073] [Medline: 29265772]
7. Bleich SN, Cutler D, Murray C, Adams A. Why is the developed world obese? Annu Rev Public Health 2008;29:273-295. [doi: 10.1146/annurev.publhealth.29.020907.090954] [Medline: 18173389]
8. Kraak VI. Comprehensive restaurant-sector changes are essential to reduce obesity risk for all Americans. Am J Public Health 2018 Feb;108(2):158-159. [doi: 10.2105/AJPH.2017.304217] [Medline: 29320299]
9. Caspi CE, Sorensen G, Subramanian SV, Kawachi I. The local food environment and diet: a systematic review. Health Place 2012 Sep;18(5):1172-1187 [FREE Full text] [doi: 10.1016/j.healthplace.2012.05.006] [Medline: 22717379]
10. Market Intel. U.S. Food Expenditures at Home and Abroad. American Farm Bureau Federation. 2019 Nov 13. URL: https://web.archive.org/web/20220328000522/https://www.fb.org/market-intel/u.s.-food-expenditures-at-home-and-abroad [accessed 2022-03-22]
11. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CA. The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. Obesity (Silver Spring) 2015 Jul;23(7):1331-1344 [FREE Full text] [doi: 10.1002/oby.21118] [Medline: 26096983]
12. Morland KB, Evenson KR. Obesity prevalence and the local food environment. Health Place 2009 Jun;15(2):491-495 [FREE Full text] [doi: 10.1016/j.healthplace.2008.09.004] [Medline: 19022700]
13. Maddock J. The relationship between obesity and the prevalence of fast food restaurants: state-level analysis. Am J Health Promot 2004;19(2):137-143. [doi: 10.4278/0890-1171-19.2.137] [Medline: 15559714]
14. Block JP, Scribner RA, DeSalvo KB. Fast food, race/ethnicity, and income: a geographic analysis. Am J Prev Med 2004 Oct;27(3):211-217. [doi: 10.1016/j.amepre.2004.06.007] [Medline: 15450633]
15. Spence JC, Cutumisu N, Edwards J, Raine KD, Smoyer-Tomic K. Relation between local food environments and obesity among adults. BMC Public Health 2009 Jun 18;9:192 [FREE Full text] [doi: 10.1186/1471-2458-9-192] [Medline: 19538709]
16. Nguyen QC, Meng H, Li D, Kath S, McCullough M, Paul D, et al. Social media indicators of the food environment and state health outcomes. Public Health 2017 Jul;148:120-128 [FREE Full text] [doi: 10.1016/j.puhe.2017.03.013] [Medline: 28478354]
17. Fried D, Surdeanu M, Kobourov S, Hingle M, Bell D. Analyzing the language of food on social media. In: Proceedings of the 2014 IEEE International Conference on Big Data. 2014 Presented at: BigData '14; October 27-30, 2014; Washington, DC, USA p. 778-783. [doi: 10.1109/bigdata.2014.7004305]
18. Culotta A. Estimating county health statistics with Twitter. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Apr Presented at: CHI '14; April 26-May 1, 2014; Toronto, Canada p. 1335-1344. [doi: 10.1145/2556288.2557139]

19. Abbar S, Mejova Y, Weber I. You Tweet what you eat: studying food consumption through Twitter. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15; April 18-23, 2015; Seoul, South Korea p. 3197-3206. [doi: 10.1145/2702123.2702153]
20. Duffey KJ, Popkin BM. Energy density, portion size, and eating occasions: contributions to increased energy intake in the United States, 1977-2006. PLoS Med 2011 Jun;8(6):e1001050 [FREE Full text] [doi: 10.1371/journal.pmed.1001050] [Medline: 21738451]
21. Poti JM, Popkin BM. Trends in energy intake among US children by eating location and food source, 1977-2006. J Am Diet Assoc 2011 Aug;111(8):1156-1164 [FREE Full text] [doi: 10.1016/j.jada.2011.05.007] [Medline: 21802561]
22. Young LR, Nestle M. The contribution of expanding portion sizes to the US obesity epidemic. Am J Public Health 2002 Feb;92(2):246-249. [doi: 10.2105/ajph.92.2.246] [Medline: 11818300]
23. Hruschka DJ, Brewis AA, Wutich A, Morin B. Shared norms and their explanation for the social clustering of obesity. Am J Public Health 2011 Dec;101 Suppl 1:S295-S300. [doi: 10.2105/AJPH.2010.300053] [Medline: 21555656]
24. Isganaitis E, Lustig RH. Fast food, central nervous system insulin resistance, and obesity. Arterioscler Thromb Vasc Biol 2005 Dec;25(12):2451-2462. [doi: 10.1161/01.ATV.0000186208.06964.91] [Medline: 16166564]
25. Pereira MA, Kartashov AI, Ebbeling CB, Van Horn L, Slattery ML, Jacobs Jr DR, et al. Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. Lancet 2005;365(9453):36-42. [doi: 10.1016/S0140-6736(04)17663-0] [Medline: 15639678]
26. Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. Med Care 1981 Feb;19(2):127-140. [doi: 10.1097/00005650-198102000-00001] [Medline: 7206846]
27. Charreire H, Casey R, Salze P, Simon C, Chaix B, Banos A, et al. Measuring the food environment using geographical information systems: a methodological review. Public Health Nutr 2010 Nov;13(11):1773-1785. [doi: 10.1017/S1368980010000753] [Medline: 20409354]
28. McKinnon RA, Reedy J, Morrissette MA, Lytle LA, Yaroch AL. Measures of the food environment: a compilation of the literature, 1990-2007. Am J Prev Med 2009 Apr;36(4 Suppl):S124-S133. [doi: 10.1016/j.amepre.2009.01.012] [Medline: 19285203]
29. Li C, Ademiluyi A, Ge Y, Park A. Using social media to understand web-based social factors concerning obesity: systematic review. JMIR Public Health Surveill 2022 Mar 07;8(3):e25552 [FREE Full text] [doi: 10.2196/25552] [Medline: 35254279]
30. Oduru T, Jordan A, Park A. Healthy vs. unhealthy food images: image classification of Twitter images. Int J Environ Res Public Health 2022 Jan 14;19(2):923 [FREE Full text] [doi: 10.3390/ijerph19020923] [Medline: 35055742]
31. Park A. Tweets related to motivation and physical activity for obesity-related behavior change: descriptive analysis. J Med Internet Res 2022 Jul 20;24(7):e15055 [FREE Full text] [doi: 10.2196/15055] [Medline: 35857347]
32. Pachucki MA, Jacques PF, Christakis NA. Social network concordance in food choice among spouses, friends, and siblings. Am J Public Health 2011 Nov;101(11):2170-2177 [FREE Full text] [doi: 10.2105/AJPH.2011.300282] [Medline: 21940920]
33. Yelp Dataset JSON. Yelp API. URL: https://web.archive.org/web/20210609105605/https://www.yelp.com/%0Adataset/documentation/main/ [accessed 2021-06-07]
34. Yelp. Wikipedia - The Free Encyclopedia. URL: https://en.wikipedia.org/wiki/Yelp [accessed 2021-06-07]
35. Yelp. URL: https://www.yelp.com/ [accessed 2022-11-08]
36. Yelp's Alexa Rank. Alexa Internet. URL: https://web.archive.org/web/20210611025251/https://www.alexa.com/siteinfo/yelp.com [accessed 2021-06-10]
37. beautifulsoup4 4.11.1. Python Package Index. URL: https://pypi.org/project/beautifulsoup4/ [accessed 2021-06-07]
38. MyFitnessPal: Calorie Counter. Google Play. URL: https://web.archive.org/web/20210211213534/https://play.%0Agoogle.com/store/apps/details?id=com.myfitnesspal.android&hl=en_US&gl=US [accessed 2022-11-07]
39. Adult Obesity Prevalence Maps. Centers for Disease Control and Prevention. 2019. URL: https://web.archive.org/web/20210611091817/https://www.%0Acdc.gov/obesity/data/prevalence-maps.html [accessed 2021-06-10]
40. Mazidi M, Speakman JR. Higher densities of fast-food and full-service restaurants are not associated with obesity prevalence. Am J Clin Nutr 2017 Aug;106(2):603-613. [doi: 10.3945/ajcn.116.151407] [Medline: 28566310]
41. Viglia G, Furlan R, Ladrón-de-Guevara A. Please, talk about it! When hotel popularity boosts preferences. Int J Hosp Manag 2014 Sep;42:155-164. [doi: 10.1016/j.ijhm.2014.07.001]
42. Zhang Z, Ye Q, Law R, Li Y. The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews. Int J Hosp Manag 2010 Dec;29(4):694-700. [doi: 10.1016/j.ijhm.2010.02.002]
43. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011 Jan 2;12:2825-2830.
44. Gordon-Larsen P. Food availability/convenience and obesity. Adv Nutr 2014 Nov;5(6):809-817 [FREE Full text] [doi: 10.3945/an.114.007070] [Medline: 25398746]
45. Fast food. Wikipedia - The Free Encyclopedia. URL: https://en.wikipedia.org/wiki/Fast_food [accessed 2021-06-01]
46. Why Poke Is Emerging as the Next Fast-Casual Restaurant Phenomenon. Money.com. 2022. URL: https://web.archive.org/web/20201030051253/https://money.com/restaurant-trends-poke-fast-casual-chipotle/ [accessed 2020-10-30]
47. Gadiraju TV, Patel Y, Gaziano JM, Djoussé L. Fried food consumption and cardiovascular health: a review of current evidence. Nutrients 2015 Oct 06;7(10):8424-8430 [FREE Full text] [doi: 10.3390/nu7105404] [Medline: 26457715]

48. Wojcik S, Hughes A. Sizing Up Twitter Users. Pew Research Center. 2019 Apr 24. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/2019/04/twitter_opinions_4_18_final_clean.pdf [accessed 2022-11-07]
49. Yelp Fast Facts. Yelp. 2021 Mar 31. URL: https://web.archive.org/web/20210707191209/https://www.%0Ayelp-press.com/company/fast-facts/default.aspx [accessed 2021-06-10]
50. Bassett MT, Perl S. Obesity: the public health challenge of our time. Am J Public Health 2004 Sep;94(9):1477. [doi: 10.2105/ajph.94.9.1477] [Medline: 15359488]
51. Chang VW, Lauderdale DS. Income disparities in body mass index and obesity in the United States, 1971-2002. Arch Intern Med 2005 Oct 10;165(18):2122-2128. [doi: 10.1001/archinte.165.18.2122] [Medline: 16217002]
52. Mobley LR, Root ED, Finkelstein EA, Khavjou O, Farris RP, Will JC. Environment, obesity, and cardiovascular disease risk in low-income women. Am J Prev Med 2006 Apr;30(4):327-332. [doi: 10.1016/j.amepre.2005.12.001] [Medline: 16530620]
53. Fleischhacker SE, Evenson KR, Rodriguez DA, Ammerman AS. A systematic review of fast food access studies. Obes Rev 2011 May;12(5):e460-e471. [doi: 10.1111/j.1467-789X.2010.00715.x] [Medline: 20149118]
54. Filos D, Lekka I, Kilintzis V, Stefanopoulos L, Karavidopoulou Y, Maramis C, et al. Exploring associations between children's obesogenic behaviors and the local environment using big data: development and evaluation of the obesity prevention dashboard. JMIR Mhealth Uhealth 2021 Jul 09;9(7):e26290 [FREE Full text] [doi: 10.2196/26290] [Medline: 34048353]

Abbreviations

API: application programming interface
BRFSS: Behavioral Risk Factor Surveillance System
GIS: geographic information system
MFP: MyFitnessPal
RQ: research question

Edited by R Kukafka; submitted 06.05.22; peer-reviewed by N Maglaveras, Z Butt, S Jia; comments to author 29.07.22; revised version received 09.09.22; accepted 26.09.22; published 13.12.22.

Please cite as:
Li C, Jordan A, Song J, Ge Y, Park A
A Novel Approach to Characterize State-level Food Environment and Predict Obesity Rate Using Social Media Data: Correlational Study
J Med Internet Res 2022;24(12):e39340
URL: https://www.jmir.org/2022/12/e39340
doi:10.2196/39340
PMID:36512396

Original Paper

Disinfection of Virtual Reality Devices in Health Care Settings: In Vitro Assessment and Survey Study

Scott C Roberts[1,2], MD, MSc; Nancy L Havill[2], MT, MHA, CIC; Rosa M Flores[2], MD, MPH; Curtis Anthony Hendrix II[1], MD; Maximillian J Williams[3], BA; Richard S Feinn[4], PhD; Steven J Choi[2], MD; Richard A Martinello[1,2,3], MD; Asher M Marks[3], MD; Thomas S Murray[3], MD, PhD

[1]Section of Infectious Diseases, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States

[2]Department of Infection Prevention, Yale New Haven Hospital, New Haven, CT, United States

[3]Department of Pediatrics, Yale School of Medicine, New Haven, CT, United States

[4]Department of Medical Sciences, Frank H Netter MD School of Medicine, Quinnipiac University, North Haven, CT, United States

Corresponding Author:
Scott C Roberts, MD, MSc
Section of Infectious Diseases
Department of Internal Medicine
Yale School of Medicine
20 York Street
Hunter 527
New Haven, CT, 06510
United States
Phone: 1 240 328 9031
Fax: 1 203 688 2823
Email: scott.c.roberts@yale.edu

Abstract

Background: Virtual reality (VR) devices are increasingly used in health care settings. The use among patients has the potential to unintentionally transmit pathogens between patients and hospital staff. No standard operating procedure for disinfection exists to ensure safe use between patients.

Objective: This study aims to determine the efficacy of disinfectants on VR devices in order to ensure safe use in health care settings.

Methods: Three types of bacteria were inoculated onto porous and nonporous surfaces of 2 VR devices: the Meta Oculus Quest and Meta Oculus Quest 2. Disinfection was performed using either isopropyl alcohol or alcohol-free quaternary ammonium wipes. A quantitative culture was used to assess the adequacy of disinfection. A survey was separately sent out to VR device technicians at other pediatric health care institutes to compare the methods of disinfection and how they were established.

Results: Both products achieved adequate disinfection of the treated surfaces; however, a greater log-kill was achieved on nonporous surfaces than on the porous surfaces. Alcohol performed better than quaternary ammonium on porous surfaces. The survey respondents reported a wide variability in disinfection processes with only 1 person reporting an established standard operating procedure.

Conclusions: Disinfection can be achieved through the use of either isopropyl alcohol or quaternary ammonium products. Porous surfaces showed lesser log-kill rates than the nonporous surfaces, indicating that the use of an added barrier may be of benefit and should be a point of future research. Given the variability in the disinfection process across health care systems, a standard operating procedure is proposed.

(J Med Internet Res 2022;24(12):e42332) doi:10.2196/42332

KEYWORDS

disinfection; healthcare-acquired infection; healthcare worker; virtual reality; disinfect; occupational health; occupational safety; infection control; infection spread

Introduction

Virtual reality (VR) devices are increasingly used in health care settings to benefit patients, and the examples include patients with posttraumatic stress disorder, anxiety, complex regional pain syndrome, and distraction therapy [1-4]. Recent data show the benefit expands to the pediatric population as well by reducing pain and anxiety during medical procedures through distraction [5]. VR can also be used to educate health care workers through training and simulation [6]. However, a lack of standardized cleaning and disinfection processes for VR devices has limited VR's use in health care settings, especially during the COVID-19 pandemic [6].

Nosocomial transmission and outbreaks have been reported with many different types of medical devices in clinical use [7], and establishing standard operating procedures (SOPs) for disinfection of VR devices between patient use is paramount. One of the most common and widely sold VR headset devices worldwide, the Meta Oculus Quest 2, specifically, recommends against the use of alcohol to clean and disinfect the device in favor of antibacterial wipes due to theoretical concerns about affecting the porous material [8]. This poses challenges in clinical settings as isopropyl alcohol (IPA) is one of the most common disinfectants used on medical devices.

A protocol to clean and disinfect VR devices used in health care settings has been proposed [6]; however, no studies have quantified the efficacy of hospital-grade disinfectants on different parts of the VR equipment. Additionally, little is known about how these machines are currently disinfected in health care settings. In this mixed methods evaluation, we sought to determine the current disinfection practices in health care settings and how they were established. We also studied the effect of commonly used disinfectant wipes on the disinfection of VR headsets experimentally contaminated with common bacterial pathogens to provide evidence for the creation of an SOP to reduce infections with multipatient VR utilization.

Methods

Survey

To learn how health care facilities disinfect VR equipment and whether infection prevention teams are involved, a voluntary Qualtrics survey was sent via an electronic link in a group chat of 50 VR technicians working at children's hospitals across the United States as a convenience sample.

Ethical Considerations

The survey was approved as exempt by the Yale University institutional review board (study #2000033075).

Laboratory Disinfection

Three types of bacteria, *Staphylococcus epidermidis* (ATCC 12228), *Pseudomonas aeruginosa* (laboratory strain PAO1), and *Staphylococcus aureus* (ATCC 25923), were chosen because of their propensity to be present on the skin and cause infection in children with compromised immune systems. The bacteria were grown overnight in 3 mL of lysogeny broth and serially diluted to quantitate the bacterial inoculum. VR headsets and controllers were inoculated by spreading 10 µL (initial inoculum 4.1×10^{6}-4.5×10^{8}) onto various sites (Figure 1) and allowed to dry for 30 minutes. This large inoculum was chosen to test whether Environmental Protection Agency (EPA)–approved disinfectants achieved sufficient log-reduction in bacteria as per their instructions for use. Two VR devices were experimentally contaminated: the Oculus Quest headset and the Oculus Quest 2 headset and hand controllers (Reality Labs, Meta Platforms). These devices were chosen as these are the most popular consumer devices, the primary ones used at our institution, and contained different surface types to trial disinfection [9]. Contaminated sites included the outer surface of the headset housing on the top side, the controller buttons, and the headband straps for each device (Figure 1). These sites were chosen as they were thought to be high–touch point areas for the hands and head during patient use. We did not study the facial interface as it is our standard practice to use a disposable barrier between the facial interface and the patient's skin. Both nonporous (Oculus Quest 1 strap, Oculus Quest 2 headset, and controller) and porous (Oculus Quest 1 headset and Oculus Quest 2 strap) surfaces were contaminated to assess disinfection efficacy.

Two products were tested for active disinfection: a 70% IPA wipe (Medium Alcohol Prep Pad, Medline) and an alcohol-free quaternary ammonium wipe (Sani-Cloth AF3 Germicidal Disposable Wipe, PDI). A positive control of inoculation without disinfection was performed for every experiment and cultured after the dry time to account for bacterial cell death from desiccation. Disinfection was performed in accordance with each product's manufacturer's instructions for use. For the IPA wipe, a 15-second scrub in a back-and-forth motion followed by a 15-second dry time was performed. For the alcohol-free quaternary ammonium wipe, the experimentally contaminated area was wiped to a point of saturation for 3 minutes and allowed to dry. Following disinfection, the cultures were obtained with sterile cotton swabs dipped in Dey-Engley (D/E) neutralizing broth (Hardy Diagnostics) and wiped across the entirety of the contaminated surface for 5 seconds in a back-and-forth motion. The swab was used to inoculate a D/E agar plate that was then incubated overnight at 37 C. Bacterial colony forming units (CFUs) were counted on the plates the following day. Between experiments, the entirety of the VR devices was disinfected with a 70% IPA spray to the point of saturation of the surface materials and dried overnight. The laboratory experiments satisfied the Yale University 100 CH.9 Clinical Quality Improvement criteria and were exempt from institutional review board approval.

Figure 1. Areas where cultures were obtained (shown by red circles) of Oculus Quest (top) and Oculus Quest 2 (bottom) devices.

Statistical Analysis

The experimental design was a 3-factor crossed design with bacterial CFUs as the outcome variable. The generalized linear model with binomial distribution and logit link was used to compare whether the proportion of trials with observable bacteria counts after disinfection differed by (1) type of disinfection, (2) type of organism, and (3) surface type. All models included the natural log of bacterial count prior to disinfection and used robust standard errors. Achieving disinfection was defined as a greater than a 6-log reduction in bacterial counts. Pairwise comparisons were performed for the type of organism. Analyses were performed in SPSS (version 27; IBM Corp), and statistical significance was set at an α level of .05.

Results

Current VR Disinfection Practices in Pediatric Hospitals

A total of 50 VR technicians across the Unites States were invited to participate in the Qualtrics survey with a response rate of 18% (9/50). One person consented and then did not answer any of the questions. The selected results of the survey are shown in Table 1 and highlight the variability of VR use and disinfection practices.

Table 1. Survey of current VR[a] disinfection practices in health care settings.

Questions and responses	Respondents (N=9), n (%)
Frequency of VR use in the hospital	
<1 times per week	2 (22)
1-3 times per week	5 (55)
4-6 times per week	1 (11)
Multiple times per day	1 (11)
VR systems used in health care[b]	
Google Daydream	4 (44)
Kind VR	1 (11)
Oculus Quest 2	4 (44)
Oculus Rift	2 (22)
PlayStation VR	5 (56)
Starlight Children's VR system	6 (67)
When disinfection is performed	
Before patient use	9 (100)
Method of disinfection[b]	
Isopropyl alcohol	3 (33)
Quaternary ammonium (PDI gray top)	2 (22)
Isopropyl alcohol/quaternary ammonium (PDI purple top)	4 (44)
Hydrogen peroxide	2 (22)
Physical barriers used to prevent infection[b]	
Any physical barriers	
Silicon covers	6 (66)
Disposable eye masks	3 (33)
Wipeable replacement head straps	2 (22)
Hair covers	2 (22)
No barriers present	3 (33)
Use of standard operating procedure	
Yes	1 (11)
Inclusion of the Department of Infection Prevention in establishing disinfection technique	
Yes	2 (22)

[a]VR: virtual reality.

[b]Some respondents gave multiple responses.

The number of VR sessions varied from less than once a week to multiple times per day. Most locations used multiple types of VR platforms with the Starlight Children's VR system most commonly used. The Starlight Children's VR system is a variation of the Lenovo Mirage Solo VR headset and is made of very similar materials as the Oculus Quest 1 and Oculus Quest 2. All participants noted that disinfection was performed before patient use (n=9, 100%). The methods of disinfection were variable with IPA and quaternary ammonium low-level disinfection wipes used most commonly. Most VR technicians used physical barriers between the VR device and the patient such as combinations of silicon covers, disposable eye masks, wipeable replacement head straps, and hair covers, while 3 (33%) participants did not use barriers (Table 1). Only 1 (11%) institution had an SOP for use and disinfection. Infection prevention teams were involved in assisting with VR disinfection protocols at 2 (22%) sites.

Effectiveness of Hospital-Grade Disinfecting Wipes on VR Decontamination

A total of 175 experiments were performed to assess disinfection (Table 2). Adequate disinfection was achieved with both the IPA and the alcohol-free quaternary ammonium wipes across all bacterial types and headset material comparing untreated

with disinfected surfaces. No bacteria were recovered in 88% (154/175) of experiments. IPA wipes performed better than the quaternary ammonium wipes at reducing overall bacterial counts (P=.001). This difference was most pronounced on porous surfaces, where the mean quantity of bacteria remaining after the alcohol-free quaternary ammonium use was more than after IPA use (Figure 2A).

When comparing the disinfection by an organism, there were more CFUs of *S aureus* and *S epidermidis* recovered after attempted disinfection than *P aeruginosa* (P=.05 and P=.03, respectively), with this difference most pronounced on porous surfaces (Figure 2B). There was no significant difference between the recovery of *S aureus* and *S epidermidis* (P=.72). Finally, fewer bacteria were recovered from porous surfaces after inoculation but prior to performing disinfection. However, more bacteria were recovered after disinfection from porous surfaces than from nonporous surfaces (P=.01) as shown in Figure 2C, confirming that porous materials inoculated (Figure 1) were more difficult to disinfect because of their availability to absorb the bacteria-containing liquid.

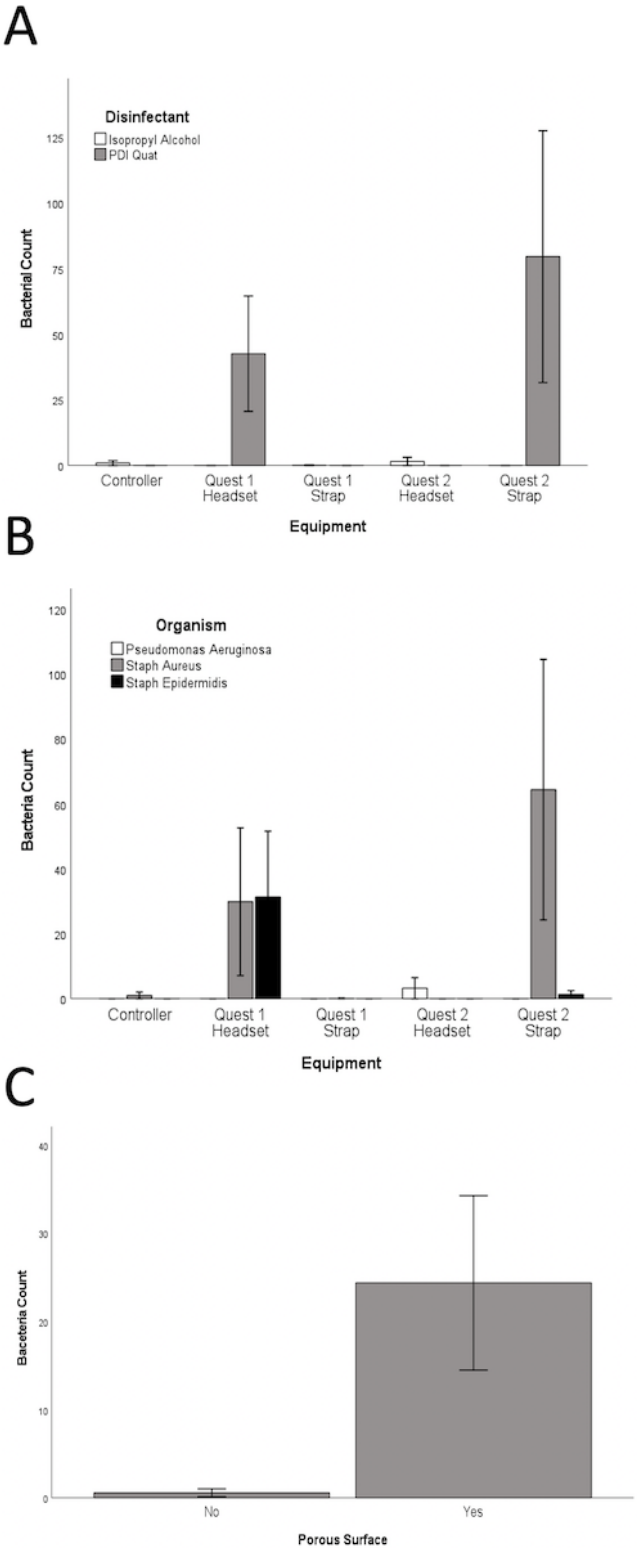
Table 2. Bacterial count by disinfection method, organism, and surface.

Factors	Observations, n	Bacterial count			
		Mean (SD)	Median	Minimum, Maximum	% Zero[a]
Disinfectant					
Isopropyl alcohol	94	0.60 (4.52)	0	0, 42	97
Alcohol-free quaternary ammonium	81	23.12 (82.4)	0	0, 500	88
Organism					
Pseudomonas aeruginosa	51	0.88 (6.06)	0	0, 42	98
Staphylococcus aureus	80	17.46 (76.5)	0	0, 500	86
Staphylococcus epidermidis	44	9.43 (41.2)	0	0, 265	82
Surface[b]					
Not porous	100	0.57 (4.4)	0	0, 42	97
Porous	75	24.4 (84.5)	0	0, 500	77

[a]The percentage of times when 0 bacterial colony-forming units were observed.

[b]Nonporous surfaces were the Quest 1 Strap, Quest 2 Headset, and controller, while the porous surfaces were the Quest 1 Headset and Quest 2 Strap.

Figure 2. The mean (SEM) raw bacterial counts recovered after disinfection according to disinfectant and equipment (A), by organism and equipment (B), and by equipment surface type (C) are displayed. Note: Non-porous surfaces were the Quest 1 Strap, Quest 2 Headset and Controller, while porous surfaces were the Quest 1 Headset and Quest 2 Strap.



Discussion

Adequate disinfection of VR devices can be achieved through the use of low-level EPA-approved hospital-grade disinfectants commonly used in clinical settings for devices that are exposed to intact skin, including IPA and quaternary ammonium wipes. We found a greater than 6-fold logarithmic reduction from initial bacteria inoculation across all pathogen types and VR device surfaces when using either product. However, we did observe the differences when evaluating raw-bacterial counts after disinfection. Notably, IPA performed better than the quaternary ammonium wipe, particularly for porous surfaces. It is possible that IPA penetrates porous surfaces better than quaternary ammonium products due to the vigorous 15-second scrub, and

future studies should evaluate how well different wipes perform on these types of surfaces. We also observed that *S aureus* and *S epidermidis* persisted on surfaces at greater densities than *P aeruginosa*, possibly reflective of a mechanism in strain type or environmental survivability. Finally, and perhaps most critically, we observed lower bacterial counts after inoculation but before disinfection, and greater bacterial counts after disinfection, on porous surfaces when compared to nonporous surfaces. This suggests that bacteria may be entering the pores in the material, potentially reducing exposure to the disinfection material. Additionally, using swabs to recover bacteria from porous surfaces is suboptimal as we do not recover bacteria that have penetrated deeper into the material as well as nonporous surfaces which have better transfer efficiency [10]. Thus, despite consistent recovery after disinfection from porous surfaces in these experiments, we likely have overestimated the efficacy of disinfection for this material. Of note, there was outstanding disinfection of all nonporous surfaces, making it the preferred material for the construction of VR devices in health care. Manufacturers should consider material in the design of both headsets and straps, and our data support the use of nonporous material, particularly in health care settings where persistent bacteria may serve as a nidus for transmission to the next VR device user. If porous surfaces are present, there should be adequate barrier protection to prevent the transmission of microbes. This also then allows for the use of IPA without concern for damage to any porous components of the device.

Nosocomial transmission and outbreaks associated with the use of medical devices are well documented and a primary concern for using VR devices in health care settings [7]. Consequently, facilities may restrict VR devices from patient use out of concern. We found substantial variability between facilities in the frequency of device use, disinfection method, and barrier protection used. Importantly, almost all sites reported that infection prevention teams were not involved in performing a risk assessment for device use during patient care, and SOPs for disinfection were absent in all but one institution. Establishing a standard process that appropriately disinfects VR devices to allow for safe and expanded use in health care settings while avoiding equipment degradation can benefit patients and health care workers alike. It is critical to ensure that when new devices such as VR equipment are introduced into patient care settings, Infection Prevention and other stakeholders are involved prior to the purchase of the devices to ensure there is an acceptable plan for device reprocessing.

Based on this generated data set, manufacturer's instructions for use [8], health care infection prevention best practices, and previous literature or expert opinion [6], we propose an SOP for VR use and disinfection in the health care setting (Textbox 1). This is particularly important as the patient population served may be undergoing chemotherapy or other immunosuppressants that can increase the risk of infection.

Textbox 1. Suggested standard operating procedure for the disinfection of virtual reality devices.

Before use

- Avoid on patients with nonintact skin or active infections on the head or hands that cannot be covered and might come into contact with the device
- Avoid use on patients known to be colonized with pathogens where specialized disinfection is required, including *Clostridioides difficile*, *Candida auris*, *Mycobacterium tuberculosis*, and nonenveloped viruses
- Patient and staff perform hand hygiene
- A nonporous cover over the face pad, a disposable face pad cover, or both should serve as a barrier between the patient's face and the device. Hair should also be covered (eg, bouffant and washable cloth surgical cap). Any porous material that makes contact with the patient's skin or hair should be covered with a barrier
- Devices should be assessed for alcohol compatibility. If the device is not alcohol compatible, a nonalcohol-based disinfectant should be used
- Perform disinfection with a device compatible Environmental Protection Agency–registered product List H [11] according to the manufacturer's instructions for use, ensuring that all surfaces of the headset (including the strap, the casing, the inner and outer facepieces, and the lens), the controller, and the nonporous, nondisposable face cover are saturated. Do not use wipes on multiple devices.

After use

- Patient and staff perform hand hygiene
- Staff don appropriate personal protective equipment, which should include nitrile gloves at a minimum unless other personal protective equipment is required per the patient's transmission-based isolation precautions
- Remove the device from the patient and placed on a clean disposable pad
- Discard the disposable face cover, if present
- Remove nonporous, nondisposable face cover from the device, if present
- Clean all visibly soiled areas with disposable wipes or paper towels
- Repeat disinfection as above
- Allow headset and controllers to dry according to the product instructions for use
- Store the device in a dry space physically separated from nondisinfected devices
- Patient and staff perform hand hygiene

In creating the SOP (Textbox 1), we considered VR devices a noncritical item requiring low-level disinfection between patients because they are most commonly exposed to intact skin only. Given disinfection success with the IPA and alcohol-free quaternary ammonium wipes, we suspect other equivalent low-level disinfectant products (eg, combination IPA or quaternary ammonium wipes) would be adequate, especially when applied to nonporous services [12]. We did not evaluate high-level disinfectants or sterilization procedures that may be required in the event of device exposure to nonintact skin or mucous membranes. We suggest avoiding VR device use on patients who have breaks in the skin on the hands or head region that cannot be appropriately covered and could come into contact with the device, thus avoiding the need for a high level of disinfection.

As disinfection was successfully achieved for a variety of pathogens, VR use is most likely safe for patients where contact isolation (gowns and gloves) is required in the hospital, including patients colonized with methicillin-resistant *S aureus* or vancomycin-resistant Enterococci. However, we suggest clinicians exercise caution when using VR devices with patients colonized with harder-to-eradicate pathogens such as *Clostridioides difficile*, *Candida auris*, and nonenveloped viruses where sodium hypochlorite or other high-level disinfection methods may be required. The pathogens tested (*S aureus*, *P aeruginosa*, and *S epidermidis*) are very common organisms in health care settings seen in both the adult and pediatric populations, and these results from disinfection are likely to be applicable to most health care settings, regardless of the patient's age.

Limitations of this study include the single-site nature limiting generalizability and the poor survey response rate. We also only tested 2 VR devices from the same company and 2 methods of surface disinfection. Preliminary experiments with an ultraviolet C (UVC) device specifically designed to decontaminate VR devices failed to produce adequate disinfection (data not shown). Since UVC disinfection depends on the angle and distance from the surface to the UVC source [13], the geometry of the headsets may make UVC disinfection more challenging. Further studies evaluating alternative disinfection methods, including UVC and other types of VR devices, are ongoing.

Acknowledgments

We acknowledge Huei-Ling Huang for her laboratory assistance with this project. This work was supported by internal funds from the Yale School of Medicine given to SJC, TSM, RAM, and SCR.

Data Availability

The data sets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

Study conception and design were performed by SCR, TSM, RAM, AMM, SJC, and MJW. Laboratory experiments were performed by SCR, NLH, RSF, CAH, and TSM. Statistical analysis was performed by RSF. All authors contributed to the writing and editing of the manuscript.

Conflicts of Interest

AMM has received funding from Meta for work unrelated to this project and was not used to fund this work. RAM has received funding from PDI for work unrelated to this project and was not used to fund this work. RAM owns shares of Meta.

References

1. Rothbaum BO. Using virtual reality to help our patients in the real world. Depress Anxiety 2009;26(3):209-211. [doi: 10.1002/da.20556] [Medline: 19253416]
2. Rothbaum BO, Price M, Jovanovic T, Norrholm SD, Gerardi M, Dunlop B, et al. A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder in Iraq and Afghanistan War veterans. Am J Psychiatry 2014 Jun;171(6):640-648 [FREE Full text] [doi: 10.1176/appi.ajp.2014.13121625] [Medline: 24743802]
3. Solcà M, Ronchi R, Bello-Ruiz J, Schmidlin T, Herbelin B, Luthi F, et al. Heartbeat-enhanced immersive virtual reality to treat complex regional pain syndrome. Neurology 2018 Jul 31;91(5):e479-e489. [doi: 10.1212/WNL.0000000000005905] [Medline: 29980635]
4. Schneider SM, Hood LE. Virtual reality: a distraction intervention for chemotherapy. Oncol Nurs Forum 2007 Jan;34(1):39-46 [FREE Full text] [doi: 10.1188/07.ONF.39-46] [Medline: 17562631]
5. Eijlers R, Utens EMWJ, Staals LM, de Nijs PFA, Berghmans JM, Wijnen RMH, et al. Systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. Anesth Analg 2019 Nov;129(5):1344-1353 [FREE Full text] [doi: 10.1213/ANE.0000000000004165] [Medline: 31136330]

6. Moore N, Dempsey K, Hockey P, Jain S, Poronnik P, Shaban RZ, et al. Innovation during a pandemic: developing a guideline for infection prevention and control to support education through virtual reality. Front Digit Health 2021;3:628452 [FREE Full text] [doi: 10.3389/fdgth.2021.628452] [Medline: 34713100]
7. Rutala WA, Weber DJ. Disinfection and sterilization in health care facilities: an overview and current issues. Infect Dis Clin North Am 2021 Sep;35(3):575-607. [doi: 10.1016/j.idc.2021.04.004] [Medline: 34362535]
8. Product care and best practices 2022. Meta. URL: https://store.facebook.com/help/quest/articles/headsets-and-accessories/product-care-and-best-practices/ [accessed 2022-11-10]
9. Meta's Oculus Quest 2 is the top selling VR headset of 2021; XR market growth skyrockets 2022. Tech Times. 2022 Mar 21. URL: https://www.techtimes.com/articles/273303/20220321/meta-s-oculus-quest-2-top-selling-vr-headset-2021.htm [accessed 2022-11-10]
10. Lopez GU, Gerba CP, Tamimi AH, Kitajima M, Maxwell SL, Rose JB. Transfer efficiency of bacteria and viruses from porous and nonporous fomites to fingers under different relative humidity conditions. Appl Environ Microbiol 2013 Sep;79(18):5728-5734 [FREE Full text] [doi: 10.1128/AEM.01030-13] [Medline: 23851098]
11. List H: EPA's registered antimicrobial products effective against methicillin-resistant Staphylococcus aureus (MRSA) and/or vancomycin resistant Enterococcus faecalis or faecium (VRE). EPA. URL: https://www.epa.gov/pesticide-registration/list-h-epas-registered-antimicrobial-products-effective-against-methicillin [accessed 2021-12-02]
12. Guideline for disinfection and sterilization in healthcare facilities, 2008. CDC. URL: https://www.cdc.gov/infectioncontrol/guidelines/disinfection/ [accessed 2022-11-10]
13. Lindblad M, Tano E, Lindahl C, Huss F. Ultraviolet-C decontamination of a hospital room: amount of UV light needed. Burns 2020 Jun;46(4):842-849 [FREE Full text] [doi: 10.1016/j.burns.2019.10.004] [Medline: 31676249]

Abbreviations

CFU: colony forming unit
D/E: Dey-Engley
EPA: Environmental Protection Agency
IPA: isopropyl alcohol
SOP: standard operating procedure
UVC: ultraviolet C
VR: virtual reality

Edited by G Eysenbach; submitted 31.08.22; peer-reviewed by N Moore, D Madhusudhan; comments to author 10.10.22; accepted 18.10.22; published 01.12.22.

Please cite as:
Roberts SC, Havill NL, Flores RM, Hendrix II CA, Williams MJ, Feinn RS, Choi SJ, Martinello RA, Marks AM, Murray TS
Disinfection of Virtual Reality Devices in Health Care Settings: In Vitro Assessment and Survey Study
J Med Internet Res 2022;24(12):e42332
URL: https://www.jmir.org/2022/12/e42332
doi:10.2196/42332
PMID:36269222

Original Paper

Measuring Digital Vaccine Literacy: Development and Psychometric Assessment of the Digital Vaccine Literacy Scale

Ilaria Montagni[1], PhD; Aude Pouymayou[1], MSc; Edwige Pereira[1], MSc; Christophe Tzourio[1], MD, PhD; Stéphane Schück[2,3], MD; Nathalie Texier[2], PharmD; Juan Luis González-Caballero[4], PhD; The CONFINS Group[5]

[1]Bordeaux Population Health UMRS1219, University of Bordeaux, Institut national de la santé et de la recherche médicale, Bordeaux, France

[2]Kappa Santé, Paris, France

[3]Kap Code, Paris, France

[4]Department of Statistics and Operational Research, University of Cádiz, Cádiz, Spain

[5]See Acknowledgments

Corresponding Author:
Ilaria Montagni, PhD
Bordeaux Population Health UMRS1219
University of Bordeaux
Institut national de la santé et de la recherche médicale
146 rue Léo Saignat
Bordeaux, 33000
France
Phone: 33 0642193363
Email: ilaria.montagni@u-bordeaux.fr

Abstract

Background: The use of the internet to look for information about vaccines has skyrocketed in the last years, especially with the COVID-19 pandemic. Digital vaccine literacy (DVL) refers to understanding, trust, appraisal, and application of vaccine-related information online.

Objective: This study aims to develop a tool measuring DVL and assess its psychometric properties.

Methods: A 7-item online questionnaire was administered to 848 French adults. Different psychometric analyses were performed, including descriptive statistics, exploratory factor analysis, confirmatory factor analysis, and convergent and discriminant validity.

Results: We developed the 7-item DVL scale composed of 3 factors (understanding and trust official information; understanding and trust information in social media; and appraisal of vaccine information online in terms of evaluation of the information and its application for decision making). The mean DVL score of the baseline sample of 848 participants was 19.5 (SD 2.8) with a range of 7-28. The median score was 20. Scores were significantly different by gender (P=.24), age (P=.03), studying or working in the field of health (P=.01), and receiving regular seasonal flu shots (P=.01).

Conclusions: The DVL tool showed good psychometric proprieties, resulting in a promising measure of DVL.

(J Med Internet Res 2022;24(12):e39220) doi:10.2196/39220

KEYWORDS

Internet; literacy; measurement; vaccination; vaccine; health information; health literacy; online; content; validity; reliability; digital literacy

Introduction

Vaccination is one of the most commonly queried topics on the internet [1]. With the COVID-19 pandemic, the number of people seeking vaccine-related information on the internet has skyrocketed [2,3]. The Increasing Vaccination Model [4] states that information sharing and rumors contribute, among other factors, to motivation to vaccinate. The 5C (complacency, constraints, calculation, confidence, collective responsibility) Model [5] asserts that vaccine hesitancy depends also on the engagement in extensive information seeking (ie, calculation), which determines deliberation on the risks and benefits of vaccination based on retrieved data and news. Thus, according to these 2 models, the contents of online information have the potential to determine the decision to get vaccinated or not.

Online sources for vaccine-related information vary. These include websites of official institutions, blogs, forums, social media, among others. The information they convey can be either reliable and valid or unscientific and misleading. On the one hand, social media have been defined as a powerful catalyst for the "anti-vax movement" [6]. This has been emphasized during the COVID-19 pandemic with a wide circulation of false information about vaccines on social media platforms [7,8]. On the other hand, websites of official institutions, such as those of governments, are considered to be more accurate [9]. Recent studies concerning the COVID-19 pandemic have confirmed that government websites are the most trusted source of information [10,11].

Hesitancy toward vaccination remains a present and growing issue [12]. Among the various reasons for this attitude, *misconception* and *misinformation* can have a strong impact [13]. Online messages can contribute to diffuse controversial information and induce indecision and skepticism about vaccines [14].

Preliminary studies have explored the influence of the internet on growing vaccine hesitancy [15,16]. According to these studies, those who search for online information more actively are usually also the most hesitant, trusting and believing science less than other sources [17]. Furthermore, the spread of fake news and misinformation on social media is blamed as a primary cause of vaccine hesitancy [18]. However, the internet is also a source of official reliable information and might provide new instruments to fight against vaccine hesitancy, because users can also access government websites, for instance.

Digital health literacy refers to the capacity of people to adequately understand and process online health information to meet their needs [19]. This set of skills affects the health of users, as well as the quality of their health care, orienting their health behavior. Vaccine literacy is defined as not only knowledge about vaccines, but also developing a simple system to communicate and offer vaccines as a sine qua non of a functioning health system [20,21]. Digital vaccine literacy (DVL) is a construct mixing digital health literacy and vaccine literacy. DVL theoretically affects both motivation and skills involving online information seeking for clear-cut elucidated decision making about getting vaccinated or not.

A valid tool for measurement of DVL is thus essential to provide inputs to train people in better navigating vaccine-related information on the internet on both social media and official online sources. This scale developed herein also allows to provide a general and population-based assessment of DVL: given the spread of the COVID-19 pandemic and the relevance of accepting vaccination, today more than ever it is pivotal to investigate the level of DVL in the population and examine its potential contribution to vaccine uptake. Furthermore, the scale can be used as an instrument to measure the effectiveness of interventions aimed at increasing DVL for reducing vaccine hesitancy.

To the best of our knowledge, no tool exists to measure DVL. The currently used questionnaires focus on vaccine literacy in general and not on online vaccine literacy (ie, DVL) [21,22]. The aim of this study was to describe the development and psychometric properties of a scale measuring DVL (Multimedia Appendix 1).

Methods

Overview of Study Phases

Our study was conducted in 3 distinct phases: (1) development of a tool to measure DVL, (2) collection of empiric cross-sectional data from a French adult population sample, and (3) assessment of the psychometric properties of the DVL tool.

We used the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) to develop the DVL tool and validate it [23].

Phase 1: DVL Tool Development

We based the conception of the DVL tool on the theories of digital health literacy and vaccine literacy, investigating the understanding, trust, appraisal, and application of vaccine-related information online [20,24], with the distinction between social media/forums and government websites. A panel of 5 public health researchers proposed a series of items inspired by the Health Literacy Questionnaire [25,26], the eHealth Literacy Scale [19], and the Vaccine Literacy Scale [22].

The construct of DVL was decided a priori and defined before any item activity. Expert judges confirmed through literature review that there were no existing instruments that will adequately serve the same purpose. A deductive method was used to identify the items through the description of the relevant field (domain), in combination with an inductive method based on the exchanges among experts. A group of 10 volunteers with characteristics similar to the target population pretested the questions. Items were worded in simple terms and unambiguously.

We narrowed the items focusing on vaccination and the digital environment to eventually obtain a total of 7 questions answered on a 4-point Likert scale (from 4 [agree] to 1 [disagree]) and an additional answer option "I do not know, I do not look for vaccine-related information." This latter option was taken into account in the descriptions, but was considered "noninformative" for the analysis of the structural validity of the scale. The total score of the DVL scale was calculated through the sum of all answers to the items. The score of the scale varied from 7 to 28. The higher the score, the better the DVL level.

We also included an item on "the online sources which were the most consulted for vaccine-related information seeking" (online journals, government websites, health institution websites, social media, forums, video platforms, other). Finally, participants had to rate the importance of the use of the internet for vaccine-related information seeking through a visual analog scale from 1 (not important at all) to 5 (very important).

Phase 2: Data Collection and Definition of the Population Under Study

We administered the DVL tool to participants from an open online cohort (CONFINS) [27]. All participants were aged more than 18 years, living in France, and were able to read and

understand French. CONFINS is a cohort collecting data on the impact of confinement on the health and well-being of the French population [28]. It included, among others, variables on opinions about vaccination and the DVL items. It also comprised sociodemographic information (age, gender, having children, being vaccinated against influenza) used in this study. Items were defined by a group of public health experts through several rounds of corrections and refinement. CONFINS consisted in a baseline questionnaire and repeated monthly follow-up questionnaires. Participants could decide whether to be contacted or not for the following phases of the survey. This study used data from the baseline questionnaire and the first follow-up questionnaire, covering the period from April to May 2020. This was a convenience sample.

CONFINS participants were recruited on a voluntary basis with no incentives through different communication channels. Posts were published on the social media (LinkedIn, Twitter, Facebook) of the University of Bordeaux and the partner contract research organization hosting the database. A total of 3 press releases were addressed to journalists. The coprinciple investigators were interviewed to promote the study. Three newsletters and weekly emails and SMS text messages were sent to the participants to remind them to complete the follow-up questionnaires. All recruitment strategies directed potential participants toward the CONFINS website including information on the objectives of the study and the investigators. Informed consent, containing details on the length of time of the survey, stored data, investigators and objectives of the study, was provided through an electronic signature.

Study Population

Concerning the population of this study, we included all participants completing all items of the DVL tool, comprising also those choosing the answer option "I do not know, I do not look for vaccine-related information" (N=2935). However, for the sake of the specific analyses required to evaluate the psychometric properties of the DVL tool, we obtained a subsample of 848 participants who did not use the answer option "I do not know, I do not look for vaccine-related information." The choice of using mainly the subsample was justified by the fact that the factor analysis mentioned later requires ordering the response modalities. As the "I do not know, I do not look for vaccine-related information" modality is difficult to classify, we decided to remove it. The subsample included those who had completed the baseline questionnaire ("test" phase). Among them, 62 participants also answered the follow-up questionnaire ("retest" phase).

Phase 3: Analysis of Other Psychometric Properties of the DVL Tool

First, a descriptive analysis of each item of the scale was performed for both the total sample of participants (N=2935) and the subsample (n=848). Participants of the subsample were also described according to their sociodemographic characteristics (ie, age, gender, working/studying in the field of health, having children, and being regularly vaccinated against flu). For quantitative variables, the mean and SD were calculated. For qualitative variables, participants were described in numbers and percentages. Answers to items were compared for each aforementioned sociodemographic characteristic. To do this, the item response options were grouped into "agree"/"rather agree" versus "disagree"/"rather disagree." The statistical tests of χ^2 independence were used to compare the responses of the participants according to their sociodemographic criteria.

Second, an exploratory factor analysis (EFA) was performed on the baseline data to identify the underlying latent factors in the set of items as well as their association. As the items were ordinal variables, the polychoric correlation matrix of observed items was explored. Two initial hypotheses were tested. The first was the test of Bartlett sphericity. If the test was significant (P<.05), the observed matrix was significantly divergent from the null matrix and an EFA had to be performed. The second hypothesis required testing the measure of sampling adequacy using the Kaiser-Meyer-Olkin index [29]. This is a measure of the proportion of variance among the observed items, equivalent to the common variance. Thus, it was used to verify for partial correlations. If the Kaiser-Meyer-Olkin index was above 0.50, the EFA was adequate. Next, the number of factors to be kept in the model had to be chosen based on different criteria using eigenvalues. The Kaiser criterion consisted of keeping factors with eigenvalues greater than 1. The Cattell criterion (also called the "elbow criterion") was based on identifying the inflection point, where the slope of the eigenvalue curve according to the number of factors in the model stabilized well below the "elbow." Thus, the number of factors above the point was retained. The third criterion was the use of a parallel analysis. In this analysis, the eigenvalues obtained were compared with those that would be obtained from random data. The number of factors extracted was the number of factors whose eigenvalues were higher than those found with random data. In addition, the item × factor matrix had to be rotated to better identify how the items were substantially related to each factor. Among the several approaches to rotation, the oblique rotation was used because it considers the correlation between factors [30]. Finally, the items were associated with a factor when their saturation weight was close or superior to 0.30 and their communalities were considered as acceptable above 0.20. We also performed a confirmatory factor analysis (CFA) considering the criteria root-mean-square error of approximation (acceptable range between 0.08 and 0.1), comparative fit index (acceptable range >0.90) and standardized root-mean-square error (acceptable range between 0 and 0.008).

Third, to complete the validation of the DVL scale, the convergent and discriminant validities of the score were assessed. The sociodemographic criteria of participants with a low DVL score were compared with those of participants with a high score, determined according to the median, using χ^2 statistical tests of independence.

Statistical significance was considered if P<.05 and all tests were 2-tailed. Statistical analyses were performed on SAS version 9.3 software (SAS Institute).

Ethics Approval

The study was approved by the French Committee for the Protection of Individuals (Comité de Protection des Personnes

[CPP], approval number 46-2020) and the French National Agency for Data Protection (Commission Nationale de l'Informatique et des Libertés [CNIL], approval number MLD/MFI/AR205600). The study follows the principles of the Declaration of Helsinki and the collection, storage, and analysis of the data comply with the European Union General Data Protection Regulation (EU GDPR).

Results

Descriptive Analysis

Responses to the 7 items on the DVL tool by the total sample and the subsample are reported in Tables 1 and 2, respectively.

Table 1. Results of all potentials items of the DVL scale[a] in the CONFINS online cohort (N=2935).

Items	Disagree, n (%)	Rather disagree, n (%)	Rather agree, n (%)	Agree, n (%)	Do not know, n (%)
1. I find vaccine-related information on social media and forums is understandable	215 (7.33)	478 (16.29)	582 (19.83)	134 (4.57)	1526 (51.99)
2. I find vaccine-related information on government websites is understandable	111 (3.78)	176 (6)	1394 (47.50)	586 (19.97)	668 (22.76)
3. I can detect vaccine-related fake news	97 (3.30)	477 (16.25)	1500 (51.11)	821 (27.97)	40 (1.36)
4. I trust vaccine-related information provided by government websites	55 (1.87)	191 (6.51)	1250 (42.59)	948 (32.30)	491 (16.73)
5. I find vaccine-related information on social networks is valid	533 (18.16)	1123 (38.26)	134 (4.53)	26 (0.89)	1119 (38.13)
6. When I read vaccination information online, I cross-reference it with other sources to verify its validity	178 (6.06)	394 (13.42)	1288 (43.88)	1060 (36.12)	15 (0.51)
7. I think the information I find online may influence my decision to get vaccinated	413 (14.07)	649 (22.11)	918 (31.28)	231 (7.97)	724 (24.67)

[a]DVL scale: Digital Vaccine Literacy scale.

Table 2. Results of all potential items of the DVL scale[a] in the CONFINS online cohort (n=848, without “do not know”).

Item	Disagree, n (%)	Rather disagree, n (%)	Rather agree, n (%)	Agree, n (%)	Test-retest reliability (n=62), intraclass correlation coefficient (95% CI)
1. I find vaccine-related information on social media and forums is understandable	139 (16.4)	287 (33.8)	342 (40.3)	80 (9.4)	0.14 (0.01 to 0.37)
2. I find vaccine-related information on government websites is understandable	49 (5.8)	82 (9.7)	492 (58.0)	225 (26.5)	0.53 (0.33 to 0.69)
3. I can detect vaccine-related *fake news*	27 (3.2)	111 (13.1)	421 (49.6)	289 (34.1)	0.70 (0.55 to 0.81)
4. I trust vaccine-related information provided by government websites	23 (2.7)	82 (9.7)	409 (48.2)	334 (39.4)	0.46 (0.24 to 0.63)
5. I find vaccine-related information on social networks is valid	224 (26.4)	529 (62.4)	83 (9.8)	12 (1.4)	0.05 (0.01 to 0.29)
6. When I read vaccination information online, I cross-reference it with other sources to verify its validity	44 (5.2)	87 (10.3)	365 (43)	352 (41.5)	0.48 (0.27 to 0.65)
7. I think the information I find online may influence my decision to get vaccinated	122 (14.4)	267 (31.5)	354 (41.7)	105 (12.4)	–0.09 (–0.33 to 0.16)

[a]DVL scale: Digital Vaccine Literacy scale.

The “I do not know, I do not look for vaccine-related information” response rates were 51.99% (1526/2935) for item 1, 22.76% (668/2935) for item 2, 1.36% (40/2935) for item 3, 16.73% (491/2935) for item 4, 38.13% (1119/2935) for item 5, 5.04% (148/2935) for item 6, and 24.67% (724/2935) for item 7. Per participant, the maximum number of “I do not know, I do not look for vaccine-related information” was 5; 24.74% (726/2935) responded “I do not know, I do not look for vaccine-related information” for at least one item; 23.51% (690/2395) for at least two items; 10.97% (322/2935) for at least three items; 7.97% (234/2935) for at least four items; and 3.92% (115/2395) for at least five items. The mean of responses per participant was 1.56 (SD 1.4). In addition, the use of a factor analysis requires ordering the response modalities. As the “I do not know, I do not look for vaccine-related information” modality is difficult to classify in view of the others, we decided to remove it from the analyses. Therefore, the study sample

contained 848 participants who responded to the items as shown in Table 2.

All item response options were used, thus qualifying them as informative. In addition, Table 2 shows that the items were discriminating because the response rates for each modality were in the average. The intraclass correlation coefficient (ICC) was calculated based on data from the 62 participants. Items 1, 5, and 7 presented a low ICC, which could be explained by nonconcordant responses between the 2 measurements, and therefore less reliability, their formulation, and possible difficulty in answering them. In fact, these items had the highest percentages of the "I do not know, I do not look for vaccine-related information" responses (Table 1).

In the subsample of 848 participants, 73.1% (620/848) were females. The mean age was 29.9 (SD 12.3). Participants working or studying in the field of health were 397/848 (46.8%). The percentage of parents was 20.9% (178/848) and 557/848 (65.7%) were not vaccinated against flu (Table 3).

The mean of the importance of the use of the internet for vaccine-related information seeking was 3.7 out of 5 (SD 1.1). The most used source for vaccine-related information seeking was websites of health institutions (395/848, 46.6%), followed by government websites (184/848, 21.7%). Online journals were consulted by 56/848 individuals (6.6%), whereas other sources by 37/848 individuals (4.4%). Social networks were consulted by 70/848 individuals (8.3%), video platforms by 16/848 (1.9%), and forums by 8/848 (0.9%).

Multimedia Appendix 2 reports data on the comparison of the answer to the DVL items according to sociodemographic characteristics.

Regarding their answers to the items, women were more in agreement with the statement of item 3 (I can detect vaccine-related fake news), item 4 (I trust vaccine-related information provided by government websites), and item 7 (I think the information I find online may influence my decision to get vaccinated) than men. Participants aged 35 or over disagreed with item 1 (I find vaccine-related information on social media and forums is understandable), which was different from those under 35 years. Participants studying or working in the field of health and those receiving regular flu shots were more in agreement with items 2 (I find vaccine-related information on government websites is understandable), item 3 (I can detect vaccine-related fake news), and item 4 (I trust vaccine-related information provided by government websites) and disagreed with item 7 (I think the information I find online may influence my decision to get vaccinated) compared with those who worked or studied in another field and those who did not get a flu shot. There was no difference in responses concerning parenthood.

Table 3. Sociodemographic characteristics of the CONFINS study population.

Characteristics	Value
Age, mean (SD)	29.9 (12.3)
Categories (n=835), years , n (%)	
18-34	653 (78.2)
≥35	182 (21.8)
Gender (n=848), n (%)	
Female	620 (73.1)
Male	228 (26.9)
Study or work in the field of health (n=763), n (%)	
No	366 (48.0)
Yes	397 (52.0)
Children (n=848), n (%)	
No	670 (79.0)
Yes	178 (21.0)
Influenza vaccine (n=848), n (%)	
No	557 (65.7)
Yes	291 (34.3)

Exploratory Factor Analysis

The interitem polychoric correlation matrix was used for the first definition of the associations between items (Table 4).

In the polychoric matrix, we observed strong correlations between items 2, 3, and 4. Item 1 was more correlated with item 5.

The hypotheses justifying the performance of an EFA were validated. The Bartlett test of sphericity showed a $P<.05$ ($\chi^2_{21}=1319.37$) and the Kaiser-Meyer-Olkin index was 0.58, indicating good sampling adequacy.

The number of factors was calculated based on the Kaiser and Cattell criteria and the parallel analysis; 3 factors were kept (Figure 1).

Finally, several EFAs were performed to test the different oblique rotations. The OBLIMIN oblique rotation was the most common. Table 5 shows that items 1 and 5 were associated with factor 2; items 2, 3, and 4 with factor 1; and items 6 and 7 with factor 3. The oblique rotation OBEAQUAMAX showed that saturation weights revealed several possible associations between items and factors. Items 3 and 7 were associated with both factors 1 and 3 based on the saturation weights close or superior to 0.30. Communalities were all acceptable.

Table 4. Interitem polychoric correlation matrix.

Item	1	2	3	4	5	6	7
1	__[a]	—	—	—	—	—	—
2	0.33	—	—	—	—	—	—
3	0.00	0.46	—	—	—	—	—
4	0.06	0.64	0.52	—	—	—	—
5	0.45	–0.02	–0.10	–0.06	—	—	—
6	0.06	0.19	0.34	0.12	–0.02	—	—
7	0.13	–0.11	–0.13	–0.15	0.21	0.20	—

[a]Dashes correspond to the absence of a correlation between items.

Figure 1. Distribution of the median simulated eigenvalues according to the number of factors and application of the parallel analysis. 7 variables, iterations, 848 observations.

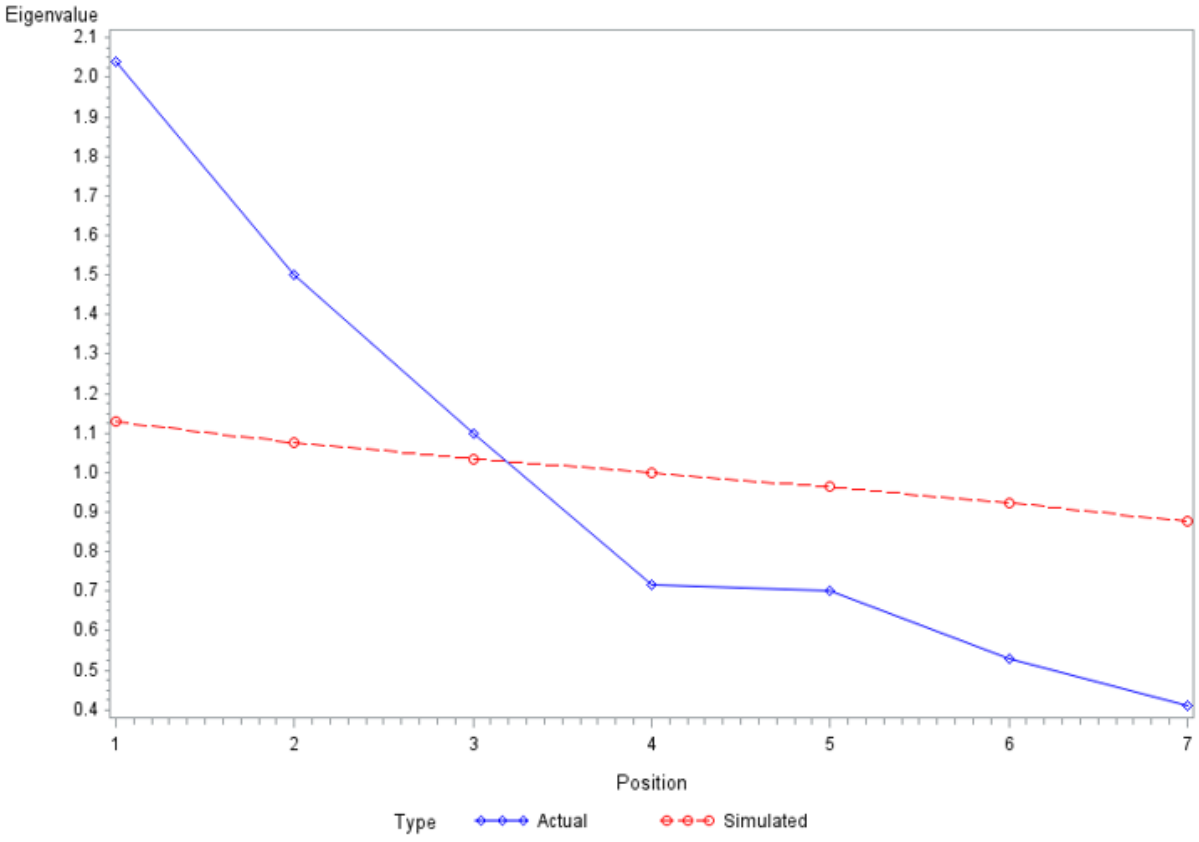


Table 5. Matrices of the saturation weights with oblique rotations and item communalities.

Item	OBLIMIN			OBEAQUAMAX			Communality
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	
1	0.19	0.69	–0.02	0.19	0.67	0.01	0.46
2	0.78	0.23	–0.01	0.74	0.21	0.13	0.63
3	0.60	–0.14	0.25	0.50	–0.15	0.37	0.47
4	0.76	0.01	–0.03	0.72	–0.01	0.12	0.57
5	–0.08	0.56	0.03	–0.07	0.57	–0.01	0.34
6	0.17	–0.05	0.49	0.03	–0.04	0.53	0.28
7	–0.23	0.20	0.33	-0.30	0.21	0.29	0.21

Table 6 shows the interfactor correlations according to the OBLIMIN and OBEAQUAMAX rotations. Correlations were low but factor 1 was negatively correlated with factor 2, and factor 3 was positively correlated with the other 2 factors.

In view of these results, the relationships between the items and the factors were interpreted as follows. Factor 1 was associated with items relating to “reliable” information about vaccination (government sites), with the label “understanding and trust official information about vaccination provided by institutional websites.” Factor 2 was associated with items related to information about vaccination of which 1 should be relatively “unreliable” (social media) with the label “understanding and trust information about vaccines as provided by social media.” Finally, factor 3 was associated with items related to the application of knowledge on vaccination consulted on the web (label of factor 3).

Finally, we also performed a CFA to confirm these 3 dimensions (Table 7).

In the CFA the criterion values were as follows: root-mean-square error of approximation 0.12 (90% CI 0.11-1.14), comparative fit index 0.80, and standardized root-mean-square error 0.08.

Table 6. Interfactor correlation matrices (OBLIMIN and OBEAQUAMAX).

Factor	OBLIMIN			OBEAQUAMAX		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
1	1	—[a]	—	1	—	—
2	–0.08	1	—	–0.09	1	—
3	0.11	0.18	1	0.19	0.16	1

[a]Dashes correspond to the absence of a correlation between items and factors.

Table 7. Weights of the relationships item-factors of the final model by confirmatory factor analysis.

Item	Model 1		
	Factor 1	Factor 2	Factor 3
1	—[a]	0.87	—
2	0.56	—	—
3	0.43	—	—
4	0.51	—	—
5	—	0.23	—
6	—	—	0.83
7	—	—	0.15

[a]Dashes correspond to the absence of a correlation between items and factors.

Convergent and Discriminant Validity

The mean DVL score of the baseline sample of 848 participants was 19.5 (SD 2.8). Participants scored between 14 and 21 points (ie, in the medium DVL range). The median was 20.

Table 8 shows the sociodemographic characteristics of the sample according to the DVL level. The score was dichotomized into <20 (low DVL score) and ≥20 (high DVL score).

Participants with a low DVL level were significantly older (30.8 years vs 29 years; *P*=.03). Those working or studying in the field of health were significantly more numerous in the group

with a higher score (*P*=.01). Those who did not receive regular flu vaccinations were significantly more likely to be in the low score group (*P*=.01). Among online sources for vaccine-related information, government websites were more used by those with a higher DVL (*P*=.03). Those with a score less than 20 considered the use of the internet for vaccine-related information less important than others, with the means being 3.4 (SD 1.1) and 4.0 (0.9), respectively.

Table 8. Sociodemographic characteristics of the baseline sample by DVL[a] level (n=848).[b]

Sociodemographics	Low DVL (score <20)	High DVL (score ≥20)	*P* value
Age (years), mean (SD)	30.8 (12.9)	29.0 (11.7)	.03
Age c ategories (n=397)			.04
18-34	298/397 (75.1)	355/438 (81.1)	
≥35	99/397 (24.9)	83/438 (18.9)	
Gender (n=404)			.24
Female	303/404 (75)	317/444 (71.4)	
Male	101/404 (25)	127/444 (28.6)	
Studying or working in the field of health (n=357)			.01
No	192/357 (53.8)	174/406 (42.9)	
Yes	165/357 (46.2)	232/406 (57.1)	
Having children (n=404)			.38
No	314/404 (77.7)	356/444 (80.2)	
Yes	90/404 (22.3)	88/444 (19.8)	
Vaccinated against flu (n=404)			.01
No	283/404 (70)	274/444 (61.7)	
Yes	121/404 (30)	170/444 (38.3)	
Online sources for vaccine-related information (n=338)			.03
Online journals	30/338 (8.9)	26/390 (6.7)	
Government websites	73/338 (21.6)	111/390 (28.5)	
Health institutions websites	185/338 (54.7)	210/390 (53.8)	
Social media	19/338 (5.6)	13/390 (3.3)	
Forums	7/338 (2.1)	1/390 (0.3)	
Video Platforms	5/338 (1.5)	11/390 (2.8)	
Other	19/338 (5.6)	18/390 (4.6)	
Importance of the use of the internet for vaccine-related information seeking (n=338), mean (SD)	3.4 (1.1)[c]	4.0 (0.9)[d]	<.001

[a]DVL: digital vaccine literacy.

[b]Values are presented as n/N (%) unless indicated otherwise.

[c]N=338.

[d]N=390.

Discussion

The DVL Scale: Dimensions, Items, and Answer Options

We conceived a scale measuring DVL and assessed its psychometric proprieties among a sample of French adults. The scale was composed of 7 items covering the overarching construct of DVL, which includes 3 subdimensions. The first subdimension (items 2 and 4) refers to understanding and trusting official information about vaccination provided by institutional websites. The second subdimension (items 1 and 5) refers to understanding and trusting information about vaccines as provided by social media. The underlying assumption for these 2 dimensions is that government websites provide valid information while social media provide fake news [31]. In this line, in our sample, the most accessed sources were health institutions and government websites, while social media and forums were less consulted.

The third subdimension (items 3, 6, and 7) refers to the appraisal of vaccine information online in terms of evaluation of the information and its application for decision making. Two items (3 and 7) are actually included in both subdimensions 1 and 2.

For the item "I can detect fake news," this ambivalence can be explained by the fact that recognizing fake news is a reflection of both the understanding/trust of official information (subdimension 1) and the appraisal and practical application of found information (subdimension 3). The possible explanation is that those who recognize fake news are more inclined to government websites and are more cautious in interpreting vaccine-related information. The inclusion of the item "I think the information I find online may influence my decision to get vaccinated" in both subdimensions 1 and 3 can be interpreted as the fact that trusting official information might correspond to a higher capacity to make correct evidence-based decisions about vaccination. This overlap of factors infers an interrelation of items, which can suggest that the scale is coherent and congruent.

Some recommendations must be considered when using the DVL scale. There are 4 response options (*disagree*, *rather disagree*, *rather agree*, and *agree*) that are used to obtain a score. However, even if it does not contribute to the calculation of the score, the fifth response option (I do not know, I do not look for vaccine-related information) provides useful information. First, this option respects the opinion of those not feeling concerned without forcing or biasing their answer. Second, it is really interesting to measure the percentage of those who do not feel concerned by seeking vaccine-related information online. In this study, one-half of the participants used the option "I do not know, I do not look for vaccine-related information" for the item on understanding information found on social media, and more than one-third for the item on trust in social media. These results confirm the fact that social media are more rarely used than government websites for this type of information. Thus, we suggest to calculate the score by considering as missing values all cases including 1 response option "I do not know, I do not look for vaccine-related information", and to complete this information with the percentage of those using this same option. These data are complementary in measuring DVL.

The DVL Scores of the Study Sample

Having a low DVL score (<20) can be interpreted as a relevant alarm in relation to the extensive use of the internet for vaccine-related contents, especially in France [15]. As is the case with health literacy, low DVL scores are associated with a higher risk of adopting an unhealthy behavior [32]; in this case this refers to the decision of *not to get vaccinated*. Not being able to navigate information on the internet could increase the chance of having a negative perception about vaccines [33]. Lower scores in the scale would also correspond to the incapacity to recognize fake news and trust in unofficial information provided by social media. There are many who consult the internet regarding vaccination and it is important to know their levels of DVL to help them navigate online information.

DVL scores were significantly different by age (participants with a low DVL score were significantly older), studying or working in the field of health (those working or studying in the field of health were significantly more numerous in the group with a high score), and being vaccinated against flu (those who did not regularly get vaccinated against influenza were significantly more numerous in the group with a low score). These results are in line with previous literature concerning general health literacy: scores of health literacy are higher in younger adults [34], health care professionals [35], and those vaccinated against flu [36].

Comparison with results from other studies is not possible because DVL has never been measured before.

Strengths and Limitations

This study is the very first to develop and validate a standardized instrument for assessing general DVL in people. It responds to the urgent need for similar scales to tackle vaccine-related misinformation [37], especially in relation to the COVID-19 pandemic. Measuring the DVL of individuals consulting the internet for information on COVID-19–related vaccination could inform health institutions, communication experts, and health care providers to plan and implement strategies to overcome gaps in DVL and promote vaccination [38]. Furthermore, analyses performed in this study are robust and based on an in-depth knowledge of psychometrics techniques. In particular, the use of the bifactorial model is justified by the fact that it considers correlations between items based on the general factor and the relations between the general factor. Items are not limited by the group factors. This model is largely applied in cognitive and psychological sciences [39].

This study is not without limitations. Items were defined a priori based on existing scales but limited to 7. A larger number of items might have provided a more exhaustive coverage of DVL factors. The population under study was not representative of French adults given that it comprised a high number of women (2971/3738, 79.48%), students (3498/3783, 93.58%), and young people (29.2 years) [40], compared with the general population [41]. However, the sample was large enough to assess the relevance of the scale. Low ICC values in some separated items might be explained by an inaccurate phrasing. The ICCs of 3 items were low, which corresponds to a low reliability. Future instruments might be based on our scale, but we propose more precise wording according to the population of interest in a specific context (eg, cultural or sociodemographic characteristics).

Conclusions

The DVL scale is the first instrument providing information on the way individuals understand, trust, and appraise vaccine-related information on the internet through 2 channels, namely, social media and government websites. The DVL scale has good psychometric properties in terms of content validity, dimensionality, and convergent and discriminant validity. Results show that the scale can be easily administered with well-grounded outcomes. It is a screening instrument contributing to detect people who need to be supported in navigating vaccine-related information online. It can be used in questionnaires to identify profiles of web users who could be influenced by anti-vax movements, for instance. Providing the instructions to look for online information and to understand its content is the key to spreading good vaccine-related information and promoting vaccination in general [42]. The

scale can be used to measure DVL in the French population and translated validated versions could be proposed internationally.

Acknowledgments

We wish to thank all members of the CONFINS group including the i-Share, Kappa Santé, and Kap Code team members: we especially acknowledge Garance Perret and Mathilde Pouriel for data analysis; Julie Arsandaux, Shérazade Kinouani, and Mélissa Macalli for paper writing; Raphaël Germain and Clothilde Pollet for regulatory affairs; and Vanessa Marie-Joseph, Adel Mebarki, Elena Milesi, and Marie Mougin for the study communication. Kevin Ouazzani Touhami is also gratefully acknowledged. The authors are also grateful to all the participants who volunteered to take part in the study. The i-Share team is currently supported by an unrestricted grant of the Nouvelle-Aquitaine Regional Council (Conseil Régional Nouvelle-Aquitaine, grant N°4370420). It has also received grants from the Nouvelle-Aquitaine Regional Health Agency (Agence Régionale de Santé Nouvelle-Aquitaine, grant N°6066R-8), Public Health France (Santé Publique France, grant N°19DPPP023-0), and The National Institute against cancer INCa (grant N°INCa_11502). The article fees were covered by the Plan Propio - UCA 2022-2023, and the RÉFLIS network. The funding bodies were neither involved in the study design, or in the collection, analysis, or interpretation of the data.

Data Availability Statement

All data generated or analyzed during this study are included in this published article. The full data set is available upon request from the CONFINS cohort team.

Authors' Contributions

IM conceived the study and wrote and revised the manuscript. JLGC conceived the study, supervised analyses, and revised the manuscript. EP and AP analyzed the data. SS, NT, and CT conceived and designed the study cohort. Also see the “Acknowledgments” section.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original items of the DVL scale (French). DVL scale: Digital Vaccine Literacy scale.
[DOCX File , 15 KB - jmir_v24i12e39220_app1.docx]

Multimedia Appendix 2
Comparison of responses to the 7 DVL items according to sociodemographic characteristics (n=848). DVL: digital vaccine literacy.
[DOCX File , 21 KB - jmir_v24i12e39220_app2.docx]

References

1. Jones AM, Omer SB, Bednarczyk RA, Halsey NA, Moulton LH, Salmon DA. Parents' source of vaccine information and impact on vaccine attitudes, beliefs, and nonmedical exemptions. Adv Prev Med 2012;2012:932741-932748 [FREE Full text] [doi: 10.1155/2012/932741] [Medline: 23082253]
2. Goel RK, Nelson MA. COVID-19 internet vaccination information and vaccine administration: evidence from the United States. J Econ Finan 2021 Jun 03;45(4):716-734. [doi: 10.1007/s12197-021-09551-x]
3. Piltch-Loeb R, Savoia E, Goldberg B, Hughes B, Verhey T, Kayyem J, et al. Examining the effect of information channel on COVID-19 vaccine acceptance. PLoS One 2021 May 12;16(5):e0251095 [FREE Full text] [doi: 10.1371/journal.pone.0251095] [Medline: 33979370]
4. Brewer NT, Chapman GB, Rothman AJ, Leask J, Kempe A. Increasing Vaccination: Putting Psychological Science Into Action. Psychol Sci Public Interest 2017 Dec 03;18(3):149-207. [doi: 10.1177/1529100618760521] [Medline: 29611455]
5. Betsch C, Schmid P, Heinemeier D, Korn L, Holtmann C, Böhm R. Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination. PLoS One 2018 Dec 7;13(12):e0208601 [FREE Full text] [doi: 10.1371/journal.pone.0208601] [Medline: 30532274]
6. 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]
7. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar 05;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]

8. Burki T. The online anti-vaccine movement in the age of COVID-19. The Lancet Digital Health 2020 Oct;2(10):e504-e505. [doi: 10.1016/s2589-7500(20)30227-2]
9. Shehata A. Health Information behaviour during COVID-19 outbreak among Egyptian library and information science undergraduate students. Information Development 2020 Dec 07;37(3):417-430. [doi: 10.1177/0266666920976181]
10. Ali SH, Foreman J, Tozan Y, Capasso A, Jones AM, DiClemente RJ. Trends and Predictors of COVID-19 Information Sources and Their Relationship With Knowledge and Beliefs Related to the Pandemic: Nationwide Cross-Sectional Study. JMIR Public Health Surveill 2020 Oct 08;6(4):e21071 [FREE Full text] [doi: 10.2196/21071] [Medline: 32936775]
11. Falcone R, Sapienza A. How COVID-19 Changed the Information Needs of Italian Citizens. Int J Environ Res Public Health 2020 Sep 24;17(19):6988 [FREE Full text] [doi: 10.3390/ijerph17196988] [Medline: 32987914]
12. Machingaidze S, Wiysonge CS. Understanding COVID-19 vaccine hesitancy. Nat Med 2021 Aug 16;27(8):1338-1339. [doi: 10.1038/s41591-021-01459-7] [Medline: 34272500]
13. 13 B. Exploring online anti-vaccination movements: the role of social media in public health communications and reporting. European Journal of Public Health 2019 Nov;29(4):ckz185.683. [doi: 10.1093/eurpub/ckz185.683]
14. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]
15. Stahl J, Cohen R, Denis F, Gaudelus J, Martinot A, Lery T, et al. The impact of the web and social networks on vaccination. New challenges and opportunities offered to fight against vaccine hesitancy. Med Mal Infect 2016 May;46(3):117-122. [doi: 10.1016/j.medmal.2016.02.002] [Medline: 26987960]
16. Vrdelja M, Kraigher A, Vercic D, Kropivnik S. The growing vaccine hesitancy: exploring the influence of the internet. Eur J Public Health 2018 Oct 01;28(5):934-939. [doi: 10.1093/eurpub/cky114] [Medline: 29982349]
17. Germani F, Biller-Andorno N. The anti-vaccination infodemic on social media: A behavioral analysis. PLoS One 2021 Mar 3;16(3):e0247642 [FREE Full text] [doi: 10.1371/journal.pone.0247642] [Medline: 33657152]
18. Carrieri V, Madio L, Principe F. Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. Health Econ 2019 Nov 20;28(11):1377-1382 [FREE Full text] [doi: 10.1002/hec.3937] [Medline: 31429153]
19. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
20. Ratzan SC. Vaccine literacy: a new shot for advancing health. J Health Commun 2011 Mar 28;16(3):227-229. [doi: 10.1080/10810730.2011.561726] [Medline: 21391044]
21. Gusar I, Konjevoda S, Babić G, Hnatešen D, Čebohin M, Orlandini R, et al. Pre-Vaccination COVID-19 Vaccine Literacy in a Croatian Adult Population: A Cross-Sectional Study. Int J Environ Res Public Health 2021 Jul 02;18(13):7073 [FREE Full text] [doi: 10.3390/ijerph18137073] [Medline: 34281009]
22. Biasio LR, Giambi C, Fadda G, Lorini C, Bonaccorsi G, D'Ancona F. Validation of an Italian tool to assess vaccine literacy in adulthood vaccination: a pilot study. Ann Ig 2020;32(3):205-222 [FREE Full text] [doi: 10.7416/ai.2020.2344] [Medline: 32266359]
23. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745. [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]
24. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
25. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13(1):658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
26. Debussche X, Caroupin-Soupoutevin J, Balcou-Debussche M, Fassier M, Boegner C, Hawkins M, et al. Health literacy needs among migrant populations in France: validity testing and potential contribution of the Health Literacy Questionnaire (HLQ). J Public Health (Berl.) 2021 Jan 08;30(10):2301-2309. [doi: 10.1007/s10389-020-01423-8]
27. CONFINS group. CONFINS. 2022. URL: https://www.confins.org/ [accessed 2022-11-29]
28. Arsandaux J, Montagni I, Macalli M, Texier N, Pouriel M, Germain R, et al. Mental health condition of college students compared to non-students during COVID-19 lockdown: the CONFINS study. BMJ Open 2021 Aug 19;11(8):e053231 [FREE Full text] [doi: 10.1136/bmjopen-2021-053231] [Medline: 34413111]
29. Rosenblad A. The Concise Encyclopedia of Statistics. Journal of Applied Statistics 2011 Apr;38(4):867-868. [doi: 10.1080/02664760903075614]
30. Jennrich RI. A simple general method for oblique rotation. Psychometrika 2002 Mar;67(1):7-19. [doi: 10.1007/bf02294706]
31. Montagni I, Ouazzani-Touhami K, Mebarki A, Texier N, Schück S, Tzourio C, CONFINS group. Acceptance of a Covid-19 vaccine is associated with ability to detect fake news and health literacy. J Public Health (Oxf) 2021 Dec 10;43(4):695-702 [FREE Full text] [doi: 10.1093/pubmed/fdab028] [Medline: 33693905]
32. Klinker CD, Aaby A, Ringgaard LW, Hjort AV, Hawkins M, Maindal HT. Health Literacy is Associated with Health Behaviors in Students from Vocational Education and Training Schools: A Danish Population-Based Survey. Int J Environ Res Public Health 2020 Jan 20;17(2):671 [FREE Full text] [doi: 10.3390/ijerph17020671] [Medline: 31968667]

33. Gendler Y, Ofri L. Investigating the Influence of Vaccine Literacy, Vaccine Perception and Vaccine Hesitancy on Israeli Parents' Acceptance of the COVID-19 Vaccine for Their Children: A Cross-Sectional Study. Vaccines (Basel) 2021 Nov 24;9(12):1391 [FREE Full text] [doi: 10.3390/vaccines9121391] [Medline: 34960137]
34. Verney SP, Gibbons LE, Dmitrieva NO, Kueider AM, Williams MW, Meyer OL, et al. Health literacy, sociodemographic factors, and cognitive training in the active study of older adults. Int J Geriatr Psychiatry 2019 Apr 14;34(4):563-570 [FREE Full text] [doi: 10.1002/gps.5051] [Medline: 30548889]
35. Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health Informatics J 2020 Mar 15;26(1):592-612 [FREE Full text] [doi: 10.1177/1460458219839613] [Medline: 30983476]
36. Zhang F, Or PP, Chung JW. The effects of health literacy in influenza vaccination competencies among community-dwelling older adults in Hong Kong. BMC Geriatr 2020 Mar 14;20(1):103 [FREE Full text] [doi: 10.1186/s12877-020-1504-5] [Medline: 32171262]
37. van der Linden S. Misinformation: susceptibility, spread, and interventions to immunize the public. Nat Med 2022 Mar 10;28(3):460-467. [doi: 10.1038/s41591-022-01713-6] [Medline: 35273402]
38. Biasio LR. Vaccine literacy is undervalued. Hum Vaccin Immunother 2019 May 21;15(11):2552-2553 [FREE Full text] [doi: 10.1080/21645515.2019.1609850] [Medline: 31013184]
39. Eid M, Krumm S, Koch T, Schulze J. Bifactor Models for Predicting Criteria by General and Specific Factors: Problems of Nonidentifiability and Alternative Solutions. J Intell 2018 Sep 07;6(3):42 [FREE Full text] [doi: 10.3390/jintelligence6030042] [Medline: 31162469]
40. Macalli M, Texier N, Schück S, Côté SM, Tzourio C. A repeated cross-sectional analysis assessing mental health conditions of adults as per student status during key periods of the COVID-19 epidemic in France. Sci Rep 2021 Nov 09;11(1):21455 [FREE Full text] [doi: 10.1038/s41598-021-00471-8] [Medline: 34753945]
41. Insee. Évolution et structure de la population en 2019. Insee. 2022. URL: https://www.insee.fr/fr/statistiques/6455209?geo=FE-1 [accessed 2022-11-29]
42. Lahouati M, De Coucy A, Sarlangue J, Cazanave C. Spread of vaccine hesitancy in France: What about YouTube™? Vaccine 2020 Aug 10;38(36):5779-5782. [doi: 10.1016/j.vaccine.2020.07.002] [Medline: 32682617]

Abbreviations

CFA: confirmatory factor analysis
CNIL: Commission Nationale de l'Informatique et des Libertés
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
CPP: Comité de Protection des Personnes
DVL: digital vaccine literacy
EFA: exploratory factor analysis
EU GDPR: European Union General Data Protection Regulation
ICC: intraclass correlation coefficient

Edited by R Kukafka; submitted 03.05.22; peer-reviewed by J Andriesen, G Seçkin; comments to author 06.07.22; revised version received 18.07.22; accepted 06.09.22; published 14.12.22.

Please cite as:
Montagni I, Pouymayou A, Pereira E, Tzourio C, Schück S, Texier N, González-Caballero JL, The CONFINS Group
Measuring Digital Vaccine Literacy: Development and Psychometric Assessment of the Digital Vaccine Literacy Scale
J Med Internet Res 2022;24(12):e39220
URL: https://www.jmir.org/2022/12/e39220
doi:10.2196/39220
PMID:36515982

Original Paper

Assessment of Clinical Information Quality in Digital Health Technologies: International eDelphi Study

Kayode Philip Fadahunsi[1], MBBS, MPH; Petra A Wark[1,2], MSc, PhD; Nikolaos Mastellos[1], BSc, MSc, PhD; Ana Luisa Neves[1], MSc, MD, PhD; Joseph Gallagher[3], BA, MBChB, BAO, MD; Azeem Majeed[1], MD; Andrew Webster[4], MBBS, MSc, PGDip; Anthony Smith[5], MSc, MBChB; Brian Choo-Kang[6], BSc, MBChB, PGDip, PhD; Catherine Leon[7], BPharm, MSc; Christopher Edwards[8], BSc, MBChB; Conor O'Shea[9], MBChB, BAO; Elizabeth Heitz[10], MBBS; Olamide Valentine Kayode[11], MBBS; Makeba Nash[10], BSc, MA, MSc; Martin Kowalski[12], MD; Mateen Jiwani[1], MBBS, PGDip; Michael Edmund O'Callaghan[13], BEng, BMBS, MSc; Nabil Zary[14], MD, PhD; Nicola Henderson[15], BSc, MSc; Niels H Chavannes[16], MD, PhD; Rok Čivljak[17,18], MD, PhD; Olubunmi Abiola Olubiyi[19], MBBS, MPH; Piyush Mahapatra[20], MA, MBBS; Rishi Nannan Panday[21], MD, PhD; Sunday O Oriji[22], MBBS; Tatiana Erlikh Fox[23], MSc, MD; Victoria Faint[4], DipHE, BSc, PGCert; Josip Car[1,23], MSc, MD, DIC, PhD

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

[2]Centre for Intelligent Healthcare, Institute for Health and Wellbeing, Coventry University, Coventry, United Kingdom

[3]gHealth Research Group, School of Medicine, University College Dublin, Dublin, Ireland

[4]Leeds Teaching Hospital National Health Service Trust, Leeds, United Kingdom

[5]Health Board, Hywel Dda University, Wales, United Kingdom

[6]Glasgow Royal Infirmary, National Health Service Greater Glasgow and Clyde, Glasgow, United Kingdom

[7]Oxford University National Health Service Foundation Trusts, Oxford, United Kingdom

[8]Leeds Children's Hospital, Leeds, United Kingdom

[9]Wheaton Hall Medical Practice, Drogheda, Co Louth, Ireland

[10]Croydon Health Services, Croydon, United Kingdom

[11]Department of Paediatrics and Child Health, Evercare Hospital Lekki, Lagos, Nigeria

[12]Department of Ophthalmology, University Eye Hospital Tübingen, Tübingen, Germany

[13]School of Medicine, University of Limerick, Limerick, Ireland

[14]Institute for Excellence in Health Professions Education, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

[15]National Health Service Forth Valley, Scotland, United Kingdom

[16]National eHealth Living Lab, Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[17]University Hospital for Infectious Diseases, Zagreb, Croatia

[18]University of Zagreb School of Medicine, Zagreb, Croatia

[19]Medical Research Council Unit The Gambia, London School of Hygiene & Tropical Medicine, Fajara, Gambia

[20]West Hertfordshire Teaching Hospitals National Health Service Trust, Watford, United Kingdom

[21]Acute Medicine Section, Department of Internal Medicine, Amsterdam University Medical Centers, VU Medical Center Location, Amsterdam, Netherlands

[22]Department of Mental Health, Nnewi Campus, Nnamdi Azikiwe University Teaching Hospital, Nnewi, Nigeria

[23]Centre for Population Health Sciences, Lee Kong Chian Medicine, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Josip Car, MSc, MD, DIC, PhD
Department of Primary Care and Public Health
Imperial College London
The Reynolds Building, St Dunstan's Road
London, W6 8RP
United Kingdom
Phone: 44 020 7594 0799
Email: josip.car@imperial.ac.uk

Abstract

Background: Digital health technologies (DHTs), such as electronic health records and prescribing systems, are transforming health care delivery around the world. The quality of information in DHTs is key to the quality and safety of care. We developed a novel clinical information quality (CLIQ) framework to assess the quality of clinical information in DHTs.

Objective: This study explored clinicians' perspectives on the relevance, definition, and assessment of information quality dimensions in the CLIQ framework.

Methods: We used a systematic and iterative eDelphi approach to engage clinicians who had information governance roles or personal interest in information governance; the clinicians were recruited through purposive and snowball sampling techniques. Data were collected using semistructured online questionnaires until consensus was reached on the information quality dimensions in the CLIQ framework. Responses on the relevance of the dimensions were summarized to inform decisions on retention of the dimensions according to prespecified rules. Thematic analysis of the free-text responses was used to revise definitions and the assessment of dimensions.

Results: Thirty-five clinicians from 10 countries participated in the study, which was concluded after the second round. Consensus was reached on all dimensions and categories in the CLIQ framework: informativeness (accuracy, completeness, interpretability, plausibility, provenance, and relevance), availability (accessibility, portability, security, and timeliness), and usability (conformance, consistency, and maintainability). A new dimension, searchability, was introduced in the availability category to account for the ease of finding needed information in the DHTs. Certain dimensions were renamed, and some definitions were rephrased to improve clarity.

Conclusions: The CLIQ framework reached a high expert consensus and clarity of language relating to the information quality dimensions. The framework can be used by health care managers and institutions as a pragmatic tool for identifying and forestalling information quality problems that could compromise patient safety and quality of care.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2021-057430

(J Med Internet Res 2022;24(12):e41889) doi:10.2196/41889

KEYWORDS

information quality; digital health technology; patient safety; perspective; digital health technologies; DHT; thematic analysis; clarity; understandable; understandability; readability; searchability; security; decision support system; framework development; framework

Introduction

Digital health technologies (DHTs), such as electronic health records, electronic prescribing systems, and clinical decision support systems, have transformed health care delivery around the world [1]. However, the quality of information obtained from DHTs varies and can compromise quality and safety of care [2-4]. Several incidents of delayed, missing, partial, or wrong information in DHTs have been documented, resulting in adverse patient outcomes, including death [3-5]. To reduce the risk of such incidents, we need a pragmatic approach to assessing the quality of clinical information in DHTs. The importance of such an information quality assessment tool continues to grow with increasing automation and use of artificial intelligence (AI) in health care, as human checks are reduced and clinical information feeds into AI tools and algorithms [6].

A systematic review of the literature identified existing frameworks and dimensions that are relevant to assessing clinical information in DHTs [7]. However, the review found that the existing frameworks did not provide assessment tools for clinical practice [7]. In addition, most of the existing frameworks were developed without input from clinicians who use clinical information from DHTs [7]. Drawing on the review's findings, we developed a clinical information quality (CLIQ) framework as a pragmatic approach to assessing the quality of clinical information in DHTs. The CLIQ framework defined 13 dimensions relevant to the quality of clinical information in DHTs and was accompanied by a questionnaire for assessing information quality. The current study explored clinicians' perspectives on the relevance, definition, and assessment of information quality dimensions in the CLIQ framework (Textbox 1 shows the original dimensions in the CLIQ framework).

Textbox 1. Information quality dimensions in the original CLIQ framework.

- Informativeness (accuracy, completeness, interpretability, plausibility, provenance, and relevance)
- Availability (accessibility, portability, security, and timeliness)
- Usability (conformance, consistency, and maintainability)

Methods

Study Design

In this study, the eDelphi method was used to obtain direct input from clinicians to contextualize the CLIQ framework to the needs of the information users. This method uses a systematic process for engaging and integrating the opinions of multiple experts to reach consensus [8,9]. Thus, the eDelphi method was suitable for this study, which sought to obtain the consensus of clinicians from different countries on the information quality dimensions that are relevant to assessing clinical information in DHTs. In addition, the asynchronous approach gave the panelists an opportunity for equal participation, in contrast to physical meetings, which are usually dominated by a few outspoken participants [10]. The iterative process of the eDelphi method enabled the participants to provide feedback and reconsider their opinions based on collective responses [11]. The flexibility of the eDelphi method allowed collection of quantitative and qualitative data, which were useful in addressing the research question.

Ethics Approval

The protocol of this study was published to promote transparency [12]. Ethics approval was obtained for the study from the Imperial College Research Ethics Committee (20IC6396).

Steering Committee

This eDelphi study was coordinated by a steering committee comprising health care researchers and clinicians with interest in digital health. The committee developed the original CLIQ framework [7] and the accompanying questionnaire from which the initial items of the eDelphi study were generated. The committee recruited the participants to the study and made decisions regarding retention, removal, or redefinition of information quality dimensions based on the input of the participants according to prespecified decision and stoppage rules.

Decision and Stoppage Rules

The decision and stoppage rules on consensus were predefined to prevent bias during analysis [11]. An information quality dimension was considered relevant and was retained in the final framework when at least 70% of the participants, in any round of the survey, chose the options “strongly relevant” or “somewhat relevant.” The choice of 70% as a cutoff was a pragmatic choice based on the literature, as most Delphi studies use 60% agreement or higher as a threshold for consensus [10]. The study was planned to be concluded whenever consensus was reached on at least 80% of the dimensions or at the end of the third round, irrespective of the level of consensus [11].

Participant Recruitment

Clinicians with information governance roles or interest were invited to participate in the eDelphi panel based on the following eligibility criteria [12]: (1) prior or current experience of using DHTs in patient care, (2) information governance role or personal interest in information governance, and (3) willingness to participate in a multiple-round eDelphi study (up to 3 rounds). The heterogeneity of the participants provided a wide range of perspectives and increased the study’s external validity. The recruitment of the participants included both purposive and snowball sampling. Clinicians with information governance roles (eg, chief clinical information officer, chief nursing information officer, or Caldicott guardian) were targeted, as they have both DHT user experience and information governance expertise. However, participation was not restricted to these roles, as they do not exist in many low- and middle-income countries. Therefore, participants with interest in information governance without any formal information governance role were also recruited, such as clinicians who have published papers relating to information governance.

The steering committee members nominated clinicians from within and beyond their professional networks. Each eligible clinician was invited by an introductory email containing a link to the survey; the email also encouraged them to share the invitation with other eligible clinicians. Two reminders were sent at least 2 weeks apart to encourage participation [8]. Thirty-five clinicians from 10 countries participated in the study, including doctors, nurses, pharmacists, and other health care professionals.

Survey Content and Administration

The initial survey (Multimedia Appendix 1) was generated from the CLIQ framework [7] and the accompanying assessment questionnaire. The accompanying assessment questionnaire was developed by the steering committee based on the findings of a systematic review of information quality frameworks [7] and further evidence from literature. The survey was administered in English.

The introductory section of the survey provided brief information about the study, a link to the participant information sheet, and the electronic consent form. Demographic data were collected from participants who gave informed consent, and only these participants were shown the remainder of the survey.

The second section of the survey consisted of questions relating to the CLIQ framework. The first part of this section included 5-point Likert scale questions on the relevance of the dimensions in the CLIQ framework to quality and safety of care. The Likert scale captured a range of options (strongly relevant, somewhat relevant, neither relevant nor irrelevant, somewhat irrelevant, and strongly irrelevant) that represent categories people naturally create and thus did not require a heavy cognitive load. The second part comprised multiple-choice and free-text questions on the definition, assessment, and categories of the dimensions in the CLIQ framework. Finally, the email addresses of participants were collected for feedback purposes and as a contact method for the next round of the survey. The survey was set up using Qualtrics software (Qualtrics) and piloted by the steering committee members before its administration. The study was conducted between June 2021 and March 2022.

Data Analysis

The data on the relevance of the dimensions were summarized using descriptive statistics and used to inform decisions on retention of dimensions and termination of the study. The data were also used to provide feedback to the participants during

the second round of the survey. The free-text suggestions were analyzed using a reflexive thematic analysis approach, which allowed the steering committee members to go beyond the text to decode the meaning intended by the participants [13]. The thematic analysis process was adapted to include the following key stages: (1) studying the free-text suggestions to become familiar with the contributions made by the participants; (2) data coding to highlight key issues identified by the participants with regards to the definition and assessment of the dimensions; and (3) identifying patterns in the suggested modifications, developing themes, reflecting on these themes in the context of the overall data set, and defining the essence of each theme.

The themes were then used to revise the definitions and the assessment of the dimensions as appropriate. Feedback from the free-text suggestions and the changes that were made were also incorporated into the second round of the survey.

Results

Statistical Summary of Findings in the First Round

Thirty-five clinicians (including 26 doctors, 5 nurses, 2 pharmacists, 1 dietician, and 1 health system specialist) from 10 countries participated in the first round of this eDelphi study, with most being doctors (n=26, 74%) and male (n=23, 66%). About half of the participants had more than 10 years of digital health experience (n=18, 51%), and about half were from the United Kingdom (n=18, 51%). Most of the countries from which the participants came were high-income countries (8/10, 80%), although 1 of the 10 countries (10%) was lower middle income (Nigeria) and 1 (10%) was low income (the Gambia). Table 1 provides more detailed information on the sociodemographic characteristics of the participants.

In the first round of the eDelphi study, 86% to 97% of the clinicians ranked each of the 13 information quality dimensions in the proposed framework as relevant. These values were above the predefined threshold of 70% for the study and indicated consensus on the relevance of all 13 proposed dimensions in the framework. The ranking of the information quality dimensions is shown in Table 2.

Table 1. Sociodemographic characteristics of the eDelphi participants (N=35).

Characteristics	Participants, n (%)
Occupation	
Doctor	26 (74)
Nurse/nurse practitioner/advanced care practitioner	5 (14)
Pharmacist/clinical pharmacist	2 (6)
Dietician	1 (3)
Health system specialist	1 (3)
Digital health experience (years)	
Less than 10	17 (49)
10 or more	18 (51)
Country	
Croatia	1 (3)
The Gambia	1 (3)
Germany	1 (3)
Ireland	5 (14)
The Netherlands	3 (9)
Nigeria	2 (6)
Singapore	1 (3)
United Arab Emirates	1 (3)
United Kingdom	18 (51)
United States of America	2 (6)
Sex	
Male	23 (66)
Female	12 (34)

Table 2. Ranking of the dimensions in the clinical information quality framework in the first round of the eDelphi study, with number of responses by participants (N =35) in selected categories.

Rank	Information quality dimension	"Strongly relevant," n (%)	"Somewhat relevant," n (%)	Combined relevance ("strongly relevant" or "somewhat relevant"), n (%)
1	Accuracy	30 (86)	2 (6)	32 (92)
2	Completeness	18 (51)	14 (40)	32 (91)
3	Interpretability	23 (66)	8 (23)	31 (89)
4	Plausibility	13 (37)	18 (51)	31 (89)
5	Provenance	27 (77)	7 (20)	34 (97)
6	Relevance	18 (51)	15 (43)	33 (94)
7	Accessibility	28 (80)	4 (11)	32 (91)
8	Portability	18 (51)	12 (34)	30 (86)
9	Security	25 (71)	5 (14)	30 (86)
10	Timeliness	25 (71)	9 (26)	34 (97)
11	Conformance	15 (43)	16 (46)	31 (89)
12	Consistency	10 (29)	20 (57)	30 (86)
13	Maintainability	20 (57)	14 (40)	34 (97)

Changes Based on Free-Text Suggestions in the First Round

The changes that were made by the steering committee members based on the suggestions of the panel members in the first round are presented in Multimedia Appendix 2. The themes from the reflective thematic analysis of the free-text suggestions during the first round that informed these changes are presented in this section and summarized in Textbox 2.

Textbox 2. Themes from the free-text suggestions in the first round.

- Avoiding ambiguity: this expresses the need to avoid ambiguous terms and phrases.
- Relatable examples: this indicates the recommendation to include examples relating to daily activities to make the questions and definitions more explicit.
- Renaming the dimensions: this relates to suggestions for naming and renaming of dimensions.
- Rephrasing for clarity: this expresses the need to rephrase aspects of the questionnaire to improve clarity.

Avoiding Ambiguity

The participants described some terms in the questionnaire as "vague," "odd," and "confusing." For example, a participant stated the following about "errors":

> *The term "errors" needs to be further defined, now it is too vague, and I have no idea what to think of when I read it.*

In addition, some definitions were considered too complex to be understood by clinicians without informatics experience, as demonstrated by this comment:

> *Just at this point, I am thinking that it is relevant to understand who your audience is with these questions. Not all clinicians would understand these questions, but clinical informatics professionals would.*

Several changes were made across the dimensions to avoid ambiguity, as recommended by the participants, including replacing or removing terms such as "free of errors," "occasionally," and "very" that were considered ambiguous by the participants.

Relatable Examples

Participants were unanimous that examples were useful in making questions more explicit. One participant advocated including an example for each option:

> *Give examples in each of the options, that would make it easier to differentiate.*

On the other hand, another participant suggested including an example in the main question:

> *Perhaps include the example within the question, rather than the choice of answers.*

Participants also advocated using specific examples that were relevant to daily activities of the clinicians. They proceeded to suggest examples they considered appropriate for each option.

> *Phone call to IT [information technology] dept is not sufficiently accessible, it's another barrier (with a potential to fail- on hold, engaged, deadline, etc).*
>
> *Pharma/tobacco or any other commercial marketing would be "very untrustworthy."*

However, participants acknowledged that it might be difficult to find suitable examples to illustrate some response options.

> *I'm struggling with the plausible/very plausible examples but can't at this time think of an alternative.*

Changes relating to this theme include introducing examples such as "two-factor authentication" to describe secure information and reassigning examples as suggested by participants, such as reassigning "access requiring phone call to IT [information technology] department" under "inaccessible" information.

Renaming Dimensions

Although all the dimensions were considered relevant, the free-text suggestions indicated a need for renaming some dimensions:

> *I don't like the use of the word "interpretable" in the context of digital health records as it is too similar to "interoperable" and easily mis-read. Comprehensibility? Information clarity?*

Some suggestions seemed to imply a need for a new dimension. A free-text suggestion on accessibility expressed concerns on how it might be difficult to search for information in a system holding the data.

> *I'd have the second option in the list, information is present in EHR [electronic health record] but have to spend time looking for it.*

Multiple suggestions on "timeliness" seemed to indicate "currency" was favored over "timeliness."

> *You could quickly log into a system that doesn't contain the most up to date patient information which would be far more concerning in terms of data quality than logging in slowly to a system with the most recent info in it.*

A new dimension, "searchability," was introduced. In addition, "timeliness," "provenance," and "consistency" were renamed "currency," "trustworthiness," and "consistency of presentation," respectively. Two suggestions from panel members that related to the renaming of dimensions but were not adopted to avoid ambiguity are presented in Multimedia Appendix 2.

Rephrasing for Clarity

Most of the suggested modifications related to the phrasing of the questionnaire. Each question and the associated options were rephrased as appropriate to clarify them. These modifications ranged from simple corrections such as typos to major changes introducing new ideas; these were addressed on a case-by-case basis.

> *The definition of an adverse event is too narrow. Consider reflecting both critical (patient safety) and non-critical (quality of care). Also, there is an implicit assumption that data will directly impact care - maybe use "contribute to" as opposed to "lead to."*

Thus, "adverse event" was replaced with an explanation of the likelihood that inaccurate information would affect quality of care and patient safety and the potential impact. Similarly, the phrase "intended task" was replaced with the term "patient care," which is more all-encompassing. Other instances of rephrasing are presented in Multimedia Appendix 2.

Results of the Second Round

A second round was conducted because the free-text suggestions indicated a need for an additional dimension. This round was also used to present the results of the first round to the participants and obtain further feedback on the modifications to the questionnaire. Full details on the modifications and point-by-point responses to the participants' full-text suggestions for each of the dimensions are included in the questionnaire for the second round (Multimedia Appendix 3).

Among clinicians who provided their email addresses during the first round, 22 of 30 (73%) completed the second round. The threshold for consensus was reached for the new dimension "searchability." Most of the participants agreed with the changes made to the definitions and assessments of the dimensions, ranging from 86% (n=19) for consistency of presentation to 100% (n=22) for accuracy, completeness, interpretability, maintainability, and searchability, with no further modifications suggested. Minor suggestions were made regarding rephrasing the definitions of plausibility, trustworthiness, accessibility, portability, security, conformance, and consistency of presentation. Multiple free-text suggestions indicated that the term "currency" was not as acceptable as "timeliness":

> *I think timeliness and currency are two different terms that could not be used interchangeably. Therefore, I would prefer timeliness was not removed. if a result of an investigation is timely, it means it would be useful for decision making.*

> *I don't like the word currency in this context (it sounds like it's referring to money).*

The dimension "currency" was therefore reverted to the original name "timeliness." The modified CLIQ framework is made up of 14 dimensions, as outlined in Table 3. The accompanying assessment questionnaire is presented in Multimedia Appendix 4.

Table 3. Clinical information quality framework for digital health.

Dimension	Description
Informativeness (the usefulness of digital information for clinical purposes)	
Accuracy	The extent to which information is accurate.
Completeness	The extent to which no required information is missing.
Interpretability	The extent to which information can be interpreted.
Plausibility	The extent to which information makes sense based on clinical knowledge.
Trustworthiness	The extent to which the source of information is trustworthy and verifiable.
Relevance	The extent to which information is useful for patient care.
Availability (the functionality of the system holding clinical information)	
Accessibility	The extent to which information is accessible.
Portability	The extent to which information can be moved or transferred between different systems.
Searchability	The extent to which needed information can be found.
Security	The extent to which information is protected from unauthorized access, corruption, and damage.
Timeliness	The extent to which information is up-to-date.
Usability (the ease of use of clinical information)	
Conformance	The extent to which information is presented in a format that complies with institutional, national, or international standards.
Consistency of presentation	The extent to which presentation of information adheres to the same set of institutional, national, or international standards.
Maintainability	The extent to which information can be maintained (eg, modified, corrected, updated, adapted, and upgraded) to achieve intended improvement.

Discussion

Principal Findings

This study was conducted to contextualize the CLIQ framework to the needs of clinicians. Consensus was reached on the relevance of all the existing dimensions and categories of the CLIQ framework, including informativeness (accuracy, completeness, interpretability, plausibility, provenance, and relevance), availability (accessibility, portability, security, and timeliness), and usability (conformance, consistency, and maintainability). A new dimension, searchability, was introduced in the "availability" category to account for the ease of finding needed information in the DHTs. "Provenance" and "consistency" were renamed "trustworthiness" and "consistency of presentation," respectively.

The questionnaire was modified based on the suggestions of the clinicians to avoid ambiguities that could confuse users and affect the validity of the questionnaire. Nonspecific terms, such as "very," "few," or "occasionally," were removed, as their meanings vary based on context. Certain dimensions, such as conformance, were redefined using nontechnical terms, making them comprehensible to clinicians without an informatics background. In addition, the clarity of the questionnaire was improved by rephrasing the questions, incorporating relatable examples, and renaming certain dimensions. Overall, these changes made the questionnaire more user-friendly and improved its face and content validity.

Comparison With Prior Work

The CLIQ framework was developed to address gaps, including a lack of a pragmatic tool for clinical information quality assessment and the noninvolvement of clinicians in the development of existing frameworks [7]. The CLIQ framework is accompanied by a pragmatic questionnaire for assessing clinical information in DHTs, unlike theoretical frameworks, which provide no means of assessment [14-20]. The involvement of clinicians across 10 countries in the development of the CLIQ framework further differentiates the framework from existing frameworks, which were developed without input from clinicians [14,16-21]. Finally, the CLIQ framework is applicable to different DHTs, while existing frameworks are only applicable to specific DHTs, such as electronic health records [16,17,19,20,22].

Strengths and Limitations

The eDelphi method afforded a systematic, practical, affordable, and transparent approach to integrating the opinions of multidisciplinary clinicians from 10 countries. The importance of multiple eDelphi rounds, which allow feedback on changes made in preceding rounds [9,10], was demonstrated in the rejection of the attempt to rename "timeliness" as "currency." In addition, this study took advantage of the clinical experience and information governance expertise of the participating clinicians, thus combining practical user experience and subject matter expertise. The heterogeneous composition of the expert panel, which consisted of people from multiple clinical professions across 10 countries, enhanced the external validity of the CLIQ framework. However, external validity may be

limited by the low proportion of participants from low- and middle-income countries. The snowball sampling technique might have contributed to the disproportionately higher number of participants who were doctors from the United Kingdom. Nevertheless, the participants in this study were actively engaged and went out of their way to scrutinize all the definitions and offer valuable suggestions to improve the CLIQ framework. Finally, the number of participants that completed the second round of the eDelphi study was modest (22/30, 73%) but this is still more than the 8 to 15 experts recommended in the literature for a Delphi study [8].

Implications for Policy, Practice, and Future Research

This study provides insight into the information quality dimensions that are considered relevant by clinicians. Such insight could be useful when developing or choosing new DHTs for health care institutions. The consideration of relevant information quality dimensions while developing or choosing new DHTs will ensure that the information is fit for purpose. The CLIQ framework is thus a potential source of vital information to policy makers, DHT developers, and health care managers. In addition, the framework could be used to identify information quality problems in existing DHTs. As part of quality improvement projects, the CLIQ questionnaire could be used to collect data on the quality of information in existing DHTs from clinicians using these DHTs in clinical practice. Insight from such projects could then be used in planning strategies to address identified information quality problems.

The modification of the CLIQ framework has made the framework user-friendly by taking into account the views of the information users, as recommended in the information quality literature [23]. However, the adopted expert panel approach mainly improved the face and content validity of the framework [24]. Face and content validity imply that an instrument measures what it is intended to measure [24]. Therefore, a follow-up study to evaluate the construct validity and reliability of the CLIQ framework is ongoing across the United Kingdom among health care professionals who use the SystmOne electronic patient record system. Similar studies could be replicated in the future in low- and middle-income countries to further assess and, if needed, improve the applicability of the framework in such settings. The CLIQ framework will be made available under a Creative Commons (CC BY) license to facilitate its use in future works by other researchers who are interested in adapting the questionnaire based on their needs.

Conclusions

The CLIQ framework reached a high expert consensus and clarity of language relating to the information quality dimensions. The study contextualized the questionnaire by obtaining direct input from clinicians who are users of clinical information in DHTs. The contextualized CLIQ framework offers a pragmatic approach to assessing clinical information in DHTs and could be used in practice to identify and forestall information quality problems that can compromise quality and safety of care.

Acknowledgments

We would like to thank all clinicians who participated in the eDelphi panel, including those who prefer not to be named in this publication. ALN is supported by the National Institute for Health and Care Research (NIHR) Patient Safety Translational Research Centre and the NIHR Biomedical Research Centre. AM and JC are supported by NIHR Applied Research Collaboration Northwest London. The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. KPF's doctoral research is funded by the Federal Government of Nigeria.

Data Availability

Researchers can apply for access to the pseudonymized data by writing to the corresponding author.

Authors' Contributions

KPF conceived the study and drafted the manuscript. KPF, PAW, ALN, NM, JG, AM, and JC were members of the steering committee. PM, NHC, NZ, MEO, RC, RNP, OAO, TEF, BCK, SOO, CO, CE, AS, AW, MN, OVK, VF, NH, CL, MK, MJ, and EH were members of the expert panel. All authors revised the manuscript for important intellectual content.

Conflicts of Interest

PM is an executive director of Open Medical Limited, a digital health company. The authors have no further interests to declare.

Multimedia Appendix 1
First round eDelphi survey.
[DOCX File , 208 KB - jmir_v24i12e41889_app1.docx]

Multimedia Appendix 2
Changes based on free-text suggestions.
[DOCX File , 16 KB - jmir_v24i12e41889_app2.docx]

Multimedia Appendix 3
Second round eDelphi survey.
[DOCX File , 31 KB - jmir_v24i12e41889_app3.docx]

Multimedia Appendix 4
Clinical information quality (CLIQ) assessment questionnaire.
[DOCX File , 18 KB - jmir_v24i12e41889_app4.docx]

References

1. World Health Organization. Global strategy on digital health 2020-2025. Geneva, Switzerland: World Health Organisation; 2021:1-45.
2. Magrabi F, Ong MS, Runciman W, Coiera E. Patient safety problems associated with heathcare information technology: an analysis of adverse events reported to the US Food and Drug Administration. AMIA Annu Symp Proc 2011;2011:853-857 [FREE Full text] [Medline: 22195143]
3. Magrabi F, Aarts J, Nohr C, Baker M, Harrison S, Pelayo S, et al. A comparative review of patient safety initiatives for national health information technology. Int J Med Inform 2013 May;82(5):e139-e148. [doi: 10.1016/j.ijmedinf.2012.11.014] [Medline: 23266061]
4. Meeks DW, Smith MW, Taylor L, Sittig DF, Scott JM, Singh H. An analysis of electronic health record-related patient safety concerns. J Am Med Inform Assoc 2014;21(6):1053-1059 [FREE Full text] [doi: 10.1136/amiajnl-2013-002578] [Medline: 24951796]
5. Kim M, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. J Am Med Inform Assoc 2017 Mar 01;24(2):246-250 [FREE Full text] [doi: 10.1093/jamia/ocw154] [Medline: 28011595]
6. Bitterman DS, Aerts HJWL, Mak RH. Approaching autonomy in medical artificial intelligence. Lancet Digit Health 2020 Sep;2(9):e447-e449 [FREE Full text] [doi: 10.1016/S2589-7500(20)30187-4] [Medline: 33328110]
7. Fadahunsi KP, O'Connor S, Akinlua JT, Wark PA, Gallagher J, Carroll C, et al. Information quality frameworks for digital health technologies: systematic review. J Med Internet Res 2021 May 17;23(5):e23479 [FREE Full text] [doi: 10.2196/23479] [Medline: 33835034]
8. Trevelyan EG, Robinson PN. Delphi methodology in health research: how to do it? Eur J Integr Med 2015 Aug;7(4):423-428. [doi: 10.1016/j.eujim.2015.07.002]
9. Diamond I, Grant R, Feldman B, Pencharz P, Ling S, Moore A, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014 Apr;67(4):401-409 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.12.002] [Medline: 24581294]
10. Hasson F, Keeney S. Enhancing rigour in the Delphi technique research. Technol Forecast Soc Change 2011 Nov;78(9):1695-1704. [doi: 10.1016/j.techfore.2011.04.005]
11. Niederberger M, Spranger J. Delphi technique in health sciences: a map. Front Public Health 2020 Sep 22;8:457 [FREE Full text] [doi: 10.3389/fpubh.2020.00457] [Medline: 33072683]
12. Fadahunsi KP, Wark PA, Mastellos N, Gallagher J, Majeed A, Car J. Clinical information quality of digital health technologies: protocol for an international eDelphi study. BMJ Open 2022 Apr 22;12(4):e057430 [FREE Full text] [doi: 10.1136/bmjopen-2021-057430] [Medline: 35459673]
13. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2020 Aug 12;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
14. Bolt T, Kano S, Kodate A. Information quality in home care coordination services. J Telemed Telecare 2016 Dec 02;13(1_suppl):7-9. [doi: 10.1258/135763307781645149]
15. McCormack J, Ash J. Clinician perspectives on the quality of patient data used for clinical decision support: a qualitative study. AMIA Annu Symp Proc 2012;2012:1302-1309 [FREE Full text] [Medline: 23304409]
16. Stetson P, Bakken S, Wrenn J, Siegler E. Assessing electronic note quality using the physician documentation quality instrument (PDQI-9). Appl Clin Inform 2012;3(2):164-174 [FREE Full text] [doi: 10.4338/aci-2011-11-ra-0070] [Medline: 22577483]
17. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 01;20(1):144-151 [FREE Full text] [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
18. Almutiry O, Wills G, Alwabel A, Crowder R, Walters R. Toward a framework for data quality in cloud-based health information system. 2013 Presented at: International Conference on Information Society (i-Society 2013); Jun 24-26, 2013; Toronto, ON.
19. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC) 2016;4(1):1244 [FREE Full text] [doi: 10.13063/2327-9214.1244] [Medline: 27713905]

20. Davoudi S, Dooling J, Glondys B, Jones T, Kadlec L, Overgaard S, et al. Data Quality Management Model (Updated). J Am Health Inf Manag Assoc 2015 Oct;86(10):62-65. [Medline: 26665585]
21. Bowen M. EMR Data Quality: Evaluation Guide. eHealth Observatory. URL: http://ehealth.uvic.ca/resources/tools/EMRsystem/2012.04.24-DataQualityEvaluationGuide-v1.0.pdf [accessed 2022-11-25]
22. Almutiry O. Data Quality Assessment Instrument For Electronic Health Record Systems in Saudi Arabia. Doctoral Thesis. University of Southampton. 2017. URL: https://eprints.soton.ac.uk/419029/1/Final_thesis_Omar_Saud_Almutiry.pdf [accessed 2022-11-17]
23. Eppler M, Wittig D. Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years. In: Proceedings of the 2000 Conference on Information Quality. 2000 Presented at: 2000 Conference on Information Quality; Oct 20-22, 2000; Cambridge, MA URL: http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202000/Papers/ConceptIQaReviewofIQFramework.pdf
24. Bolarinwa O. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Niger Postgrad Med J 2015;22(4):195-201 [FREE Full text] [doi: 10.4103/1117-1936.173959] [Medline: 26776330]

Abbreviations

AI: artificial intelligence
CLIQ: clinical information quality
DHT: digital health technology

Edited by G Eysenbach; submitted 13.08.22; peer-reviewed by J Davis, N O'Brien; comments to author 07.10.22; revised version received 29.10.22; accepted 17.11.22; published 06.12.22.

Please cite as:
Fadahunsi KP, Wark PA, Mastellos N, Neves AL, Gallagher J, Majeed A, Webster A, Smith A, Choo-Kang B, Leon C, Edwards C, O'Shea C, Heitz E, Kayode OV, Nash M, Kowalski M, Jiwani M, O'Callaghan ME, Zary N, Henderson N, Chavannes NH, Čivljak R, Olubiyi OA, Mahapatra P, Panday RN, Oriji SO, Fox TE, Faint V, Car J
Assessment of Clinical Information Quality in Digital Health Technologies: International eDelphi Study
J Med Internet Res 2022;24(12):e41889
URL: https://www.jmir.org/2022/12/e41889
doi:10.2196/41889
PMID:36472901

Original Paper

Web-Based Short Video Intervention and Short Message Comparison of Repeat Blood Donation Behavior Based on an Extended Theory of Planned Behavior: Prospective Randomized Controlled Trial Study

Qiuyue Hu[1], MD; Wei Hu[1], MD; Wenjuan Han[1], MD; Lingling Pan[1], MPH

Blood Center of Zhejiang Province, Hangzhou, China

Corresponding Author:
Lingling Pan, MPH
Blood Center of Zhejiang Province
789 Jianye Road, Binjiang District
Hangzhou, 310052
China
Phone: 86 13588809060
Email: panbbs@163.com

Abstract

Background: Although blood is an indispensable and important resource for clinical treatment, an imbalance between supply and demand may occur as the population ages and diversifies. Studies indicate that repeat blood donors are safe blood sources because of their voluntary blood donation education and frequent blood screening. However, the high rate of reduction in the number of first-time voluntary blood donors and low rate of repeated blood donation are common problems worldwide.

Objective: This study aimed to evaluate the effect of an intervention in nonregular blood donors using web-based videos and SMS text messages, in which the former was guided by the extended theory of planned behavior, to discover effective intervention methods to improve repeat blood donation rates among nonregular blood donors.

Methods: A total of 692 nonregular blood donors in Zhejiang province were randomly divided into intervention and control groups. The control group received regular, short reminder messages for a 6-month period, whereas the intervention group received web-based videos on the WeChat platform. The intervention group was guided by an extended theory of planned behavior, which included 9 factors: the respondents' attitude, subjective behavioral norms, perceived behavioral control, the willingness to donate blood, outcome expectations, self-identity, blood donation–related anxiety, cognition of the blood donation environment, and previous blood donation experience. The intervention group was divided into 2 stages: those with an intervention at 3 months and those with a follow-up 3 months later. After 6 months, the redonation rate was evaluated for the 2 groups, and the scale in the intervention group was determined both before and after the intervention. A *t* test, chi-square test, logistic stepwise regression, and ANOVA were performed.

Results: The intervention group's redonation rate was 16.14%, which was significantly higher than the control group's redonation rate of 5.16%; $P<.001$. Men who were aged 31 to 45 years and had donated blood twice had a higher redonation rate after the web-based video intervention than after the SMS text messages; $P<.05$. The repeat donors' improved blood donation anxiety ($P=.01$), outcome expectations ($P=.008$), and cognition of the blood donation environment ($P=.005$) after the intervention were significantly higher than those of the nonrepeat donors.

Conclusions: The web-based short video intervention based on the extended theory of planned behavior can effectively improve redonation rates. Outcome expectations, blood donation anxiety, and cognition of the blood donation environment can directly influence irregular blood donors to redonate blood.

(J Med Internet Res 2022;24(12):e37467) doi:10.2196/37467

KEYWORDS

extended theory of planned behavior; repeated blood donation intervention; randomized controlled trial; mobile phone

Introduction

Background

Although blood is an indispensable and important resource for clinical treatment, an imbalance between supply and demand may occur as the population ages and diversifies [1]. Studies indicate that repeat blood donors are safe blood sources because of their voluntary blood donation education and frequent blood screening [2-4]. However, the high rate of reduction in the number of first-time voluntary blood donors and the low rate of repeated blood donation among are common problems worldwide. The rate of repeated blood donation in Zhejiang province from 2006 to 2015 was 30.8%, which was lower than the global rate of 50% and fell within the average range of 24.3% to 38.8% in China [5]. Hence, a common challenge—in Zhejiang province as well as nationally and internationally—involves the question of how the rate of repeat blood donations can be increased while ensuring an ample blood supply and safety. Current research on the evaluation, prediction, and behavioral intervention of repeated blood donation behavior is in its infancy. Furthermore, the relationship between personal, psychological, and socioenvironmental factors, among others, and repeated blood donation behavior has been clarified, nor has an authoritative evaluation system been developed to index repeated blood donation intentions [6-8]. The literature has primarily focused on changing blood donation knowledge, attitude, and willingness through education [9,10]. Most methods involve traditional SMS text messages, phone calls, and brochures [11] and lack robustness in methodological reporting [12] and intervention studies on repeated blood donation behavior. Few studies have addressed prospective randomized controlled trials of repeated blood donation intervention.

Objectives

The theory of planned behavior is the most widely used theory to explain behavioral motivation and has consistently demonstrated the ability to predict blood donation intention and behavior [13-17]. This theory posits that human behavior is determined by 3 aspects: the first factor is the consequences of a behavior and the evaluation of these results, which can generate positive or negative attitudes toward the behavior. The second factor comes from the normative expectations of others and the motivation to follow these expectations, namely normative beliefs, which lead to social pressure and subjective norms. The resources and opportunities required for this behavior, as well as their ease of access, are the control beliefs that lead to the third factor, that is, perceived behavioral control. Although a majority of studies have confirmed that the theory of planned behavior can effectively predict behavioral intentions and can significantly improve the explanatory and predictive power of behavioral research, such works also have various shortcomings, such as the omission of socioenvironmental factors and insignificant intervention effects [18,19]. Ajzen [20] observed that if a factor was found to enhance the prediction of an intention or behavior, the theory of planned behavior can extend the factor, forming an extended theory of planned behavior (ETPB). Therefore, this study's initial stage first considers a literature review and a Delphi expert consultation based on the theory of planned behavior's 4 dimensions: attitude, subjective behavioral norms, perceived behavioral control, and willingness. It also explores the expected outcome, self-identity, and blood donation anxiety and environment and ultimately forms an ETPB; further research is incorporated to form a repeat blood donation intention–assessment scale with this theory as the overall guiding framework (Textbox 1).

The "Statistical Report on Internet Development in China" indicates that as of December 2020, China's short videos reached an audience of 873 million people or 88.3% of all netizens [21]. The widespread popularity of these short videos suggests that people generally accept and enjoy them. Currently, videos are widely used in behavioral health interventions, such as patient education for different diseases and patient family care [22-28], but few studies have examined their application in blood donation environments. On the basis of the previous research results on the factors influencing repeated blood donation as guided by the ETPB [28-30], this study designed short videos based on the ETPB; these short videos were presented on the web to nonregular blood donors as repeated blood donation interventions. An exploratory, prospective, randomized, and controlled experiment was conducted to analyze the changes in intermediary variables before and after the intervention period. The results from repeated blood donation behavior were compared with those from the SMS control group to not only analyze the intervention effect but also provide a reference for empirical research in determining the next intervention strategy to ensure repeat blood donation behavior.

Textbox 1. The influencing factor scale of repeated blood donation based on the extended theory of planned behavior.

Factors and the corresponding items

1. Attitude
   a. I think donating blood can save lives.
   b. I think donating blood is a kind of blood storage protection for me and my family.
   c. I feel that giving blood demonstrates my courage.
   d. I think many people in the hospital need blood transfusions and need me to donate blood.
2. Subjective behavioral norms
   a. Most of the people who are important to me think I should donate blood or donate again.
   b. Most of the people who are important to me will support and encourage me to donate again.
   c. Most people I know will evaluate me based on whether I donate blood or donate again.
   d. I think donating blood is about everyone.
3. Perceived behavioral control
   a. The standardized process of voluntary blood donation will not be infected with diseases.
   b. I will pay attention to information on voluntary blood donations (such as those presented on the television, internet, newspapers, or magazines) and will actively acquire knowledge about voluntary blood donation.
   c. Each voluntary blood donation of 200-400 mL is in the normal range and will not damage the body.
   d. I will take the initiative to donate blood because my family, friends, or colleagues donate blood.
   e. I will encourage my family, friends, or colleagues to voluntarily donate blood.
   f. It is my decision to donate blood or continue to donate blood again.
   g. I can meet the necessary conditions, such as good health or a convenient time, among others, to increase the number of blood donations.
   h. If the blood donation experience will be positive, I will donate blood or donate blood again.
   i. If my family can prioritize transfusions as necessary after I donate blood, I will donate blood or donate blood again.
   j. I am confident I will overcome the factors that may prevent me from donating or continuing to donate blood.
   k. The preferential blood donation policy affirmed and encouraged me.
   l. in the next year, I plan to donate blood (or donate again).
4. Blood donation willingness
   a. I believe that I will be able to donate blood or donate blood again within the next year.
   b. Blood donation souvenirs or awards will motivate me to donate blood or donate blood again.
   c. In the next year, I will definitely donate blood (or donate again).
5. Outcome expectations
   a. If I donate blood again, more patients will be treated.
   b. If I donate blood again, I can set a good example for others.
   c. If I donate blood again, I will gain more recognition and respect.
   d. Voluntarily donating blood at regular intervals (6 months or more) is good for your health.
   e. If you do not donate blood, or do not continue to donate blood, you are likely to regret it in the future.
6. Self-identity
   a. I am the type of person who will donate blood (or continue to donate blood).
   b. I believe it is appropriate in every way for someone like me to donate blood (or donate blood again).
   c. Donating blood is a way of realizing one’s self-worth.
7. Donation anxiety
   a. I am concerned that my physical condition does not meet the blood donation requirements.

b. Dissatisfaction with the blood donation experience, whether when I have donated blood or heard from others, causes me to worry about donating blood.

c. If I am asked to donate blood or donate blood again, I will feel distressed and anxious.

8. Cognition of the blood donation environment

a. The blood donation environment looks clean and comfortable.

b. The blood donation environment looks safe.

c. The blood collection staff at the donation site are highly skilled.

d. The blood donation site's hours of operation are convenient for me.

e. The blood donation site's staff were friendly.

f. The blood donation site's location was convenient for me.

g. I have seen promotional materials for blood donation in the media.

9. Previous blood donation experience

a. Have you ever felt unbearable pain when donating blood?

b. Have you ever experienced dizziness, weakness, or a mild headache during or after donating blood?

c. Have you ever felt nervous when donating blood?

Methods

Research Design

This was a prospective, single-blind, randomized study. SMS text messages from the Zhejiang provincial blood management information system were sent to eligible, nonregular blood donors, inviting them to participate in the study. The text messages included an invitation letter and a research link. Blood donors who were willing to participate could click the link to obtain detailed information, such as the research objective and content, notice of informed consent, and the research group's contact information. Participants were randomly assigned to either a web-based intervention group or a SMS control group. As blood donors in China have a minimum interval of 6 months between donations, the study's SMS control group received a regular reminder SMS within the 6-month interval. The web-based intervention group was analyzed across 2 phases: the intervention period, or the first 3 months, and the follow-up period, or the next 3 months. A baseline survey was conducted using the scale before the intervention and reassessed using the same scale at the end of the intervention period. This scale's outcome measures were the 9 ETPB factors: attitude, subjective behavioral norms, perceived behavioral control, willingness, outcome expectations, self-identity, blood donation anxiety, the blood donation environment, and previous blood donation experience. At the end of the 3-month follow-up period, blood donation results of the intervention and control groups were tracked using the Zhejiang provincial blood management information system. We hypothesized that the blood donors who received the web-based intervention would donate again more often than those in the SMS control group, as mediated by increases in the 9 ETPB factors. The study protocol was approved by the ethics review committee of the Zhejiang provincial blood center.

Study Participants, Exclusion and Inclusion Criteria, and the Recruitment Method

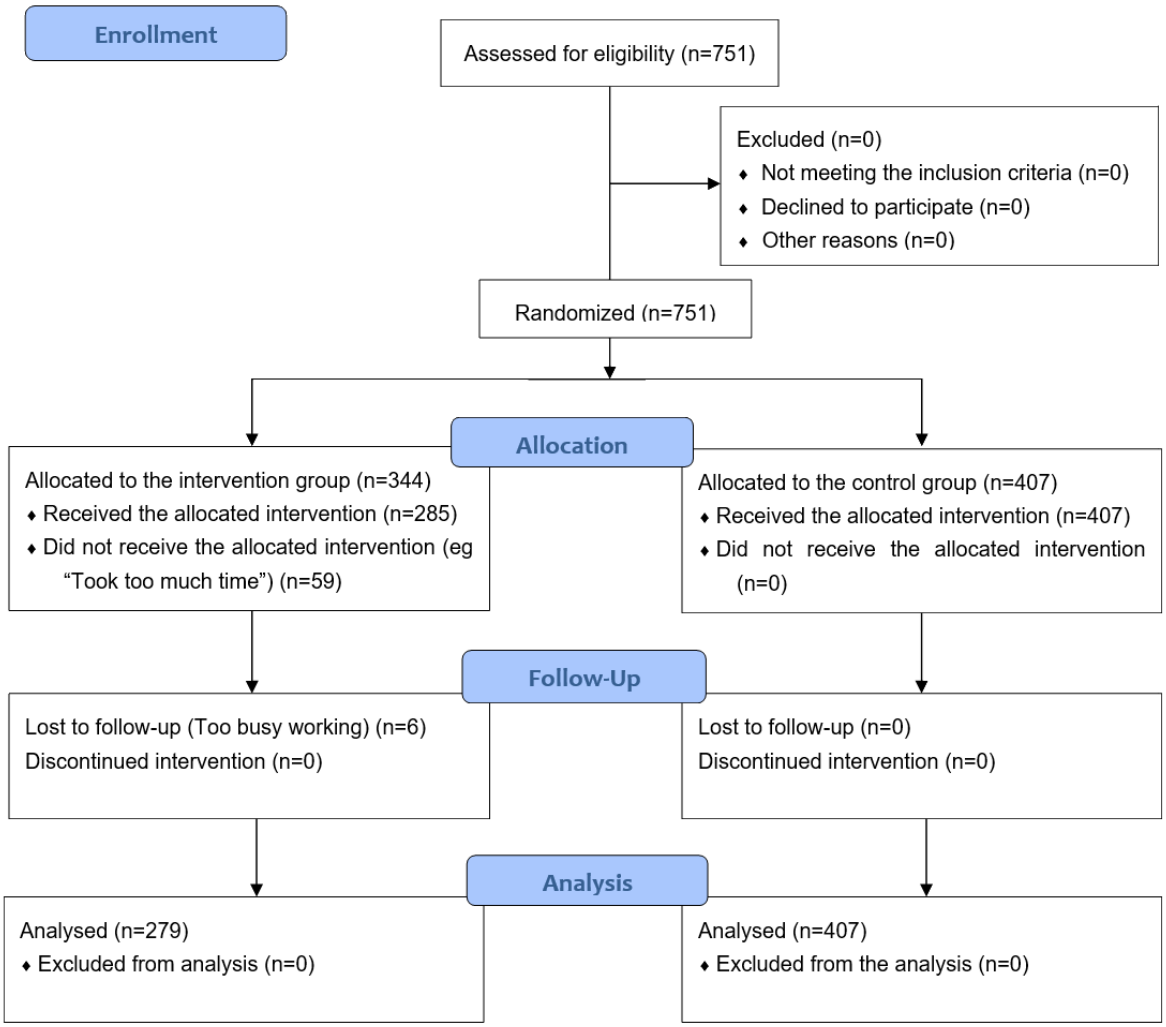
According to the World Health Organization's definition, regular blood donors are those who donated blood >3 times and at least once in recent year [31]. This study examines nonregular blood donors; for the convenience of observation, the inclusion criteria were blood donors aged 18 to 55 years, with current physical conditions meeting the requirements for blood donation, and who meet at least one of the following conditions: (1) donated whole blood in 2019 and did not donate again in 2020, consistent with the category of "*lost donor*" or "*those who donated blood at least once in the past 24 months but did not donate blood in the past 12 months*" [32]; or (2) whole-blood donors with fewer than 3 blood donations and who have not donated blood in the last 6 months, including first-time blood donors who had not donated blood in the past. Respondents were excluded if (1) their current physical condition did not meet the blood donation requirements and (2) they were "regular" blood donors or had donated blood at least 3 times and at least once in recent year.

According to the Zhejiang province's blood donation statistics in 2017 based on the Zhejiang blood information system, approximately 5% of blood donors had repeatedly donated blood within 6 months after meeting the blood donation interval requirement in Zhejiang province, or specifically, the control group's repeat blood donation rate was 5%. This study assumes that the intervention could consequently increase the repeat blood donation rate by at least 10% [33]. To detect a minimum 10% difference between the control and intervention groups, the repeat blood donation rate of the latter group was expected to be 15%. The target sample size of each group was calculated to be 141 (α=.05, β=.8). The estimated loss to follow-up rate was 25%; thus, the minimum sample size of each group was 176.

From March 9 to 15, 2021, an invitation was texted to all the research participants who met the inclusion criteria. All the 751

respondents who agreed to the invitation were coded by a computer and randomly divided into the web-based intervention (344 people) and SMS control groups (407 people; Figure 1).

Figure 1. Flow diagram.



Measurement

Web-Based Intervention Methods

Two outcome indicators were used for the web-based intervention: (1) the 3-month period from March 15 to June 15, 2021, was the intervention period, and the changes in the 9 influencing factors before and after the intervention were measured using the same scale and (2) from June 15 to September 15, 2021, the 3-month follow-up period after the intervention period ended, included an investigation of whether the participants donated blood again.

Baseline Measurement and Postintervention Reassessment

On the basis of the previous research results, this study adopted the "Repeated Blood Donation Influencing Factors Scale Based on ETPB" (or the "ETPB scale" hereafter) [30], which consists of 9 factors and 44 items. The responses were measured on a 5-point Likert scale and ranged from "strongly disagree" to "strongly agree." On March 15, 2021, the ETPB scale was sent to the web-based intervention group to collect the participants' baseline data. On June 15, the day the intervention ended, the same scale was sent again to measure the postintervention results.

Short Videos Based on the ETPB Elements and Short Videos Regularly Sent on the Web

The primary web-based intervention method involved sending weekly short videos, which were designed based on the ETPB elements, and timely web-based responses to questions from blood donors. This study used smartphones as the carrier because they are characterized as convenient, low cost, and unlimited by time and space, with positive effects, strong communication ability, and high acceptance [34-37]. Moreover, WeChat was chosen because it is easy to operate and free to use and because China's mainstream social media platform is the most widely used instant messaging tool [38,39]. This study's web-based intervention WeChat group was equivalent to a small internet-based community. The respondents could view this study's videos in real time, which facilitates the reception and reading of information and reduces disturbances to daily life while being highly interactive. The group could publicly respond to various frequently asked questions, such as those regarding blood donation locations and policies and how

long after vaccination one can donate blood, and eliminate similar doubts among other blood donors.

This study adopted group announcements, real-time communication, and group agency methods in WeChat groups after sending videos. The respondents click a button to complete a group chat after watching a video to let the researchers know that they have watched it. The short videos used in this study were between 45 seconds and 2 minutes in length. Studies have indicated that periodic reminders can encourage the occurrence and persistence of healthy patient behaviors [40-42]. Table 1 displays this study's short video content and the arrangement of the web-based intervention group.

Table 1. Web-based intervention videos' content and distribution schedule.

Intervention factor	Corresponding short video content	Description	Implementation date
Attitude	1. Why donate blood?	Letting blood donors understand the practical significance of donating blood to save others and promoting change in blood donation attitudes, from opposition and indifference to approval and understanding	Week 1
Self-identity	1. College students' blood donation stories 2. The blood donor family's donation story	Self-identity is an important part of self-awareness and the core self-regulatory system in self-awareness. In the human social environment, the process of becoming a qualified social member is inseparable from the growing maturity of self-awareness. This study uses college students' blood donation and family blood donation experiences to stimulate blood donors' self-identity regarding blood donation	Week 2
Cognition of blood donation environment	1. The blood donation environment (such as the most beautiful and digital blood donation site)	Sending a video depicting the most esthetic, state-of-the-art blood donation environment to convey the concepts of safety, hygiene, cleanliness, warmth, and convenience	Week 3
Blood donation anxiety	1. Responses regarding blood donation misconceptions 2. Why is there a charge for donating blood? 3. Blood donation knowledge	In providing relief to potential donors by eliminating misunderstandings, this study adopts a face-to-face attitude, with open and candid communication and response methods to reduce or alleviate the blood donors' anxiety	Week 4
Subjective behavioral norms	1. Stories of regular blood donor representatives 2. Volunteer service	Addressing blood donors' perceived normative expectations set by others and their motivation to follow those expectations; the video demonstrates that donating blood, as a part of service and selfless dedication to others, can bring spiritual satisfaction and joy and relieve the external pressure that blood donors experience	Weeks 5 and 6
Outcome expectations	1. Blood donation care policy 2. Blood donors get direct fee waived after transfusion 3. One blood recipient's college car accident story and a Rh-negative recipient's story	Sending videos of real cases where blood recipients have had their lives saved because of blood transfusions and communicating that timely blood transfusions can avoid the negative consequence of patient death; furthermore, post–blood donation results can include social honors and other care policies that can be enjoyed after donation	Weeks 7 and 8
Previous blood donation experience	1. Precautions taken for donating blood 2. The donation process 3. From one blood vessel to another, 3 topics: the blood source, detection, and blood preparation and supply	Sending videos to reshape the blood donors' scientific concept of blood donation and view such experiences as adverse reactions in previous blood donation experiences from a scientific perspective	Weeks 9 and 10
Perceived behavioral control	1. Reach out to donate blood 2. People who have donated blood many times show up 3. The first blood donation experience	By addressing the blood donor's awareness of whether they can donate blood again, the video enhanced the blood donor's confidence in their ability to donate blood again	Week 11
Blood donation willingness	1. Call for blood donations	The final week's video reinforces the significance of blood donation as conveyed in the discussion of the first factor. This will hopefully spur recipients to action and change blood donors' awareness and influence them to donate blood again	Week 12

Method for the SMS Control Group

In the SMS control group, only regular interval reminder messages were sent during the 6-month period. The content primarily thanked the blood donors for their selfless dedication, warmly reminded them that they have met the minimum required donation interval, and invited them to donate blood again. No other intervention methods were used, such as communication through the telephone or internet. The respondents promised

not to watch blood donation–related videos or read similar material during the study.

Statistical Methods

As SPSS (version 23.0; IBM Corp) software was used to organize the data, the measurement data were expressed as "mean (SD); x [s])," and the count data were expressed as a percentage (%). Furthermore, this study's statistical analysis was conducted through a chi-square test, 2 independent sample *t* tests, an ANOVA, and a logistic stepwise regression, among other methods. The results were statistically significant ($P<.05$).

Ethics Approval

This study was approved by the Regional Ethics Committee of the Blood Center of Zhejiang Province (approval number 2019-019). This study was conducted with the framework of randomized controlled trial, which was in full compliance with the CONSORT guidelines. In terms of content, it has no clinical trials, no human trials, no human samples, no medical records and other information, no human blood samples, pathological phenomena, disease etiology and pathogenesis, no disease prevention, diagnosis, treatment and rehabilitation information, and will not have any adverse effects on the human body. As this study was an observational study, we did not register in the Chinese Clinical Trial Registry. No personal privacy or medical information that can identify the blood donors and commercial interests will be disclosed.

Results

Overview

The average age of the participants in this study was 30.47 (SD 9.76) years, approximately 70.4% (487/692) of the participants were male, and the frequency of blood donation was mostly once (481/692, 69.5%) and twice (173/692, 25%). Responses of "*never*" and "*three or more times*" were included only in the intervention group. At the end of 6 months, the intervention group's blood donation rate was 16.1% and that of the SMS control group was 5.2%. The chi-square test ($c^2_1=23.1$; $P<.001$) results indicated that the difference in repeated donation rates between the groups was statistically significant (Table 2).

Table 2. Participants' demographic and donation information collected during the intervention period[a].

Items	Web-based intervention group (n=285), n (%)	SMS group (n=407), n (%)
Sex		
Male	198 (69.5)	289 (71)
Female	87 (30.5)	118 (29)
Intersex	0 (0)	0 (0)
Age (years)		
18-25	159 (55.8)	98 (24.1)
26-30	48 (16.8)	71 (17.4)
31-35	27 (9.5)	70 (17.2)
36-40	17 (6)	70 (17.2)
41-45	14 (4.9)	84 (20.6)
46-55	20 (7)	14 (3.4)
Number of previous blood donations		
0 time	24 (8.4)	0 (0)
1 time	130 (45.6)	351 (86.2)
2 times	117 (41.1)	56 (13.8)
≥3 times	14 (4.9)	0 (0)
Number of people who donated blood again within 6 months of observation period	46 (16.1)	21 (5.2)

[a]Chi-square test of the blood donation rate for the intervention and control groups during the observation period; $c^2_1=23.1$; $P<.001$.

Comparative Analysis of the 2 Groups

According to whether the participants in the web-based intervention and SMS control groups donated blood again during the study period, they were divided into the "redonating" and "nonredonating" groups, respectively. The results revealed that male blood donors who were aged 31 to 45 years and had donated twice in the past exhibited significant differences in their response to the text messages and web-based intervention, and the redonation rate was higher among such participants in the web-based intervention group (Table 3).

Table 3. Comparative analysis of repeat and nonrepeat donors in the SMS control and web-based intervention groups during the observation period.

Items	SMS group		Web-based intervention group		Chi-square (*df*)	*P* value
	Repeat donation (n=21), n (%)	Nonrepeat donation (n=386), n (%)	Repeat donation (n=46), n (%)	Nonrepeat donation (n=239), n (%)		
Sex						
Male	13 (62)	276 (71.5)	36 (78)	162 (67.8)	24.3 (1)	<.001
Female	8 (38)	110 (28.5)	10 (22)	77 (32.2)	1.4 (1)	.24
Intersex	0 (0)	0 (0)	0 (0)	0 (0)	N/A[a]	N/A
Age (years)						
18-25	5 (24)	80 (20.7)	17 (37)	142 (59.4)	1.6 (1)	.21
26-30	7 (33)	64 (16.6)	10 (22)	38 (15.9)	2.8 (1)	.09
31-35	3 (14)	68 (17.6)	6 (13)	21 (8.8)	5.6 (1)	.02
36-40	1 (5)	67 (17.4)	4 (9)	13 (5.4)	8.3 (1)	.004
41-45	5 (24)	68 (17.6)	7 (15)	7 (2.9)	14.9 (1)	<.001
46-55	0 (0)	39 (10.1)	2 (4)	18 (7.5)	1.6 (1)	.22
Blood donation times						
0 time	0	0	0	24 (10.0)	N/A	N/A
1 time	17 (81)	334 (86.5)	12 (26)	118 (49.4)	3.2 (1)	.07
2 times	4 (19)	52 (13.5)	30 (65)	87 (36.4)	8.2 (1)	.004
≥3 times	0 (0)	0 (0)	4 (9)	11 (4.6)	N/A	N/A

[a]N/A: not applicable.

Results of the Theory of Planned Behavior Scale Comparison in the Web-Based Intervention Group Before and After the Intervention

After the intervention and verification of the respondents' information, it was determined that 279 people completed both the baseline and postintervention surveys. The statistical results presented in Table 4 indicate the clear effects of the web-based intervention. The 9 factors—specifically, participants' attitude, subjective behavioral norms, perceived behavioral control, blood donation willingness, expectation of the results, self-identity, blood donation anxiety, cognition of the blood donation environment, and previous blood donation experience—were significantly improved.

Table 4. Comparison of the survey results of the theory of planned behavior scale before and after intervention in the web-based intervention group.

Factor	Before, mean (SD)	After, mean (SD)	*t* test (*df*)	*P* value	Improved, mean (SD)	95% CI
Attitude	17.36 (2.338)	18.19 (2.234)	4.770 (278)	<.001	0.84 (2.292)	0.49-1.18
Subjective behavioral norms	14.9 (2.898)	16.22 (2.974)	5.940 (278)	<.001	1.32 (2.91)	0.88-1.76
Perceived behavioral control	51.82 (5.772)	54.44 (6.165)	5.858 (278)	<.001	2.62 (5.848)	1.74-3.50
Blood donation willingness	20.79 (3.114)	22.04 (3.168)	5.697 (278)	<.001	1.25 (2.859)	0.81-1.68
Outcome expectations	12.99 (1.71)	13.66 (1.535)	5.732 (278)	<.001	0.66 (1.507)	0.43-0.89
Self-identity	16.2 (3.274)	17.55 (6.138)	2.755 (278)	.007	1.35 (6.348)	0.38-2.31
Blood donation anxiety	12.46 (1.941)	13.27 (1.802)	6.234 (278)	<.001	0.82 (1.717)	0.56-1.08
Cognition of the blood donation environment	27.25 (4.464)	29.82 (4.138)	9.128 (278)	<.001	2.57 (3.686)	2.02-3.13
Previous blood donation experience	10.82 (2.646)	11.25 (2.855)	2.613 (278)	.01	0.22 (2.745)	0.2-0.63
Total	184.56 (21.124)	196.19 (21.986)	8.931 (278)	<.001	11.64 (17.039)	9.07-14.21

Comparison of Variables Before and After the Intervention for Blood Donors With Different Blood Donation Times in the Web-Based Intervention Group

The respondents in the web-based intervention group were further divided into groups based on the number of times they had donated blood in the past: none, once, twice, or ≥3 times. Statistically significant differences were observed between the groups in the factors blood donation anxiety, cognition of the blood donation environment, and previous blood donation experience. The willingness to donate, blood donation anxiety, and cognition of the blood donation environment improved the

most in the group with 2 donations, and the difference between the groups was statistically significant. In terms of cognitive improvement regarding the respondents' past blood donation experiences, the group with 1 donation showed greater improvement than the other groups, with a statistically significant difference between the groups (Table 5).

Table 5. Comparison of variable changes in blood donors with different blood donation times after the web-based intervention.

Factor	0 time, mean (SD)	1 time, mean (SD)	2 times, mean (SD)	≥3 times, mean (SD)	*F* test (*df*)	*P* value
Attitude	0.63 (2.018)	0.69 (2.303)	0.98 (2.201)	−0.40 (3.376)	1.673 (282)	.17
Subjective behavioral norms	1.21 (3.464)	1.04 (3.002)	1.63 (2.705)	0.20 (4.873)	1.358 (282)	.26
Perceived behavioral control	3.17 (5.239)	2.46 (5.203)	3.04 (4.800)	−1.07 (11.835)	2.484 (282)	.06
Blood donation willingness	0.33 (3.046)	0.89 (3.148)	1.57 (2.595)	−0.40 (4.469)	2.845 (282)	.04
Outcome expectations	0.42 (1.613)	0.45 (1.576)	0.71 (1.527)	0.27 (1.668)	0.822 (282)	.48
Self-identity	−0.67 (2.729)	0.00 (4.334)	0.01 (3.121)	0.40 (2.501)	0.315 (282)	.82
Blood donation anxiety	−0.46 (2.126)	0.66 (1.749)	1.00 (1.698)	−0.13 (1.685)	5.637 (282)	.001
Cognition of the blood donation environment	1.25 (3.904)	1.17 (3.657)	3.16 (3.626)	1.73 (4.284)	6.166 (282)	<.001
Previous blood donation experience	−1.54 (3.189)	0.34 (3.087)	−0.17 (2.857)	−0.33 (2.870)	2.796 (282)	.04
Total	4.33 (13.786)	7.7 (18.078)	11.93 (14.852)	0.27 (26.980)	3.210 (282)	.02

Postintervention Variable Comparison of Repeat and Nonrepeat Donors in the Web-Based Intervention Group

According to whether they donated blood again after the intervention in the subsequent 6-month period, the respondents in the web-based intervention group were divided into 2 groups, and the differences in the changes in the 9 variables were compared and analyzed. Table 6 reveals that the blood donors who chose to donate blood again after the intervention exhibited a greater improvement in the "outcome expectation" and "blood donation anxiety" variables; compared with nonrepeat donors, the difference was statistically significant.

Furthermore, with the blood donation result again as the dependent variable, age, gender, blood donation frequency, and the 9 intermediary variables were included as independent variables. A logistic stepwise regression indicated that improvements to the "outcome expectations" and "blood donation environment" factors can increase the possibility that nonregular blood donors will donate again, with statistical significance (Table 7).

Table 6. Analysis of the degree of change in variables among the repeat and nonrepeat donors in the web-based intervention group after the intervention.

Factor	Nonrepeat donors, mean (SD)	Repeat donors, mean (SD)	*t* test (*df*)	*P* value
Attitude	18.19 (2.096)	18.34 (2.854)	0.441 (283)	.66
Subjective behavioral norms	16.36 (2.971)	16.24 (3.226)	−0.25 (283)	.80
Perceived behavioral control	54.19 (5.543)	54.85 (7.430)	0.693 (283)	.49
Blood donation willingness	21.88 (3.184)	22.35 (3.466)	0.889 (283)	.38
Outcome expectations	13.45 (1.679)	14.04 (1.264)	2.739 (283)	.008
Self-identity	15.22 (3.391)	16.11 (3.295)	1.626 (283)	.11
Blood donation anxiety	13.06 (1.898)	13.70 (1.412)	2.603 (283)	.01
Cognition of the blood donation environment	30.18 (4.333)	29.52 (3.650)	−1.082 (283)	.28
Previous blood donation experience	13.32 (3.493)	14.26 (3.022)	1.701 (283)	.09
Total	195.86 (20.573)	199.41 (20.824)	1.068 (283)	.29

Table 7. Results of the logistic stepwise regression analysis of the web-based intervention group after the intervention.

	B	SE	Chi-square (df)	*P* value	Exp (B)	95% CI
Outcome expectations	0.560	0.174	10.3 (1)	.001	1.751	1.244-2.465
Blood donation environment	0.165	0.059	7.8 (1)	.005	0.848	0.754-0.952
Constant	−3.886	1.871	4.3 (1)	.04	0.021	N/A[a]

[a]N/A: not applicable.

Discussion

Principal Findings

This study aimed to not only address the possible influencing factors of blood donor losses after an initial blood donation but also discover theoretical intervention strategies given these factors to improve repeat donation rates. This study first provided data on interventions for nonregular blood donors based on the ETPB, which was important in providing reasons for continuous blood donation and future intervention directions; the results then indicated the effects of 2 different intervention methods: web-based videos and SMS text messages.

There are 3 factors that have a significant influence on repeated blood donation behavior: outcome expectations, blood donation anxiety, and blood donation environment. The first step in planning blood donation interventions involves having knowledge regarding the preventive factors in blood donation [43]. This study applied an ETPB to a prospective randomized controlled trial of an intervention in repeat blood donation behavior and received positive results. After the intervention, participants' attitudes, subjective behavioral norms, perceived behavioral control, blood donation willingness, self-identity, blood donation anxiety, outcome expectations, cognition of the blood donation environment, and cognition of the previous blood donation experience all significantly improved, with positive changes.

However, improvements in such perceptions as attitudes were not always reflected in the respondents' actions [44-46], and actual blood donors were far fewer than self-reported blood donors. Therefore, what factors have a significant impact on repeated blood donation behavior? This study further observed that 3 factors—outcome expectations, blood donation anxiety, and cognition of the blood donation environment–significantly differed between those who chose to donate blood again after the web-based video intervention and those who did not donate blood again after the intervention. Clearly, these 3 factors significantly impacted repeat blood donation behaviors.

First, outcome expectations can be divided into positive outcome expectations and negative outcome expectations. Similar to other studies, negative outcome expectations, such as anticipatory regret, have been shown to predict blood donation behavior [47-49]. Simultaneously, studies have demonstrated that in promoting healthy behaviors, the persuasion effect to avoid loss will be better than that to obtain gains [50]. This study adopted a negative outcome expectation, and 2 videos of negative outcome expectation were presented. One was about college students in a car accident who required substantial blood transfusions during surgery. The video indicated that if everyone actively donated blood, the blood supply would be sufficient to avoid any negative consequences, including amputations. The other one was about a Rh-negative mother in childbirth who urgently needed a transfusion. If everyone actively donated blood, an adequate supply of blood would ensure a smooth delivery, and the mother would avoid the negative consequences of stillbirth or infant death. These videos aroused donors' empathy, generated positive emotions, and psychologically matched the act of donating blood with the individual in need of help, prompting people to donate blood again. In support of the suggestion that the transfusion story videos should be promoted more in the future, these videos also helped people realize that repeatedly donating blood could avoid loss of life for the recipients because of insufficient blood supply; the viewers could avoid regret and be encouraged to donate blood again.

Second, this study verified that blood donation anxiety was an important factor affecting repeat donation behaviors; blood donation anxiety can prevent repeat donations. Other studies have shown that blood donation anxiety was critical in blood donors' decision to donate blood again [51]. The main reasons for not donating blood were concerns about safety and fear of donation [52,53]. Previous studies have revealed the fear of donating blood, needles in particular, and the belief that blood donation will adversely affect one's health are primary anxiety factors [54-56]. Hence, this study's short video of blood donation anxiety factors was aimed at explaining the above major anxiety factors in a straightforward manner and refuting the common fears and misunderstandings in donating blood. The video also details the entire donation process and the practices of blood donation testing, blood donation preparation, and delivery of the donated blood to the hospital, thereby reducing misunderstandings, alleviating blood donors' fears, and enhancing safety as well as confidence in the blood supply.

Third, the research discovered the connections between these environmental factors and blood donation behavior. A good blood donation environment may promote repeated blood donation behavior. Therefore, the possibility of irregular blood donors' repeat donations can be increased by improving the blood donation environment and providing a warm and comfortable blood donation environment, mitigating blood donation anxiety and strengthening outcome expectations. This is an important finding in research on repeat blood donors after expanding the theory of planned behavior in this study, and it offers significance and guidance for blood collection and supply institutions in implementing their own interventions for nonregular blood donors in subsequent steps.

In addition, the study also found that men aged 31 to 45 years and had donated blood twice in the past and irregular blood

donors who had donated twice in the past were more likely to donate blood again after the web-based intervention than after the SMS text message. This may be related to the fact that those who have donated blood twice have had a certain donation experience and a particular foundation in blood donation knowledge, including the process, experience, and perceptions. The group with ≥3 donations had the most experience in donating blood; with a similar "ceiling effect" [57], there was limited room for cognitive improvement. Therefore, those who had donated twice in the past were the most likely to become regular donors. Blood donors aged over 31 years generally had steady employment and were more mature. After receiving the relevant video interventions, they exhibited a higher action-based conversion rate after a cognitive change.

Web-based video interventions were effective. A major issue for blood donation workers involves the question of how to not only best convey information on coping with the obstacles to blood donation but also choose the best intervention method. This study combines the currently most effective web-based short video methods for dissemination with guidance from an ETPB to conduct an exploratory study of behavioral interventions. The study's results—specifically, that the web-based short video intervention method was more effective than SMS text messages for nonregular blood donors—were consistent with the research findings that video can effectively improve patients' knowledge, self-efficacy, satisfaction, and self-management levels in other areas, such as diabetes, heart disease, and patient family care [22-28,58].

However, this study differs from the findings of Karacaoğlu and Öncü [59]. Karacaoğlu and Öncü [59] began with first understanding new blood donors' fears and concerns and compared 6-minute educational videos with the brochures in use at that time; the videos addressed how to handle stress and anxiety among those experiencing the blood donation procedure for the first time. Considering the increase in knowledge and decrease in anxiety as outcome indicators after the intervention, the results revealed no difference between the brochure and video intervention groups. In the study by Masser et al [11], the video content was relatively simple, with only a video providing content from a precaution manual on the process before, during, and after blood donation. In contrast to these studies, this study first provided a short video with rich content, which was theoretically guided, driven by influencing factors, and provided on the web; measured the degree of psychological change from the intervention; and then tracked blood donation behavior rather than blood donation intention as an outcome variable, which more intuitively reflects the overall situation from the change of consciousness to the occurrence of behavior. The study also found that when those viewing the web-based video intervention chose to donate blood again, most of them uploaded photos of the time when they donated blood again to the WeChat group. This also played a role in donors' taking the initiative by example and encouraged undecided blood donors to donate blood again.

Conclusions

In conclusion, this study developed and verified an ETPB-driven web-based video intervention method to promote nonregular blood donors to donate again. The method addressed the individual donors' psychological and environmental factors, with remarkable results that can be popularized and applied in nonregular blood donor interventions to consequently improve repeat blood donation rates. Among the factors presented in this study, blood donation anxiety, result expectation, and improvements to the blood donation environment can positively impact repeat blood donation behaviors; hence, these are recommended as directions of focus in subsequent key interventions.

However, this study also has some limitations. First, this was an exploratory study, and its video sequence and frequency corresponding to the influencing factors were the first attempt and exploration, with no comparative study of other sequences. Second, the web-based intervention was based on the WeChat social media platform. Although group announcements and tasks were available to urge participants to click and watch the videos, no exact, effective means were used to understand more specific information, such as a particular viewing time. The respondents in the SMS group promised not to watch videos or other blood donation recruitment material during the study period, but these were limited by the respondents' self-awareness. As we could not discern whether they actually accessed intervention videos, errors may exist in that some respondents could have still accessed such videos. Third, this study used multiple comparisons, which may have caused a type 1 error. Fourth, this study has only been conducted for 6 months, and a longer follow-up period is needed for more comprehensive results. Fifth, the data collected in this study were Chinese, and the results may be different from those of other countries with different regions and cultures.

In subsequent research, we will continue to track these respondents' repeat donation behaviors after 1 and 1.5 years to further improve this study. Simultaneously, using the 9 variables discovered in this study—especially outcome expectations, blood donation anxiety, and cognition of the blood donation environment—specific improvement measures were designed and applied to a larger number of nonregular blood donors to observe the results of repeat blood donations. Further research will be conducted on the web-based video intervention method driven by the ETPB created in this study and focusing on factors such as video length, playback order, and sending frequency to further improve the web-based video intervention method.

Acknowledgments

This study was supported by Natural Science Foundation of Zhejiang Province (grant LY20G030003).

Authors' Contributions

All the authors contributed to the preparation and editing of the manuscript. QH and W Hu designed the study, collected the data, and drafted the manuscript. W Han and LP analyzed the data. QH, W Hu, W Han, and LP contributed to data interpretation and critical revisions of the manuscript. All authors approved the final version of the paper.

Conflicts of Interest

None declared.

Editorial notice: This randomized study was not registered because, as explained by the authors, it is observational in nature. The editor granted an exception from ICMJE rules mandating registration of randomized trials because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[DOC File , 197 KB - jmir_v24i12e37467_app1.doc]

References

1. France JL, France CR, Carlson BW, Kessler DA, Rebosa M, Shaz BH, et al. Motivating first-time, group O blood donors to return: rationale and design of a randomized controlled trial of a post-donation telephone interview. Contemp Clin Trials 2015 Sep;44:64-69 [FREE Full text] [doi: 10.1016/j.cct.2015.07.020] [Medline: 26247570]
2. Xianming C, Julin L, Ming L, Lifei L, Shibin L, Yanhua T. Effects of repeated blood donation on blood safety, Clinical transfusion and examination. Clin Transfusion Examination 2012 Jan;14(1):62-64. [doi: 10.3969/j.issn.1671-2587.2012.01.021]
3. Glynn SA, Kleinman SH, Schreiber GB, Busch MP, Wright DJ, Smith JW, et al. Trends in incidence and prevalence of major transfusion-transmissible viral infections in US blood donors, 1991 to 1996. Retrovirus Epidemiology Donor Study (REDS). JAMA 2000 Jul 12;284(2):229-235. [doi: 10.1001/jama.284.2.229] [Medline: 10889598]
4. Atsma F, Veldhuizen I, de Vegt F, Doggen C, de Kort W. Cardiovascular and demographic characteristics in whole blood and plasma donors: results from the Donor InSight study. Transfusion 2011 Feb;51(2):412-420. [doi: 10.1111/j.1537-2995.2010.02867.x] [Medline: 20804526]
5. Qiuyue H, Jian X, Xianwen L, Xianguo Q, Wei H. Analysis on the characteristics of demography and blood donation behavior of repeat and first-time blood donors——Based on the data of blood donors in Zhejiang province from 2006 to 2015. Chin J Blood Transfusion 2017;30(07):772-775. [doi: 10.13303/j.cjbt.issn.1004-549x.2017.07.042]
6. Xiaoling C. Outlining of the behavior characteristics of repeat blood donors in Fuzhou, China. Chinese J Blood Transfusion 2020;33(9):956-959. [doi: 10.13303/j.cjbt.issn.1004-549x.2020.09.026]
7. Shurong W, Yan C, Peng L. The intention to redonate among college first-time donors and its influencing factors. Chinese J Blood Transfusion 2020;33(10):1087-1090. [doi: 10.13303/j.cjbt.issn.1004-549x.2020.10.022]
8. Lingling P, Jian X, Xianwen L, Xianguo Q, Wei H. Analysis on the re-donate behavior of repeat blood donors in Zhejiang Province. Zhejiang Med 2018;40(19):2171-2174. [doi: 10.12056/j.issn.1006- 2785.2018.40.19.2017-1721]
9. Ugwu N, Uneke C, Ugwu C, Oti WJ, Agbo U, Akamike I. Effect of blood donor educational intervention on the knowledge and attitude towards voluntary blood donation among medical students at a Nigerian University. Niger Med J 2020;61(3):163-168 [FREE Full text] [doi: 10.4103/nmj.NMJ_177_19] [Medline: 33100469]
10. Francis KL, Polonsky MJ, Jones SC, Renzaho AM. The effects of a culturally-tailored campaign to increase blood donation knowledge, attitudes and intentions among African migrants in two Australian States: Victoria and South Australia. PLoS One 2017 Nov 30;12(11):e0188765 [FREE Full text] [doi: 10.1371/journal.pone.0188765] [Medline: 29190655]
11. Masser BM, France CR, Himawan LK, Hyde MK, Smith G. The impact of the context and recruitment materials on nondonors' willingness to donate blood. Transfusion 2016 Dec 25;56(12):2995-3003. [doi: 10.1111/trf.13805] [Medline: 27667318]
12. Godin G, Vézina-Im LA, Bélanger-Gravel A, Amireault S. Efficacy of interventions promoting blood donation: a systematic review. Transfus Med Rev 2012 Jul;26(3):224-37.e6. [doi: 10.1016/j.tmrv.2011.10.001] [Medline: 22126711]
13. France JL, France CR, Himawan LK. Re-donation intentions among experienced blood donors: does gender make a difference? Transfus Apher Sci 2008 Apr;38(2):159-166. [doi: 10.1016/j.transci.2008.01.001] [Medline: 18331813]
14. Robinson NG, Masser BM, White KM, Hyde MK, Terry DJ. Predicting intentions to donate blood among nondonors in Australia: an extended theory of planned behavior. Transfusion 2008 Dec;48(12):2559-2567. [doi: 10.1111/j.1537-2995.2008.01904.x] [Medline: 18717776]
15. Masser BM, White KM, Hyde MK, Terry DJ, Robinson NG. Predicting blood donation intentions and behavior among Australian blood donors: testing an extended theory of planned behavior model. Transfusion 2009 Feb;49(2):320-329. [doi: 10.1111/j.1537-2995.2008.01981.x] [Medline: 19040598]

16. Masser BM, Bednall TC, White KM, Terry D. Predicting the retention of first-time donors using an extended Theory of Planned Behavior. Transfusion 2012 Jun;52(6):1303-1310. [doi: 10.1111/j.1537-2995.2011.03479.x] [Medline: 22257135]
17. Veldhuizen I, Ferguson E, de Kort W, Donders R, Atsma F. Exploring the dynamics of the theory of planned behavior in the context of blood donation: does donation experience make a difference? Transfusion 2011 Nov;51(11):2425-2437. [doi: 10.1111/j.1537-2995.2011.03165.x] [Medline: 21569046]
18. Yihan Z, Qinjian Y. The theory of planned behavior and its application and prospect in information system research. J Modern Inform 2019;39(12):138-177. [doi: 10.3969/j.issn.1008-0821.2019.12.016]
19. Guohua Z, Wangli W, Xuezheng Q. Research progress on the theory and mechanism of health behavior intervention. J Modern Inform 2021;35(8):1428-1434. [doi: 10.12102/j.issn.1009-6493.2021.08.020]
20. Ajzen I. The theory of planned behavior. Organizational Behav Human Decision Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
21. Statistical report on internet development in China. China Internet Network Information Center. 2020 Sep. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/202012/P020201201530023411644.pdf [accessed 2021-02-03]
22. Li XQ, Hu GH, Shen WY, Yang HM, Zhao QH. Application of dialect video education combined with teach-back method in rehabilitation of elderly patients after total hip arthroplasty. J Nurs 2019;26(24):53-56 [FREE Full text] [doi: 10.16460/j.issn1008-9969.2019.24.053]
23. Hong W. Effect of modular rehabilitation training combined with video health education on the recovery of lower limb function in patients with tibial plateau fractures. Int Med Health Guidance News 2020;26(14):2143-2145. [doi: 10.3760/cma.j.issn.1007-1245.2020.14.045]
24. Rui-qing D, Xing-dan L. The effect of a short video based on internet platform on swallowing function of patients with nasopharyngeal carcinoma by radiotherapy. Chin J Nurs Educ 2020;17(1):84-88. [doi: 10.3761/j.issn.1672-9234.2020.01.018]
25. Müller T, Müller A, Hübel C, Knipel V, Windisch W, Cornelissen CG, et al. Optimizing inhalation technique using web-based videos in obstructive lung diseases. Respir Med 2017 Aug;129:140-144 [FREE Full text] [doi: 10.1016/j.rmed.2017.06.009] [Medline: 28732821]
26. Zhike Y. Effects of micro-video combined with scenario simulation training on safe eating at home for stroke patients with dysphagia. Chinese Nursing Res 2020;34(5):908-911. [doi: 10.12102/j.issn.1009-6493.2020.05.039]
27. Maciel RR, Chiavegato LD, Marin LS, Portella DD, de Souza MC, Camelier FW, et al. Development of an e-health education program at the workplace using formative research - technologies for improving quality of life. Eval Program Plan 2019 Apr;73:129-137. [doi: 10.1016/j.evalprogplan.2018.12.009] [Medline: 30622061]
28. Tuong W, Larsen ER, Armstrong AW. Videos to influence: a systematic review of effectiveness of video-based education in modifying health behaviors. J Behav Med 2014 Apr 28;37(2):218-233. [doi: 10.1007/s10865-012-9480-7] [Medline: 23188480]
29. Qiuyue H, Wenjuan H, Li Z, Yiping W, Yechen J, Yinfang L, et al. Delphi method was applied to construct the evaluation index system of voluntary blood donors' intention to repeat blood donation. Preventive Med 2021;33(8):841-843. [doi: 10.19485/j.cnki.issn2096-5087.2021.08.023]
30. Qiuyue H, Li Z, Wenjuan H. Reliability and validity analysis of repeated blood donation influencing factors and behavioral intention Assessment Scale—Based on expand theory of planned behavior. Zhejiang Medical J 2021;43(11):1196-1200. [doi: 10.12056/j.issn.1006-2785.2021.43.11.2021-600]
31. Veldhuizen IJ. Blood donor profiling using donation patterns. ISBT Sci Series 2013 May 31;8(1):233-237. [doi: 10.1111/voxs.12017]
32. Veldhuizen I, Folléa G, de Kort W. Donor cycle and donor segmentation: new tools for improving blood donor management. Vox Sang 2013 Jul 26;105(1):28-37. [doi: 10.1111/vox.12018] [Medline: 23438134]
33. Godin G, Sheeran P, Conner M, Germain M. Asking questions changes behavior: mere measurement effects on frequency of blood donation. Health Psychol 2008 Mar;27(2):179-184. [doi: 10.1037/0278-6133.27.2.179] [Medline: 18377136]
34. Jinting W, Yuanyuan W, Yanyan L, Changrong Y. Application and prospect of smart phone application in health management of patients with chronic diseases. Chinese J Nursing 2014;49(8):994-997. [doi: 10.3761/j.issn.0254-1769.2014.08.025]
35. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
36. 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]
37. Hao B, Jingjing H. Research on the application of wechat public platform in the field of higher education. Educ Inform China 2013;2:78-81. [doi: 10.3969/j.issn.1673-0410.2014.08.222]
38. Yuan C. Application of wechat in nursing work. Digital Technol Applications 2014;2:45. [doi: 10.19695/j.cnki.cn12-1369.2014.02.031]
39. Chongming H, Feiwei M. Study on the practice of public micro message platform of a hospital in Zhejiang. Med Soc 2014;27(3):46-48. [doi: 10.3870/YXYSH.2014.03.015]

40. Juan W, Ye Y, Jinjing Y. Wechat intervention research on postpartum PFMEs compliance of natural delivery women. Primary Care Forum 2017;21(27). [doi: 10.19435/j.1672-1721.2017.27.027]
41. 42 YJ, Mingna Z. Influence of wechat platform of information health education on cognition, attitude and behavior of maternal pelvic floor function exercise. Chinese J Practical Nursing 2019;35(10):758-763. [doi: 10.3760/cma.j.issn.1672-7088.2019.10.008]
42. Denny MC, Vahidy F, Vu KY, Sharrief AZ, Savitz SI. Video-based educational intervention associated with improved stroke literacy, self-efficacy, and patient satisfaction. PLoS ONE 2017 Mar 23;12(3):e0171952. [doi: 10.1371/journal.pone.0171952] [Medline: 28333925]
43. Chell K, Waller D, Masser B. The Blood Donor Anxiety Scale: a six-item state anxiety measure based on the Spielberger State-Trait Anxiety Inventory. Transfusion 2016 Jun;56(6 Pt 2):1645-1653. [doi: 10.1111/trf.13520] [Medline: 27021408]
44. Özgür S, Ürek H, Kösal K. Turkish university students' opinions towards blood donation. Univ J Educ Res 2018 May;6(5):897-908. [doi: 10.13189/ujer.2018.060511]
45. Raghuwanshi B, Pehlajani NK, Sinha MK. Voluntary blood donation among students - a cross-sectional study on knowledge and practice vs. attitude. J Clin Diagn Res 2016 Oct;10(10):EC18-EC22 [FREE Full text] [doi: 10.7860/JCDR/2016/21957.8733] [Medline: 27891345]
46. Charles KS, Chisholm K, Gabourel K, Philip K, Ramdath S, Abdul-Hakeem H, et al. A follow-up survey of knowledge, attitudes and practices surrounding blood donation in Trinidad and Tobago. ISBT Sci Series 2017 May 25;12(3):349-356. [doi: 10.1111/voxs.12358]
47. Godin G, Conner M, Sheeran P, Bélanger-Gravel A, Germain M. Determinants of repeated blood donation among new and experienced blood donors. Transfusion 2007 Sep;47(9):1607-1615. [doi: 10.1111/j.1537-2995.2007.01331.x] [Medline: 17725724]
48. Conner M, Godin G, Sheeran P, Germain M. Some feelings are more important: cognitive attitudes, affective attitudes, anticipated affect, and blood donation. Health Psychol 2013 Mar;32(3):264-272. [doi: 10.1037/a0028500] [Medline: 22612559]
49. Faqah A, Moiz B, Shahid F, Ibrahim M, Raheem A. Assessment of blood donation intention among medical students in Pakistan--An application of theory of planned behavior. Transfus Apher Sci 2015 Dec;53(3):353-359. [doi: 10.1016/j.transci.2015.07.003] [Medline: 26254256]
50. Chou EY, Murnighan JK. Life or death decisions: framing the call for help. PLoS One 2013 Mar 6;8(3):e57351 [FREE Full text] [doi: 10.1371/journal.pone.0057351] [Medline: 23483903]
51. Ferguson E, France CR, Abraham C, Ditto B, Sheeran P. Improving blood donor recruitment and retention: integrating theoretical advances from social and behavioral science research agendas. Transfusion 2007 Nov;47(11):1999-2010. [doi: 10.1111/j.1537-2995.2007.01423.x] [Medline: 17958528]
52. Zucoloto ML, Gonçalez T, Menezes NP, McFarland W, Custer B, Martinez EZ. Fear of blood, injections and fainting as barriers to blood donation in Brazil. Vox Sang 2019 Jan 28;114(1):38-46. [doi: 10.1111/vox.12728] [Medline: 30485453]
53. Delépine-Farvacques S, Terrade F, Danic B, Lebaudy JP. [Commitment in plasmapheresis donation: a study of determinants among regular donors in plasmapheresis]. Transfus Clin Biol 2017 Jun;24(2):38-46. [doi: 10.1016/j.tracli.2017.02.003] [Medline: 28336244]
54. Nilsson Sojka B, Sojka P. The blood-donation experience: perceived physical, psychological and social impact of blood donation on the donor. Vox Sang 2003 Feb;84(2):120-128. [doi: 10.1046/j.1423-0410.2003.00271.x] [Medline: 12609018]
55. Tariq S, Tariq S, Jawed S, Tariq S. Knowledge and attitude of blood donation among female medical students in Faisalabad. J Pak Med Assoc 2018 Jan;68(1):65-70 [FREE Full text] [Medline: 29371721]
56. Argan MT, Argan M. Prospectives on attitudes, thinking styles and fears towards blood donation: a research on Turkish sample. Pazarlama Teorisi ve Uygulamaları Dergisi 2017;3:56.
57. France CR, France JL, Wissel ME, Kowalsky JM, Bolinger EM, Huckins JL. Enhancing blood donation intentions using multimedia donor education materials. Transfusion 2011 Aug;51(8):1796-1801. [doi: 10.1111/j.1537-2995.2010.03033.x] [Medline: 21303366]
58. Steffen AM, Gant JR. A telehealth behavioral coaching intervention for neurocognitive disorder family carers. Int J Geriatr Psychiatry 2016 Feb 15;31(2):195-203 [FREE Full text] [doi: 10.1002/gps.4312] [Medline: 26077904]
59. Karacaoğlu Y, Öncü E. The effect of the video and brochure via donor recruitment on fear, anxiety and intention: randomized controlled trial. Transfus Apher Sci 2020 Apr;59(2):102698. [doi: 10.1016/j.transci.2019.102698] [Medline: 31980334]

Abbreviations

ETPB: extended theory of planned behavior

Edited by G Eysenbach; submitted 24.02.22; peer-reviewed by S Quinn, L Guo; comments to author 22.06.22; revised version received 13.08.22; accepted 15.11.22; published 23.12.22.

Please cite as:
Hu Q, Hu W, Han W, Pan L
Web-Based Short Video Intervention and Short Message Comparison of Repeat Blood Donation Behavior Based on an Extended Theory of Planned Behavior: Prospective Randomized Controlled Trial Study
J Med Internet Res 2022;24(12):e37467
URL: https://www.jmir.org/2022/12/e37467
doi:10.2196/37467
PMID:36379691

Original Paper

Using Deep Transfer Learning to Detect Hyperkalemia From Ambulatory Electrocardiogram Monitors in Intensive Care Units: Personalized Medicine Approach

I-Min Chiu[1,2], MD; Jhu-Yin Cheng[2*], MD; Tien-Yu Chen[3], MD; Yi-Min Wang[2], MD; Chi-Yung Cheng[1,2], MD; Chia-Te Kung[2], MD; Fu-Jen Cheng[2], MD, PhD; Fei-Fei Flora Yau[2], MD; Chun-Hung Richard Lin[1*], PhD, Prof Dr

[1]Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung City, Taiwan

[2]Department of Emergency Medicine, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung City, Taiwan

[3]Division of Cardiology, Department of Internal Medicine, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung City, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chun-Hung Richard Lin, PhD, Prof Dr
Department of Computer Science and Engineering
National Sun Yat-sen University
No 70, Lienhai Rd
Kaohsiung City, 804
Taiwan
Phone: 886 07 5252000 ext 4301
Email: lin@cse.nsysu.edu.tw

Abstract

Background: Hyperkalemia is a critical condition, especially in intensive care units. So far, there have been no accurate and noninvasive methods for recognizing hyperkalemia events on ambulatory electrocardiogram monitors.

Objective: This study aimed to improve the accuracy of hyperkalemia predictions from ambulatory electrocardiogram (ECG) monitors using a personalized transfer learning method; this would be done by training a generic model and refining it with personal data.

Methods: This retrospective cohort study used open source data from the Waveform Database Matched Subset of the Medical Information Mart From Intensive Care III (MIMIC-III). We included patients with multiple serum potassium test results and matched ECG data from the MIMIC-III database. A 1D convolutional neural network–based deep learning model was first developed to predict hyperkalemia in a generic population. Once the model achieved a state-of-the-art performance, it was used in an active transfer learning process to perform patient-adaptive heartbeat classification tasks.

Results: The results show that by acquiring data from each new patient, the personalized model can improve the accuracy of hyperkalemia detection significantly, from an average of 0.604 (SD 0.211) to 0.980 (SD 0.078), when compared with the generic model. Moreover, the area under the receiver operating characteristic curve level improved from 0.729 (SD 0.240) to 0.945 (SD 0.094).

Conclusions: By using the deep transfer learning method, we were able to build a clinical standard model for hyperkalemia detection using ambulatory ECG monitors. These findings could potentially be extended to applications that continuously monitor one's ECGs for early alerts of hyperkalemia and help avoid unnecessary blood tests.

(J Med Internet Res 2022;24(12):e41163) doi:10.2196/41163

KEYWORDS

deep learning; transfer learning; hyperkalemia; electrocardiogram; ECG monitor; ICU; personalized medicine

Introduction

Hyperkalemia is a metabolic condition that contributes to more than 800,000 emergency department visits in the United States annually [1]. It is associated with life-threatening ventricular arrhythmias and sudden cardiac arrest, and it is especially common among patients with chronic kidney disease due to

their impaired renal potassium homeostasis and long-term use of renin-angiotensin-aldosterone system inhibitors [2,3].

Patients under critical care may receive regular blood tests for electrolytes every few hours or days [4]. Many potential factors can affect potassium levels in between monitoring periods, such as diets, metabolic acidosis, and alterations in the intracellular/extracellular potassium distribution. Therefore, noninvasive monitoring techniques of potassium levels can help fill the gap between blood tests for early detection of this potentially deadly condition.

It is well known that a variety of changes on the electrocardiogram (ECG) can be associated with hyperkalemia, including but not limited to peaked T waves, shortened QT interval, lengthening of PR interval, and QRS duration [5]. Accurate human interpretations of these ECG patterns requires a steep learning curve, and the sensitivity of physician diagnoses has been estimated to be as low as 34% to 43% [6], not to mention the impossibility of self-detection of hyperkalemia using only symptoms and signs. There have been several successes in leveraging deep learning models to detect electrolyte abnormalities on ECGs [7-10], and previous studies have proven the feasibility of this approach for detecting subtle signals from ECGs. However, low specificity and a high false-positive rate could cause alert fatigue among physicians and anxiety in patients.

A recent study exploring a personalized deep learning–based system to detect hypoglycemia via ECG data has yielded promising results [11]. The study collected dozens of personal blood glucose and corresponding ECG data and adopted a convolution neural network to develop a personalized deep learning model to predict the hypoglycemia event. Since it is nearly impossible to gather enough personal data in a real-world setting, our study proposes a personalized transfer learning method by first training a general model and then refining it with personal data to improve the accuracy of hyperkalemia predictions, diminish the intersubject heterogeneity, and advance toward personalized medicines.

Methods

Ethics Approval

The data collection and study protocols were approved by the Institutional Review Board of Chang Gung Medical Foundation (202001217B0; date of approval July 21, 2020). The study was conducted following the standards issued by the World Medical Association's Declaration of Helsinki. The data that support the findings of this study are openly available in the Medical Information Mart From Intensive Care III (MIMIC-III) Waveform Database Matched Subset [12,13].

Data Set Collection

This study used data from the Waveform Database Matched Subset. The data set contains 22,317 waveform records and 22,247 numeric records for 10,282 distinct intensive care unit (ICU) patients who were admitted to critical care units of medical centers in the United States between 2001 and 2012. These recordings typically include digitized signals, such as ECG, arterial blood pressure, and respiration; additionally, they include periodic measurements such as heart rate, oxygen saturation, and blood pressure. The data set's ECG signals were usually of leads I, II, or V. This subset represents records for which the patients have been identified and whose corresponding clinical records are available in its matched clinical database.

Patient Population

All patients with a plasma potassium level during admission, from the MIMIC-III data set, were included. However, patients without lead II ECG signals at the time of the potassium level test and patients with atrial fibrillation, pacing rhythm, or other medical conditions for which a complete heartbeat cycle could not be distinctly identified in the ECG were excluded.

This study aimed to distinguish hyperkalemia from a normal level based on ECG features. Patients with at least 8 records of hyperkalemia and normokalemia each were adopted for personalized transfer learning. The others were selected for generic model training.

Data Preprocessing

ECG excerpts from 10 minutes before the time of serum potassium tests were annotated as corresponding to hyperkalemia or normokalemia according to the test results. Hyperkalemia was defined as serum potassium concentration values above 5.5 mEq/L and normokalemia as serum concentration between 3.5 mEq/L and 5 mEq/L. We excluded serum potassium levels between 5 and 5.5 mEq/L to ensure that no consecutive heartbeats would be considered as both hyper and normal, therefore reducing overfitting of the model.

After collection, ECG excerpts were filtered using finite impulse response techniques and underwent manual inspection to exclude those containing too much ECG signal noise. This process helps to reduce overfitting of the model and deviating to noisy data.

After retrieving the corresponding ECG signals, each ECG record was segmented into heartbeats of 120 samplings based on the fiducial point, which was the R peak. Each heartbeat segment contained 40 samples preceding the R peak and 80 samples after, in which the R peak was the 41st sample.

Generic Model Training

The goal was to train a generic model using a large set of heartbeats and to leverage a transfer learning algorithm that could refine the generalized model into a subject-specific model. The purpose of training the deep learning model during this step was to obtain a pretrained weight for transfer learning, as it helps the model to learn the shape of the ECG features for hyperkalemia.

All the data were randomly split into a training set, validation set, and test data set in a 6:2:2 ratio. This study used the ResNet architecture as the baseline architecture [14]. ResNet stands for residual network; it is an innovative neural network first introduced in 2015 that won the top position at the ImageNet classification competition, with an error of only 3.57%. The ResNet structure is widely used in ECG classification tasks since the residual block component allows the model to add more layers that help to detect the complex pattern of ECG morphology. In 2019, a study demonstrated a cardiologist-level

arrhythmia detection task in ambulatory ECGs using a 1D ResNet model [15]. Since ECG data in the MIMIC-III database is 1D, we substituted the 2D convolution layers with 1D for detailed feature extraction. Besides that, we did not change the size of the convolutional block, stride, and number of filters. After training, model weights for the best performance in the validation set were saved as pretrained for personal transfer learning.

Personalized Transfer Learning

Transfer learning applies knowledge obtained by solving one problem to a related problem. The general procedure for transfer learning is to pretrain a deep learning model with a large data set (eg, generic population for hyperkalemia), then refine it using a much smaller target data set (eg, personal data for hyperkalemia).

In personalized training steps, we adopted the same architecture of the 1D ResNet model in the general population for allowing the weight to be preserved [16]. Before refining it with personal data, we replaced its classification layer with a fully connected layer, the weights of which were randomly initialized. We froze the pretrained weights in the first few blocks of the convolution layers and trained the last few blocks of layers for 5 epochs using personal ECG data.

For each patient in the personalized group, 25% of potassium records and their corresponding ECG data were preserved as the test data set. During the training process, ECG data representing one record each of hyperkalemia and normokalemia were used as inputs. The training process continued for five rounds at most, depending on the total number of potassium records for each patient. The model performance after each training process was assessed to measure the performance changes. The deep learning models were trained with the TensorFlow application programming interface on the Google Colab platform.

Statistical Analysis

Continuous variables with normal distribution were presented either as means (SDs) or medians (IQRs). Continuous variables were analyzed using the Mann-Whitney *U* test, and the final model was validated using a majority voting scheme that runs through all the single heartbeats in a 10-minute ECG strip to determine the prediction result. All performance predictions were assessed using accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, and specificity. All statistical analyses were performed on SPSS 26 for Mac (IBM Corp).

Results

Characteristics of Data Sets

In this study, of the 41,291 patients in the MIMIC-III database, 5385 who fulfilled the criteria were included for analysis; 16 were chosen for personalized model development and validation and 5369 for pretrained general model development. To avoid deviation of general model prediction toward normokalemia, balanced ECG samples of hyperkalemia and normokalemia retrieved from 1439 patients were used. These included 1341 hyperkalemia records of 721 patients and 1325 normokalemia records of 718 patients. The inclusion flowchart is shown in Figure 1. Demographics and clinical characteristics of the two development populations are shown in Table 1. For the personal model development, patients' median age was 50 (IQR 43-60) years, and 13 (81%) of them were male. Concerning ethnic groups, 7 (44%) patients and 3 (19%) patients were White and African American, respectively. The mean serum potassium levels of hyperkalemia and normokalemia were 6.3 (SD 0.64) mEq/L and 4.3 (SD0.40) mEq/L and 6.2 (SD 0.70) mEq/L and 4.1 (SD 0.44) mEq/L in the general and personalized groups, respectively.

Figure 1. Patient inclusion flowchart. ECG: electrocardiogram; MIMIC-III: Medical Information Mart From Intensive Care III.

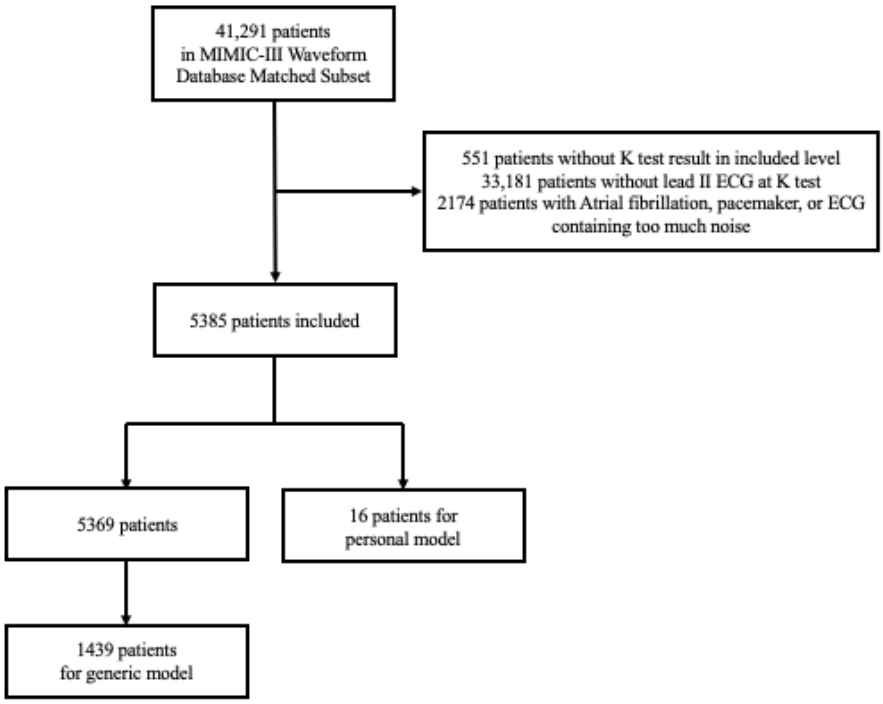


Table 1. Patient demographic in generic population and personal population.

Variables	Generic population (n=1439)	Personal population (n=16)
Age (years), median (IQR)	64 (52-76)	50 (43-60)
Gender, n (%)		
Male	610 (59.6)	13 (81.2)
Female	413 (40.4)	3 (18.8)
Ethnicity, n (%)		
White	907 (63.0)	7 (43.8)
African American	176 (12.2)	3 (18.8)
Hispanic	71 (4.6)	0 (0.0)
Asian	45 (3.1)	1 (6.2)
American Indian	6 (0.4)	0 (0.0)
Other	50 (3.5)	1 (6.2)
Unknown	184 (12.8)	4 (25.0)
Body index, median (IQR)		
Height (cm)	170.0 (160.2-178.0)	176.5 (169.2-186.8)
Weight (kg)	79.2 (66.9-94.6)	89.2 (71.4-108.3)
Serum level (mEq/L), mean (SD)		
Normokalemia	4.3 (0.40)	4.1 (0.44)
Hyperkalemia	6.3 (0.64)	6.2 (0.70)

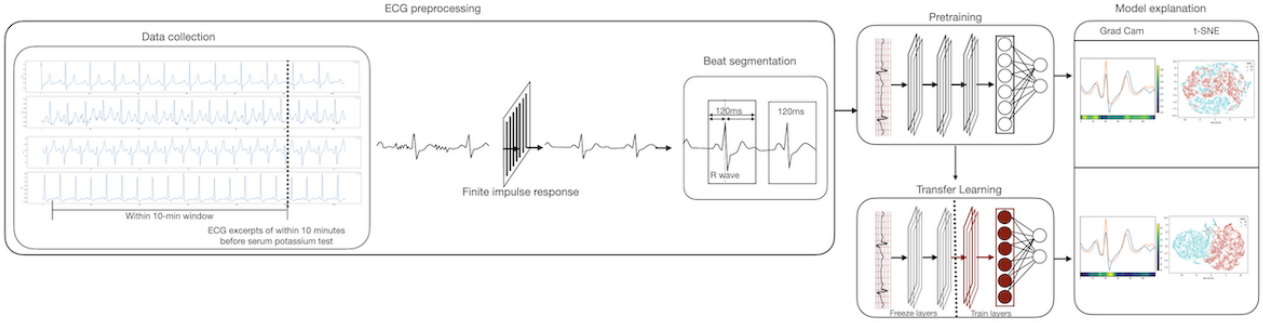
Development of a Generic Model

The proposed transfer learning method is depicted in Figure 2. For generic model development, in the training set, 152,322 and 172,388 normokalemic and hyperkalemic heartbeats, respectively, were contributed by 881 patients. In the validation set, 51,468 normokalemic and 53,488 hyperkalemic heartbeats were segmented from 280 patients. In the test set, 52,134 normokalemic and 53,412 hyperkalemic heartbeats were present in 278 patients.

This study adopted the ResNet-50 model for training the classification task. ResNet-50 is a network that has 50 layers in depth. Rather than its shallow version, ResNet-18 or

ResNet-34, ResNet-50 combines the structure of residual and bottleneck blocks in the convolutional layer to reduce computing resources. Before each convolutional layer, we applied batch normalization and a rectified linear activation, adopting the original design of the preactivation block. The network was trained with random initialization of the weights. We used the Adam optimizer with the default parameters and a mini-batch size of 1024. We initialized the learning rate to 5×10^{-5} and reduced it by a factor of 2 when the developmentally set loss stopped improving for 3 consecutive epochs. The model was trained for 50 epochs using the training set. We saved the model that achieved minimal loss in the validation set during training as the generic, or the so-called pretrained, model. The final pretrained model's prediction accuracy was 0.724, 0.639, and 0.627 in the training, validation, and test sets, respectively.

Figure 2. Proposed transfer learning algorithm for recognizing hyperkalemia from ambulatory ECG monitoring. ECG: electrocardiogram; Grad Cam: gradient-weighted class activation mapping; t-SNE: t-distributed stochastic neighbor embedding.

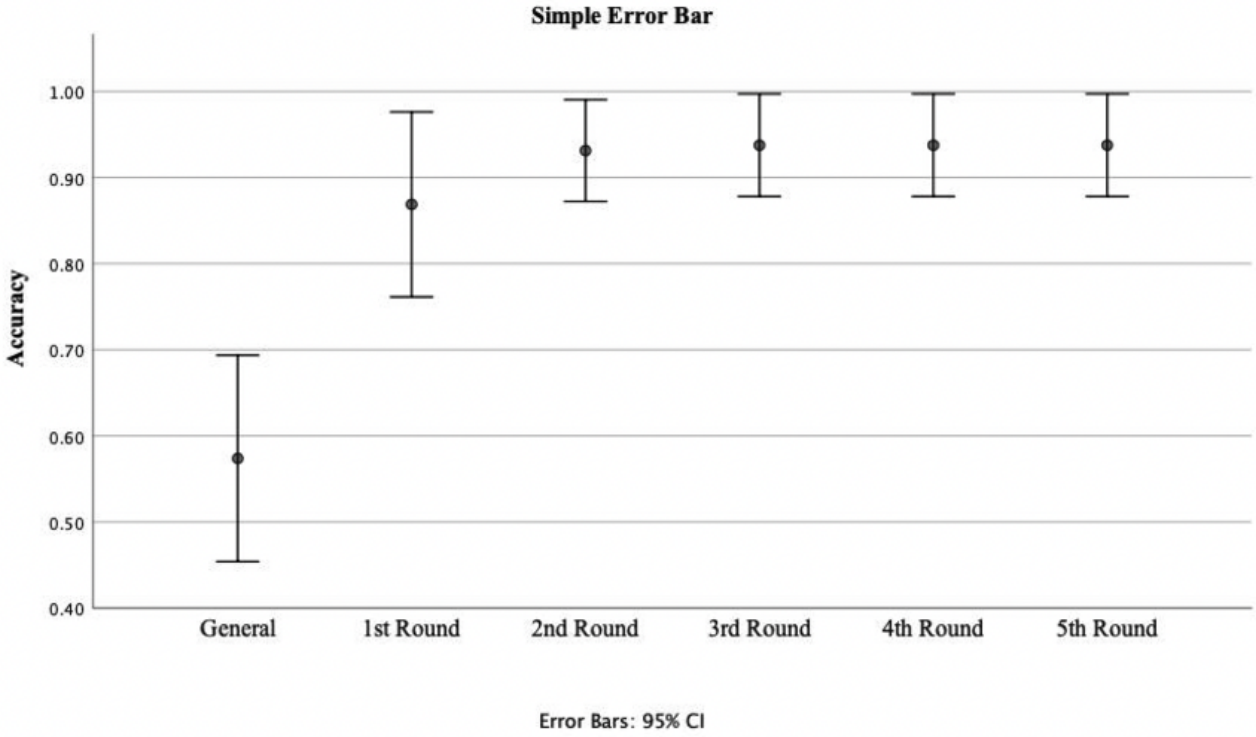


Development of a Personalized Model

The complete list of patients included for personalized model development is shown in Multimedia Appendix 1. Before starting the personalized training process, the performance of the pretrained model was assessed on each patient to obtain baseline metrics. The same convolutional neural network (CNN) architecture as ResNet-50 was used, the pretrained weight in the first several layers were frozen, and the model was fine-tuned in each training round. The improvement in predictions after each training round for all personalized group patients are shown in Multimedia Appendix 2.

On average, accuracy improved from 0.604 (SD 0.211) to 0.895 (SD 0.189; $P<.001$) after the first round of training (Figure 3) and achieved a plateau at 0.942 (SD 0.104) after the second round of training. In addition, the AUC level improved from an average of 0.729 (SD 0.240) to 0.918 (SD 0.149) after the first round of training and continued to increase after the second round to 0.945 (SD 0.094), being maintained at that level thereafter. After five rounds of training, the personalized model was able to predict hyperkalemia with an average accuracy of 0.980 (SD 0.078). Moreover, the sensitivity and specificity of model prediction improved after personalized transfer learning. There was a significant increase in average sensitivity after the second round of personalized transfer learning (0.674, SD 0.456 vs 0.953, SD 0.160; $P=.03$) and improvement in average specificity after the first round of personalized transfer learning (0.628, SD 0.417 vs 0.907, SD 0.321; $P=.03$; Multimedia Appendix 2).

Figure 3. Accuracy improvement on number of personalized training rounds.



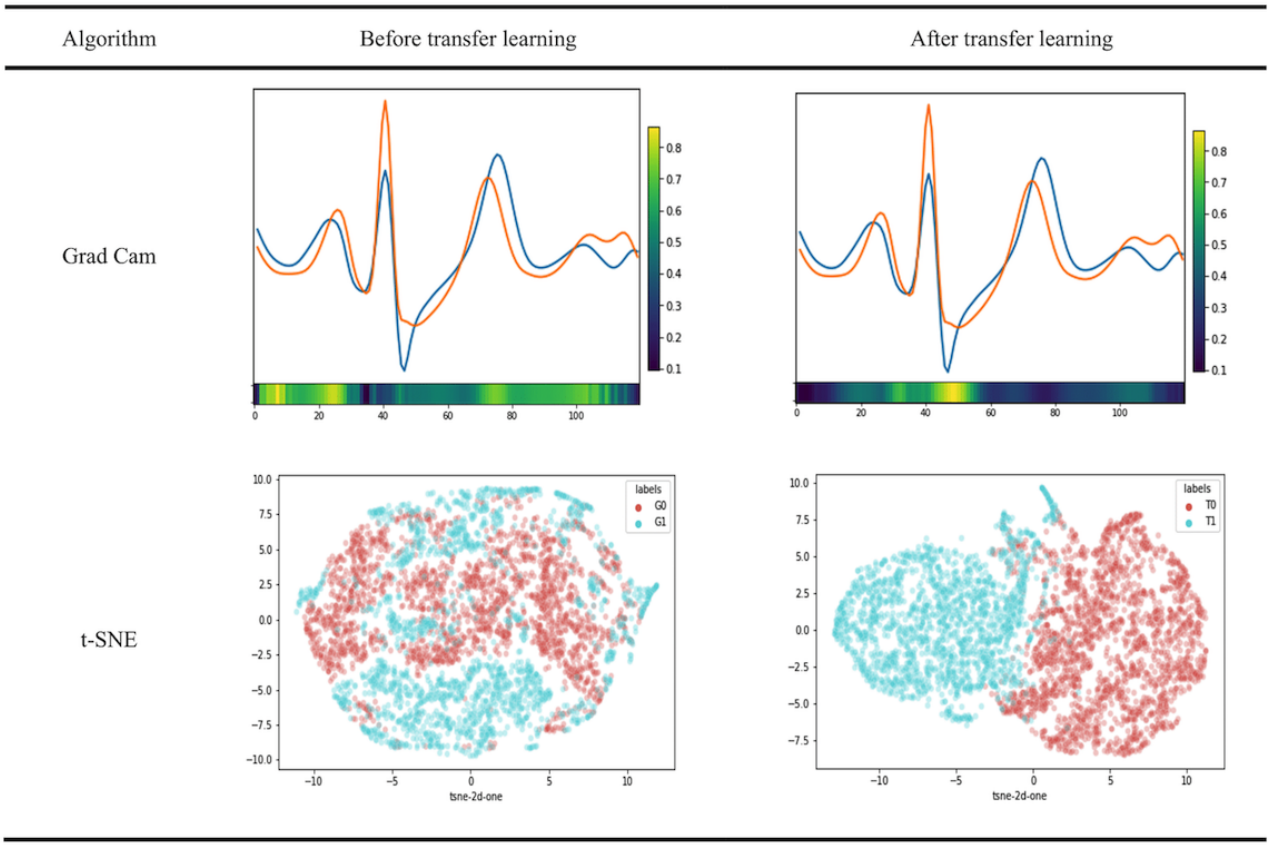
Model Interpretation and Visualization

After refining the model for each candidate, we generated the gradient-weighted class activation mapping (Grad Cam) for obtaining visual explanations from the model [17]. We selected one patient for demonstration (Figure 4). The blue and orange lines represented the average plot of segmented heartbeats from hyperkalemia and normokalemia in the test set of that patient, respectively. Grad Cam uses the gradients of the final convolutional layer to produce a coarse localization map, highlighting important regions in the image for predicting the concept. We overlapped the activation maps on the original ECG reading to highlight the emphasized areas in the CNN. In this patient, the region of interest was dispersed before transfer learning and thereafter became more focused over the QRS segment.

To further address the effect of personalized transfer learning and to visualize the learned embeddings [18], we used the t-distributed stochastic neighbor embedding (t-SNE) method that extracted features from the last convolutional layer and converted these high-dimension features to 2D features, which we could analyze using the scatter plot.

This study presents a visualization of the same patient's heartbeats using a t-SNE scatter plot and Grad Cam (Figure 4). In the figure, the light blue dot represents the hyperkalemia ECG data and the red dot the normokalemia data. It can be observed that the dots presenting each heartbeat corresponding to the hyperkalemia and normokalemia were quite muddled at the beginning and gradually organized into two clusters that were easier to separate by a straight line in a 2D space after personalized transfer learning. The t-SNE plots imply that after transfer learning, it is easier to distinguish the two classes clearly through the CNN model.

Figure 4. Visualization of model prediction before and after personalized transfer learning. Grad Cam: gradient-weighted class activation mapping; t-SNE: t-distributed stochastic neighbor embedding.



Discussion

Principal Findings

Our study showed the feasibility of detecting hyperkalemia by using transfer learning on personalized single-lead ECG readings. In previous research, one of the biggest challenges of deep learning for ECG classification was the scalability of a single model in different populations. The reasons for this generalization problem included individually varying ECG signal properties that depended on various factors such as weight, height, age, and physical conditions [19], not to mention data collected from different institutions and by other technicians. Therefore, expecting a generalized framework to be functional for the general population can be problematic [20]. The novelty of this study is that it considers the same classification task in the generic and personal predictions as different domains, which is a novel area in dyskalemia prediction using ECGs. Since other medical conditions may affect patients' ECG, the ECG signal of hyperkalemia is not specific. Patients' different medical diseases contribute to how their ECGs manifest. Due to large interpersonal variation, we believe that one generic model cannot cover a variety of illnesses that affect ECG signals. By using personalized transfer learning, we can load personal information into the model and make it better fit the personal ECG change.

In our study, when comparing the results with our pretrained model, transfer learning on personalized data showed a considerable increase in accuracy and AUC, and significantly boosted both sensitivity (mean 0.674, SD 0.456 vs mean 0.953, SD 0.160; *P*=.03) and specificity (mean 0.628, SD 0.417 vs mean 0.907, SD 0.321; *P*=.03), which demonstrated a good precision by ambulatory ECG monitors.

Previous studies applying deep learning to ECG classification focused on arrhythmia detection [21-23]. One of the reasons may be the availability of benchmark data sets that have reliable annotations that are mostly limited to arrhythmia, such as atrial fibrillation. Within the deep learning arrhythmia detection domain, transfer learning had been extensively explored to enhance the performance of CNNs; in a study, Weimann and Conrad [24] showed that transfer learning effectively reduces the number of annotations required to achieve the same performance as CNNs that are not pretrained. However, apart from arrhythmias, many other metabolic conditions such as hyperkalemia, hypokalemia, and hypomagnesemia can also be manifested in ECG readings [25]; hence, successes in arrhythmia detection should be expandable for detection of metabolic diseases through ECGs. In 2019, a study by Galloway et al [10] used 2 to 4 leads of the 12 leads of an ECG to develop a deep learning model to predict hyperkalemia and demonstrated a sensitivity of 88.9%-91.3% and a specificity of 54.7%-63.2%. The study proved that screening for hyperkalemia in patients with chronic kidney disease was feasible. A recent study using all 12 leads from complete ECGs to predict both hyperkalemia and hypokalemia showed better results with a balanced accuracy of around 79.9%-82.8% [26]. Nevertheless, a higher standard of prediction performance may be desired to aid in clinical practice.

Personalized medicine, mostly discussed in genomic research, refers to the application of specific patient information to make a more informed choice regarding their optimal therapeutic treatments or precise diagnoses, rather than relying on population-based trends [27]. Recent studies have shown that a personalized medicine approach could benefit disease diagnosis with ECGs. A 2018 study proved that by acquiring about 5% of personal ECG data from each new patient, the personalized deep learning model was able to substantially improve the precision of disease detection in contrast with the generic model [28]. Another study using a personalized deep learning system to detect hypoglycemic events from ECG rhythms found that the model overcame the limitations of intersubject variability in conventional systems [11]. The concept of adopting a personalized approach on ECG interpretation could improve specificity, which could prevent alert fatigue and overinterventions in a real-world setting. Our study demonstrated that transfer learning through a personalized approach required fewer ECG data queries and could bring about substantial improvements on sensitivity and specificity, which are useful in minimizing false alerts.

To enhance the interpretability of our model, we applied the Grad Cam method that allows one to easily scrutinize the areas the model is relying on (Figure 4). Between our generalized model and personalized model trained from transfer learning, the former focuses more on broad characteristics such as the P, QRS, and T segments that were considered as a well-defined series of changes by previous literature [29]; the latter focused more on a few localized areas that we believe are related to interpersonal differences. In each patient in the personalized group, the highlighted area of the ECG section from the Grad Cam visualization was different from the others. This explains how the personalized model performed better than the generic model and the interpersonal variation of ECGs occurred according to the level of hyperkalemia.

We used the t-SNE method for visualizing our learned embeddings in a lower dimension. This particular approach finds a joint probability distribution in a low dimension that closely represents the data points in the original high dimension by using gradient descent [18]. From our t-SNE results regarding the last convolution layer of the deep learning model, the personalized model showed two discrete groups compared to a more heterogeneous appearance in the generalized model (Figure 4). This shows that learned embedding can better separate the heartbeats according to potassium levels.

The study results demonstrate the potential of leveraging transfer learning on a personalized data set with fewer data while producing comparable results. Even in an individual participant, by using only four sets of data, the AUC increased significantly and many even plateaued. However, our study did have limitations that warrant future investigations and validations. First, since its database came from ICU records, only 16 patients who had multiple potassium drawings and corresponding ECG readings were included. However, in reality, it is difficult for a single patient to undergo so many blood tests, the results of each having its own corresponding ECG data. Second, by extending the framework to be multimodal, including other physiological signals such as blood pressure, age, gender, underlying disease, and weight, we could further enhance our model's performance, and this should, therefore, be the future research goal.

Conclusion

Using personalized transfer learning on single-lead ECG readings is sufficient to yield high AUC and accurate results for hyperkalemia detection. The visualization of the model interpretation demonstrated the interpersonal differences on ECG change according to hyperkalemia. This finding could potentially be extended to applications that continuously monitor one's ECGs, thus serving as a surveillance system for patients at high risk of hyperkalemia and avoiding unnecessary blood tests.

Acknowledgments

This work was supported by the Ministry of Science and Technology of Taiwan (grant MOST 109-2314-B-182A-174).

Data Availability

The data that support the findings of this study are openly available in the Medical Information Mart From Intensive Care III Waveform Database Matched Subset [13].

Authors' Contributions

IMC devised the project and main conceptual ideas, worked out most of the technical details, and drafted the manuscript. JYC drafted the manuscript and performed the numerical calculations. JYC and TYC verified and edited the article. YMW performed the graphic interpretation. CYC performed the statistical analysis. CTK, FJC, and FFFY contributed to the data collection and preparation. CHRL supervised the findings of this study. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Complete list of patients included for the personalized model development.
[PDF File (Adobe PDF File), 72 KB - jmir_v24i12e41163_app1.pdf]

Multimedia Appendix 2
The improvement in predictions after each training round for all patients in the personalized group.
[PDF File (Adobe PDF File), 70 KB - jmir_v24i12e41163_app2.pdf]

References

1. Rafique Z, Peacock F, Armstead T, Bischof JJ, Hudson J, Weir MR, et al. Hyperkalemia management in the emergency department: An expert panel consensus. J Am Coll Emerg Physicians Open 2021 Oct;2(5):e12572 [FREE Full text] [doi: 10.1002/emp2.12572] [Medline: 34632453]
2. Luo J, Brunelli SM, Jensen DE, Yang A. Association between serum potassium and outcomes in patients with reduced kidney function. Clin J Am Soc Nephrol 2016 Jan 07;11(1):90-100 [FREE Full text] [doi: 10.2215/CJN.01730215] [Medline: 26500246]
3. Clase CM, Carrero J, Ellison DH, Grams ME, Hemmelgarn BR, Jardine MJ, Conference Participants. Potassium homeostasis and management of dyskalemia in kidney diseases: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. Kidney Int 2020 Jan;97(1):42-61 [FREE Full text] [doi: 10.1016/j.kint.2019.09.018] [Medline: 31706619]
4. Thomas A, Silver SA, Perl J, Freeman M, Slater JJ, Nash DM, et al. The frequency of routine blood sampling and patient outcomes among maintenance hemodialysis recipients. Am J Kidney Dis 2020 Apr;75(4):471-479. [doi: 10.1053/j.ajkd.2019.08.016] [Medline: 31732233]
5. Surawicz B. Relationship between electrocardiogram and electrolytes. Am Heart J 1967 Jun;73(6):814-834. [doi: 10.1016/0002-8703(67)90233-5] [Medline: 5338052]
6. Wrenn KD, Slovis CM, Slovis BS. The ability of physicians to predict hyperkalemia from the ECG. Ann Emerg Med 1991 Nov;20(11):1229-1232. [doi: 10.1016/s0196-0644(05)81476-3] [Medline: 1952310]
7. Kwon J, Jung M, Kim K, Jo Y, Shin J, Cho Y, et al. Artificial intelligence for detecting electrolyte imbalance using electrocardiography. Ann Noninvasive Electrocardiol 2021 May;26(3):e12839 [FREE Full text] [doi: 10.1111/anec.12839] [Medline: 33719135]
8. Attia ZI, DeSimone CV, Dillon JJ, Sapir Y, Somers VK, Dugan JL, et al. Novel bloodless potassium determination using a signal-processed single-lead ECG. J Am Heart Assoc 2016 Jan 25;5(1):e002746 [FREE Full text] [doi: 10.1161/JAHA.115.002746] [Medline: 26811164]
9. Yasin OZ, Attia Z, Dillon JJ, DeSimone CV, Sapir Y, Dugan J, et al. Noninvasive blood potassium measurement using signal-processed, single-lead ecg acquired from a handheld smartphone. J Electrocardiol 2017;50(5):620-625 [FREE Full text] [doi: 10.1016/j.jelectrocard.2017.06.008] [Medline: 28641860]
10. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. JAMA Cardiol 2019 May 01;4(5):428-436 [FREE Full text] [doi: 10.1001/jamacardio.2019.0640] [Medline: 30942845]
11. 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. [doi: 10.1038/s41598-019-56927-5] [Medline: 31932608]
12. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035. [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
13. Moody B, Moody G, Villarroel M, Clifford GD, Silva I. MIMIC-III Waveform Database Matched Subset. PhysioNet 2020:1. [doi: 10.13026/c2294b]
14. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition; June 27-30, 2016; Las Vegas, NV. [doi: 10.1109/cvpr.2016.90]
15. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan;25(1):65-69 [FREE Full text] [doi: 10.1038/s41591-018-0268-3] [Medline: 30617320]
16. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 26, 2016; Las Vegas, NV p. 770-778.
17. Selvaraju R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 Presented at: 2017 IEEE International Conference on Computer Vision; October 22-29, 2017; Venice, Italy. [doi: 10.1109/iccv.2017.74]
18. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Machine Learning Res 2008;9(11):2579-2605.
19. Hoekema R, Uijen G, van Oosterom A. Geometrical aspects of the interindividual variability of multilead ECG recordings. IEEE Trans Biomed Eng 2001 May;48(5):551-559. [doi: 10.1109/10.918594] [Medline: 11341529]

20. Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. Expert Syst Applications X 2020 Sep;7:100033. [doi: 10.1016/j.eswax.2020.100033]
21. Mohebbanaaz, Kumar LVR, Sai YP. A new transfer learning approach to detect cardiac arrhythmia from ECG signals. Signal Image Video Processing 2022 Feb 18;16(7):1945-1953. [doi: 10.1007/s11760-022-02155-w]
22. Pal A, Srivastva R, Singh Y. CardioNet: an efficient ECG arrhythmia classification system using transfer learning. Big Data Res 2021 Nov;26:100271. [doi: 10.1016/j.bdr.2021.100271]
23. Salem M, Taheri S, Shiun-Yuan J. ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features. arXiv. Preprint posted online on December 11, 2018. [FREE Full text]
24. Weimann K, Conrad TOF. Transfer learning for ECG classification. Sci Rep 2021 Mar 04;11(1):5251. [doi: 10.1038/s41598-021-84374-8] [Medline: 33664343]
25. Wald DA. ECG manifestations of selected metabolic and endocrine disorders. Emerg Med Clin North Am 2006 Feb;24(1):145-57, vii. [doi: 10.1016/j.emc.2005.08.010] [Medline: 16308117]
26. Lin C, Lin C, Fang W, Hsu C, Chen S, Huang K, et al. A deep-learning algorithm (ECG12Net) for detecting hypokalemia and hyperkalemia by electrocardiography: algorithm development. JMIR Med Inform 2020 Mar 05;8(3):e15931 [FREE Full text] [doi: 10.2196/15931] [Medline: 32134388]
27. Stone A, Bornhorst J. An introduction to personalized medicine. Therapeutic Drug Monitoring 2012:121-143. [doi: 10.1016/b978-0-12-385467-4.00006-3]
28. Wu MH, Chang EJ, Chu TH. Personalizing a generic ECG heartbeat classification for arrhythmia detection: a deep learning approach. 2018 Presented at: 2018 IEEE Conference on Multimedia Information Processing and Retrieval; April 10-12, 2018; Miami, FL. [doi: 10.1109/mipr.2018.00024]
29. Dittrich K, Walls R. Hyperkalemia: ECG manifestations and clinical considerations. J Emerg Med 1986;4(6):449-455. [doi: 10.1016/0736-4679(86)90174-5] [Medline: 3559133]

Abbreviations

AUC: area under the receiver operating characteristic curve
CNN: convolutional neural network
ECG: electrocardiogram
Grad Cam: gradient-weighted class activation mapping
ICU: intensive care unit
MIMIC-III: Medical Information Mart From Intensive Care III
t-SNE: t-distributed stochastic neighbor embedding

Edited by G Eysenbach; submitted 17.07.22; peer-reviewed by Z Li, M Kapsetaki, Y Wang; comments to author 19.10.22; revised version received 07.11.22; accepted 17.11.22; published 05.12.22.

Please cite as:
Chiu IM, Cheng JY, Chen TY, Wang YM, Cheng CY, Kung CT, Cheng FJ, Yau FFF, Lin CHR
Using Deep Transfer Learning to Detect Hyperkalemia From Ambulatory Electrocardiogram Monitors in Intensive Care Units: Personalized Medicine Approach
J Med Internet Res 2022;24(12):e41163
URL: https://www.jmir.org/2022/12/e41163
doi:10.2196/41163
PMID:36469396

Original Paper

Model for Predicting In-Hospital Mortality of Physical Trauma Patients Using Artificial Intelligence Techniques: Nationwide Population-Based Study in Korea

Seungseok Lee[1*], BE; Wu Seong Kang[2*], MD, PhD; Sanghyun Seo[3*], MD; Do Wan Kim[4*], MD; Hoon Ko[1], MSc; Joongsuck Kim[2], MD; Seonghwa Lee[5], MD, PhD; Jinseok Lee[1], PhD

[1]Department of Biomedical Engineering, Kyung Hee University, Yong-in, Republic of Korea

[2]Department of Trauma Surgery, Jeju Regional Trauma Center, Cheju Halla General Hospital, Jeju, Republic of Korea

[3]Department of Radiology, Wonkwang University Hospital, Iksan, Republic of Korea

[4]Department of Thoracic and Cardiovascular Surgery, Chonnam National University Hospital, Chonnam National University Medical School, Gwangju, Republic of Korea

[5]Department of Emergency Medicine, Jeju Regional Trauma Center, Cheju Halla General Hospital, Jeju, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jinseok Lee, PhD
Department of Biomedical Engineering
Kyung Hee University
1732, Deogyeong-daero, Giheung-gu
Yong-in, 17104
Republic of Korea
Phone: 82 312012570
Fax: 82 312012571
Email: gonasago@khu.ac.kr

Abstract

Background: Physical trauma–related mortality places a heavy burden on society. Estimating the mortality risk in physical trauma patients is crucial to enhance treatment efficiency and reduce this burden. The most popular and accurate model is the Injury Severity Score (ISS), which is based on the Abbreviated Injury Scale (AIS), an anatomical injury severity scoring system. However, the AIS requires specialists to code the injury scale by reviewing a patient's medical record; therefore, applying the model to every hospital is impossible.

Objective: We aimed to develop an artificial intelligence (AI) model to predict in-hospital mortality in physical trauma patients using the International Classification of Disease 10th Revision (ICD-10), triage scale, procedure codes, and other clinical features.

Methods: We used the Korean National Emergency Department Information System (NEDIS) data set (N=778,111) compiled from over 400 hospitals between 2016 and 2019. To predict in-hospital mortality, we used the following as input features: ICD-10, patient age, gender, intentionality, injury mechanism, and emergent symptom, Alert/Verbal/Painful/Unresponsive (AVPU) scale, Korean Triage and Acuity Scale (KTAS), and procedure codes. We proposed the ensemble of deep neural networks (EDNN) via 5-fold cross-validation and compared them with other state-of-the-art machine learning models, including traditional prediction models. We further investigated the effect of the features.

Results: Our proposed EDNN with all features provided the highest area under the receiver operating characteristic (AUROC) curve of 0.9507, outperforming other state-of-the-art models, including the following traditional prediction models: Adaptive Boosting (AdaBoost; AUROC of 0.9433), Extreme Gradient Boosting (XGBoost; AUROC of 0.9331), ICD-based ISS (AUROC of 0.8699 for an inclusive model and AUROC of 0.8224 for an exclusive model), and KTAS (AUROC of 0.1841). In addition, using all features yielded a higher AUROC than any other partial features, namely, EDNN with the features of ICD-10 only (AUROC of 0.8964) and EDNN with the features excluding ICD-10 (AUROC of 0.9383).

Conclusions: Our proposed EDNN with all features outperforms other state-of-the-art models, including the traditional diagnostic code-based prediction model and triage scale.

(J Med Internet Res 2022;24(12):e43757) doi:10.2196/43757

KEYWORDS

artificial intelligence; trauma; mortality prediction; international classification of disease; injury; prediction model; severity score; emergency department; Information system; deep neural network

Introduction

Physical trauma–related mortality places a heavy burden on individuals and society. Accurately estimating mortality risk enhances treatment efficiency and reduces this burden. To date, there are various models to predict the severity of physical trauma patients [1-7]. Among them, the most popular and accurate model is the Injury Severity Score (ISS) developed in the 1970s and based on the Abbreviated Injury Scale (AIS), an anatomical injury severity scoring system [1,8]. However, the AIS requires specialists to code the injury scale by reviewing a patient's medical record; therefore, applying the model to every hospital is impossible. To overcome these shortcomings, the following International Classification of Diseases (ICD)–based severity models have been introduced: ICD-based Injury Severity Score (ICISS)[9], trauma mortality models using International Classification of Disease 10th Revision (ICD-10) (TMPM-ICD10) [10], and Mortality Ratio-adjusted Injury Severity Score (EMR-ISS) [11]. However, ICD-based models are not as accurate as AIS-based models [8]. Since 2016, all emergency medical institutions in Korea have introduced the Korean Triage and Acuity Scale (KTAS), an emergency department (ED) triage system composed of 5 levels [12]. However, the KTAS relies on the practitioner’s judgment and may introduce bias and be prone to human error [13].

Artificial intelligence (AI) is widely used to find complex associations between various features in medical applications [14-16], such as individual injuries and mortality. We recently proposed AI technology utilizing AIS codes that outperformed conventional ISS [1], providing a favorable area under the receiver operating characteristic (AUROC) of 0.908 [17]. Tran et al [18] also used AI technology for mortality prediction using the ICD-10 from the National Trauma Database (NTDB) data set, but the AUROC value was not as high as that of our previous proposed AI model.

We aimed to construct an AI model to predict in-hospital mortality in physical trauma patients using the National Emergency Department Information System (NEDIS) data set. We hypothesized that an AI model based on ICD-10 with other clinical features is a useful alternative. We compared the predictive performance of our model with other ICD-10-based models, such as the ICISS [9], EMR-ISS [11], and the AI-driven ICD-10-only based model. Finally, we deployed our AI-driven public website to predict in-hospital mortality in physical trauma patients to benefit end users.

Methods

Ethics Approval

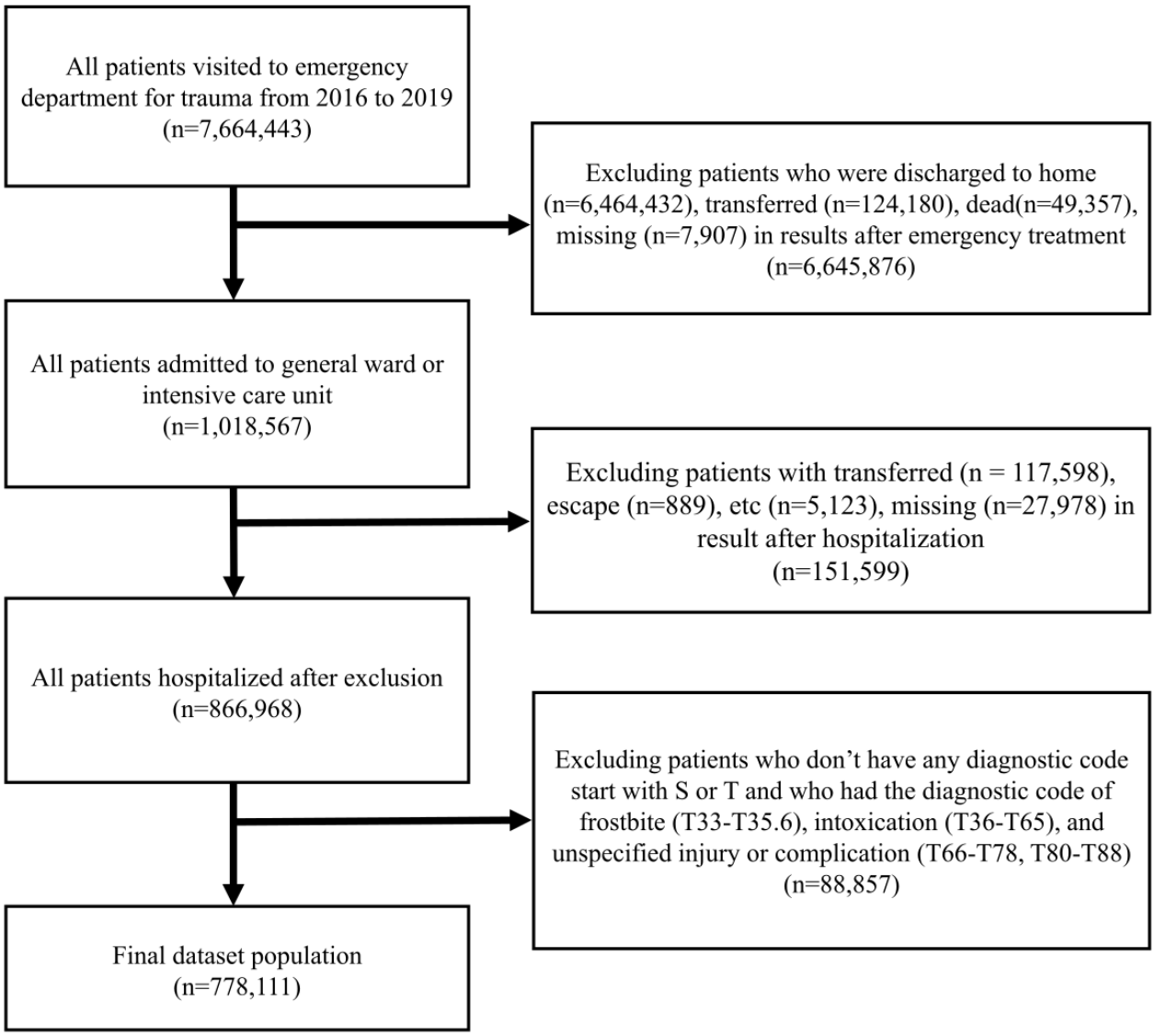
This study was conducted according to the TRIPOD (Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis) statement [19]. NEDIS data were provided by the National Emergency Medical Center (data acquisition number N20212920825).

Patients and Data Set for AI Model

The NEDIS data set was collected mandatorily from 2016 to 2019 from over 400 hospitals in South Korea. The inclusion criteria were as follows: (1) physical trauma patients (but not psychological) with a diagnostic code of S or T based on the Korean version of the ICD-10; (2) patients admitted to the intensive care unit (ICU) or general ward from the ED; and (3) patients admitted to the ICU or general ward after surgery or a procedure from the ED. The exclusion criteria were as follows: (1) patients without diagnostic codes starting with S or T (eg, S001, T063; all physical traumatic patients include S or T code. The S code represents the trauma in a single body region, and the T code represents the trauma in multiple or unspecified regions); (2) patients with diagnostic code of frostbite (T33-T35.6), intoxication (T36-T65), and unspecified injury or complication (T66-T78, T80-T88); (3) patients transferred to another hospital or discharged from the ED after treatment; (4) patients transferred to another hospital or discharged without notification to staffs at hospitals; (5) patients who died in the ED before ICU or general ward admission; and (6) missing information.

More specifically, we first collected 7,664,443 patients with a nondisease identifier comprising trauma patients. Since our primary outcome was to predict in-hospital mortality in trauma patients, we had to exclude unrelated patients. We then excluded all nonhospitalized patient information (n=6,464,432, 84.34%). The second most commonly excluded data were from patients transferred to another hospital (n=241,778, 3.15%). For transferred patients, the NEDIS policy of deidentification is to assign a new anonymous ID number; thus, the data is redundant. In addition, we excluded deceased ED patients (n=49,357, 0.64%) due to insufficient information about diagnostic codes, procedure codes, and other clinical features. Moreover, we excluded escaped patients during hospitalization (n=889, 0.01%) and patients with missing data (n=35,885, 4.68%), not including mortality information. A final total of 778,111 patient data were used for training and testing our AI model (Figure 1).

Figure 1. Flowchart of the patient selection process.



We used the following variables in NEDIS data: age, gender, intentionality, injury mechanism, emergent symptom, Alert/Verbal/Painful/Unresponsive (AVPU) scale, initial KTAS, altered KTAS, ICD-10 codes, procedure codes of surgical operation or interventional radiology, and in-hospital mortality. All included variables for the AI model are summarized in Table 1. A total of 938 AI model input features (categories) were considered from 10 variables. The AVPU scale is a simplified version of the Glasgow Coma Scale (GCS) [20,21] and includes 4 categories: A, alert; V, verbal responsive (drowsy); P, painful response (stupor, semicoma); and U, unresponsive (coma). KTAS was developed as a severity triage in the ED in 2012, based on the Canadian Triage and Acuity Scale (CTAS) [12]. KTAS is a standardized triage tool to avoid complexity and ambiguity and includes 5 categories: level 1, resuscitation; level 2, emergent; level 3, urgent; level 4, less urgent; level 5, nonurgent. According to NEDIS policy, KTAS should be conducted by a certified faculty, and the initial KTAS should be assessed within 2 minutes of ED admission. The altered KTAS should be assessed when the ED patient deteriorates before moving to the operating room, ICU, or general ward. Regarding ICD-10, we considered 856 codes starting with S or T. The procedure codes, which are used to claim from the National Health Insurance Review and Assessment Service, include surgery and angioembolization and are more specifically categorized as follows and summarized in Table S1 in Multimedia Appendix 1: (1) head procedure; (2) torso procedure-vascular; (3) torso procedure-abdomen; (4) torso procedure-chest; (5) torso procedure-heart; and (6) extracorporeal membrane oxygenation (ECMO). The primary outcome was in-hospital mortality, defined as a patient with a dead result code and discharged with medical futility in NEDIS.

Table 1. Included variables of the Korean National Emergency Department Information System (NEDIS) for the artificial intelligence (AI) model.

Value, n	Variables	Type	Description
1	Age	26 categories	• 5-year-old unit, classification
2	Gender	2 categories	• M: male • F: female
3	Intentionality	5 categories	• 1: accidental, unintentional • 2: self-harm, suicide • 3: violence, assault • 4: other specified • 5: unspecified • 6: no data
4	Injury mechanism	16 categories	• 1: traffic accident-car • 2: traffic accident-bike • 3: traffic accident-motorcycle • 4: traffic accident-etc • 5: traffic accident-unspecified • 6: fall • 7: slip down • 8: struck • 9: firearm/cut/pierce • 10: machine • 11: fire, flames or heat • 12: drowning or nearly • 13: poisoning • 14: choking, hanging • 15: etc • 16: unknown • 17: no data
5	Emergent symptoms	2 categories	• Y: emergency • N: nonemergency
6	AVPU[a] scale	4 categories	• A: alert • V: verbal response (drowsy) • P: painful response (semicoma) • U: unresponsive (coma) • N: unknown
7	Initial KTAS[b]	7 categories	• 1: Level 1 (resuscitation) • 2: Level 2 (emergency) • 3: Level 3 (urgency) • 4: Level 4 (less urgency) • 5: Level 5 (nonurgency) • 6: etc • 7: unknown • 8: no data
8	Altered KTAS	5 categories	• 1: Level 1 (resuscitation) • 2: Level 2 (emergency) • 3: Level 3 (urgency) • 4: Level 4 (less urgency) • 5: Level 5 (nonurgency) • 6: etc • 7: no data
9	Diagnostic code at discharge	865 categories	• ICD-10[c] codes starting S[d] or T[e]
10	Procedure code after hospitalized	6 categories	• Procedure code including surgery or interventional radiology

[a]AVPU: Alert/Verbal/Painful/Unresponsive.

[b]KTAS: Korean Triage and Acuity Scale.

[c]ICD-10: International Classification of Disease 10th Revision.

[d]Represents trauma in a single body region.

[e]Represents trauma in multiple or unspecified regions.

Data Split, Data Balancing, and Cross-Validation

The data set in this study comprised both training and testing data (Table S2 in Multimedia Appendix 1). Data from 778,111 patients were divided into training and testing data with a ratio of 8:2 in a stratified fashion. The testing set was used only to independently test our developed AI model and not for training or internal validation.

We first performed 5-fold cross-validation using the training data to confirm its generalization ability. The training data set (n=622,488, 80%) was randomly shuffled and stratified into 5 equal groups, of which 4 groups were selected from training the model, and the remaining group was used for internal validation. This process was repeated 5 times by shifting the internal validation group. Our finalized AI model is described in the subsequent sections and was used to evaluate performance using the isolated testing data.

Since the number of survived patients (n=611,481, 98.23%) was much higher than that of deceased patients (n=11,007, 1.77%), we upsampled the survived patient data using the Synthetic Minority Oversampling Technique (SMOTE) during the model update [22]. By balancing the 2 groups, we prevented bias toward the survived patient data.

Feature Analysis

To analyze the effects on mortality prediction from 914 features, we applied 3 machine-learning algorithms: Adaptive Boosting (AdaBoost) [23], Extreme Gradient Boosting (XGBoost) [24], and light gradient boosting machine (LightGBM) [25]. We also considered 4 ensemble models: AdaBoost with XGBoost, AdaBoost with LightGBM, XGBoost with LightGBM, and a combination of the 3 models. Finally, among 7 machine learning models, we chose the best prediction model and presented its feature importance analysis, listing features in the order that they contributed to the mortality prediction.

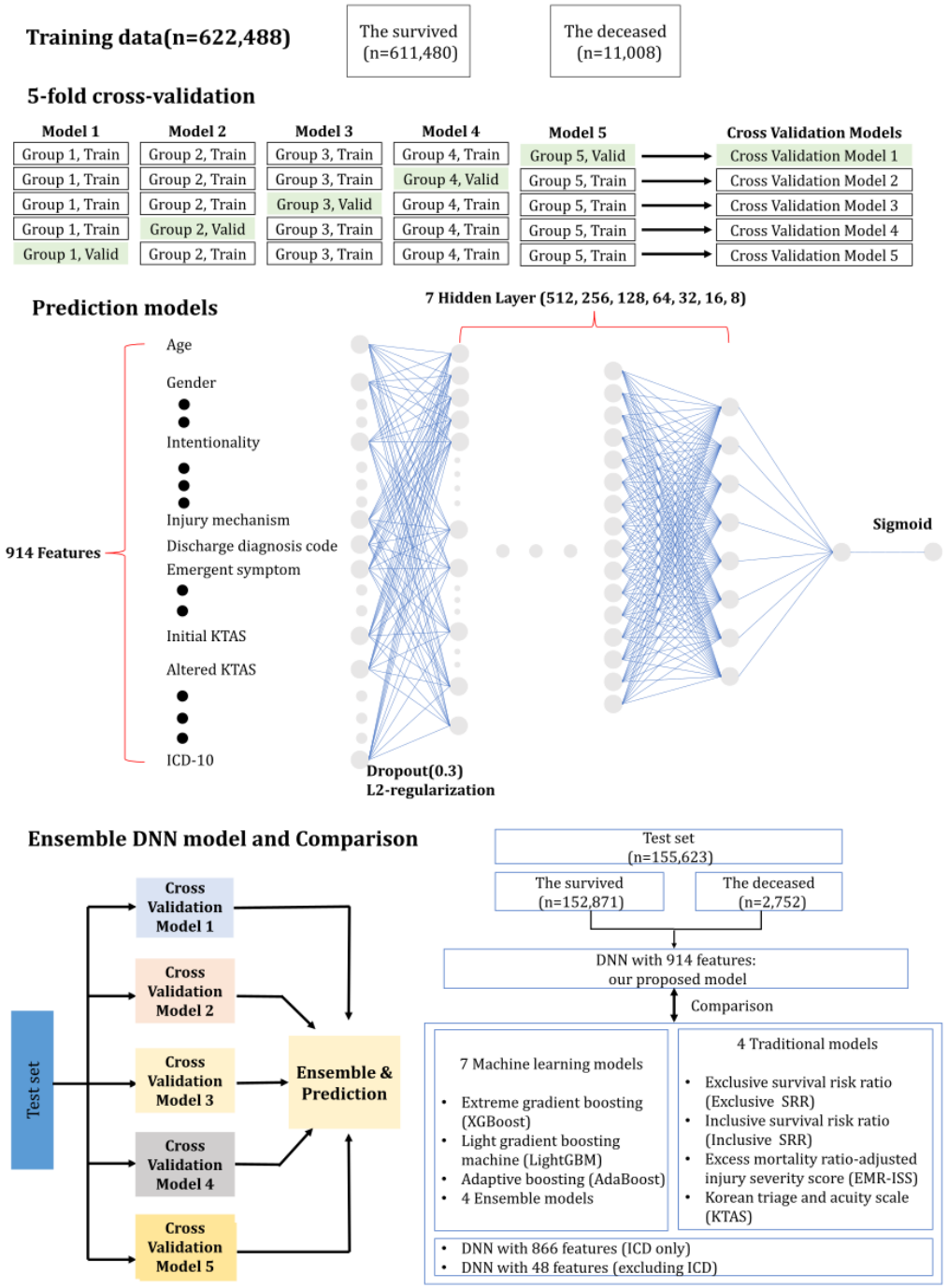
Performance evaluations were based on 5-fold cross-validation using 5 metrics: sensitivity, specificity, accuracy, balanced accuracy, and AUROC.

AI Prediction Model Development and Statistical Analysis

We developed a deep neural network (DNN)–based AI model using 914 features, including ICD-10 as an input layer. To find the best model, we searched hyperparameters, such as layer depth and width for fully connected (FC) layers. The last FC output layer was fed into a sigmoid layer, which provided the mortality probability. After the hyperparameter search, we found the best model with a 9-layer DNN, which comprised an input layer, 7 FC layers as hidden layers, and an output layer. The input layer was fed into a series of 7 FC layers, consisting of 512, 256, 128, 64, 32, 16, and 8 nodes, respectively. We applied dropout with a rate of 0.3 and L2 regularization for the FC hidden layers. Figure 2 shows the process flow of the AI development and DNN architecture. The prediction performance of our proposed 9-layer DNN model was evaluated with 5-fold cross-validation. Subsequently, for the final DNN-based AI model, we adopted an ensemble approach to combine the 5 models from the 5-fold cross-validation. The 914 features were inputs to 5 cross-validation models, and each provided mortality probabilities. A total of 5 probabilities were averaged, known as soft voting. Based on the ensemble DNN model, the prediction performance was evaluated with the isolated testing data set (n=155,623, 20%).

We trained the models with an Adam optimizer and binary cross-entropy cost function with a learning rate of 0.001 and a batch size of 32. We implemented the models using Python (version 3.7.13) with TensorFlow (version 2.8.0), Keras (version 2.8.0), NumPy (version 1.21.6), Pandas (version 1.3.5), Matplotlib (version 3.5.1), and Scikit-learn (version 1.0.2). All statistical analyses were performed using R software version 4.1.2 (R Foundation for Statistical Computing). As appropriate, proportions were compared using the chi-square test or Fisher exact test. A P value <.05 was considered statistically significant.

Figure 2. Process flow of our artificial intelligence (AI) model development: data, deep neural network (DNN) architecture, ensemble DNN model, and performance comparison. AdaBoost: Adaptive Boosting; EMR-ISS: mortality ratio-adjusted Injury Severity Score; ICD: International Classification of Diseases; KTAS: Korean Triage and Acuity Scale; LightGBM: light gradient boosting machine; SRR: survival risk ratio; XGBoost: Extreme Gradient Boosting.



Conventional Metrics Based on Diagnostic Code

We applied conventional metrics based on ICD-10. ICISS utilizes survival risk ratios (SRRs) to calculate the probability of survival [9]. SRR is defined as the number of survived patients with a specific injury code divided by the number of all patients with the specific same injury code. A patient's probability of survival (Ps) is determined by multiplying all SRRs of the injury codes from the patient [9]. The traditional ICISS was calculated as the product of Ps for as many as 10 injuries [26]. Two different methods were performed to calculate ICISS. First, the inclusive SRR was calculated for each injury irrespective of the associated injury [9]. Second, the exclusive SRR was calculated by the number of survivors who had an isolated specific injury divided by the total number of patients who only had that injury [9]. Thus, patients with multiple injuries were excluded from the calculation of exclusive SRR [9]. Regarding EMR-ISS, an injury severity grade similar to AIS was produced from ICD-10 codes based on the quintile of the EMR for each ICD-10 code [11]. The EMR-ISS was calculated from 3 maximum severity grades using data from the National Health Insurance data set, the Industrial Accident Compensation Insurance data set, and the National Death Certificate database from 2001 to 2003 [11].

Results

Initial Findings

Of the 778,111 patients included in the final analysis, 13,760 (1.77%) died during hospitalization (13,667 had a deceased code, and 93 were discharged with a medical futility code). Table 2 shows a comparison of included variables between deceased and surviving patients, and Table S3 in Multimedia Appendix 1 shows the ICD-10 comparison between deceased and surviving patients.

Table 2. Comparison of included variables of the Korean National Emergency Department Information System (NEDIS) between deceased and survived patients.

Variables	Deceased (N=13,760), n (%)	Survived (N=764,351), n (%)	*P* value
Age (years)			<.001
<1	16 (0.1)	2139 (0.3)	
1-4	63 (0.5)	9706 (1.3)	
5-9	38 (0.3)	16,345 (2.1)	
10-14	36 (0.3)	16,788 (2.2)	
15-19	171 (1.2)	25,776 (3.4)	
20-24	229 (1.7)	31,000 (4.1)	
25-29	214 (1.6)	33,241 (4.3)	
30-34	176 (1.3)	32,151 (4.2)	
35-39	276 (2)	38,611 (5.1)	
40-44	326 (2.4)	42,013 (5.5)	
45-49	583 (4.2)	55,126 (7.2)	
50-54	768 (5.6)	66,276 (8.7)	
55-59	1055 (7.7)	78,447 (10.3)	
60-64	1110 (8.1)	66,899 (8.8)	
65-69	1155 (8.4)	52,900 (6.9)	
70-74	1388 (10.1)	50,396 (6.6)	
75-79	2028 (14.7)	58,334 (7.6)	
80-84	1925 (14)	48,440 (6.3)	
85-89	1320 (9.6)	27,670 (3.6)	
90-94	665 (4.8)	9723 (1.3)	
95-99	189 (1.4)	21,08 (0.3)	
100-104	25 (0.2)	226 (0)	
105-109	4 (0)	25 (0)	
110-114	0 (0)	9 (0)	
115-119	0 (0)	2 (0)	
Procedure code			
Head procedure	2473 (18)	6419 (0.8)	<.001
Torso procedure-vascular	880 (6.4)	4961 (0.6)	<.001
Torso procedure-abdomen	810 (5.9)	4544 (0.6)	<.001
Torso procedure-chest	1209 (8.8)	7228 (0.9)	<.001
Torso procedure-heart	39 (0.3)	127 (0)	<.001
ECMO[a]	183 (1.3)	39 (0)	<.001
Initial KTAS[b]			
Level 1	3800 (27.6)	2812 (0.4)	<.001
Level 2	4209 (30.6)	49,234 (6.4)	<.001
Level 3	3306 (24)	270,574 (35.4)	<.001
Level 4	2020 (14.7)	346,663 (45.4)	<.001
Level 5	235 (1.7)	49,892 (6.5)	<.001
Not classified	4 (0)	301 (0)	.698
Unspecified	0 (0)	13 (0)	>.99

Variables	Deceased (N=13,760), n (%)	Survived (N=764,351), n (%)	*P* value
Missing data	186 (1.4)	44,862 (5.9)	<.001
Altered KTAS			
Level 1	2938 (21.4)	1921 (0.3)	<.001
Level 2	3173 (23.1)	35,356 (4.6)	<.001
Level 3	2784 (20.2)	241,201 (31.6)	<.001
Level 4	873 (6.3)	189,314 (24.8)	<.001
Level 5	108 (0.8)	27,355 (3.6)	<.001
Not classified	0 (0)	5 (0)	>.99
Missing data	3884 (28.2)	269,199 (35.2)	<.001
Intentionality			
Accidental, unintentional	12078 (87.8)	574,556 (75.2)	<.001
Suicide, intentional self-harm	248 (1.8)	6235 (0.8)	<.001
Assault, violence	113 (0.8)	12,989 (1.7)	<.001
Other specified	132 (1)	1694 (0.2)	<.001
Unspecified	548 (4)	12,225 (1.6)	<.001
Missing data	641 (4.7)	156,652 (20.5)	<.001
Injury mechanism			
Traffic accident-car	1154 (8.4)	98,320 (12.9)	<.001
Traffic accident-bike	450 (3.3)	20,692 (2.7)	<.001
Traffic accident-motorcycle	1020 (7.4)	31,957 (4.2)	<.001
Traffic accident-pedestrian, train, airplane, ship, etc	1925 (14)	35,898 (4.7)	<.001
Traffic accident-unknown	18 (0.1)	197 (0)	<.001
Fall down	2374 (17.3)	76,714 (10)	<.001
Slip down	3859 (28)	16,8677 (22.1)	<.001
Struck by person or object	713 (5.2)	60,518 (7.9)	<.001
Firearm/cut (sharp or object)/piece	159 (1.2)	39,515 (5.2)	<.001
Machine	54 (0.4)	16,991 (2.2)	<.001
Fire, flames, or heat	207 (1.5)	6587 (0.9)	<.001
Drowning or nearly drowning	20 (0.1)	203 (0)	<.001
Poisoning	62 (0.5)	1811 (0.2)	<.001
Choking, hanging	146 (1.1)	436 (0.1)	<.001
Others-rape, electric	323 (2.3)	35,461 (4.6)	<.001
Unknown	635 (4.6)	13,722 (1.8)	<.001
Missing data	641 (4.7)	156,652 (20.5)	<.001
Emergent symptom			
Yes	13351 (97)	69,7118 (91.2)	<.001
No	409 (3)	67,228 (8.8)	<.001
Unspecified	0 (0)	5 (0)	>.99
AVPU[c] scale			
Alert	5403 (39.3)	579,669 (75.8)	<.001
Verbal response (drowsy)	1393 (10.1)	12,085 (1.6)	<.001
Painful response (stupor, semicoma)	3218 (23.4)	5581 (0.7)	<.001

Variables	Deceased (N=13,760), n (%)	Survived (N=764,351), n (%)	*P* value
Unresponsive (coma)	3049 (22.2)	847 (0.1)	<.001
Unspecified response	697 (5.1)	166,169 (21.7)	<.001
Sex			
Male	9050 (65.8)	434,280 (56.8)	<.001

[a]ECMO: extracorporeal membrane oxygenation.

[b]KTAS: Korean Triage and Acuity Scale.

[c]AVPU: Alert/Verbal/Painful/Unresponsive.

K-Fold Cross-Validation Results

Table 3 summarizes the 5-fold cross-validation results. Our model used all 914 features, including ICD-10, and provided the highest balanced accuracy (0.8718) and AUROC (0.9513) values. Among the machine learning models, AdaBoost provided the highest balanced accuracy (0.8603) and AUROC (0.9442). Any ensemble models from the combination of AdaBoost, XGBoost, and LightGBM did not improve accuracy above our model or AdaBoost. Compared to our model, traditional methods produced lower balanced accuracy and AUROC values. More specifically, inclusive SRR resulted in a lower balanced accuracy of 0.7888 and AUROC of 0.8266, while exclusive SRR resulted in 0.7931 and 0.8737, and EMR-ISS yielded 0.7571 and 0.6108, respectively. KTAS resulted in an even lower balanced accuracy of 0.5372 and AUROC of 0.1057.

Of the models considering 866 features of ICD-10 only, DNN demonstrated the highest balanced accuracy (0.8234) and AUROC (0.8975), followed by AdaBoost, the ensemble of AdaBoost and XGBoost, and the ensemble of AdaBoost and LightGBM. However, the models generated much lower balanced accuracy and AUROC values compared to models considering 48 features, excluding ICD-10.

Table 3. Results of the 5-fold cross-validation.

Model	Sensitivity, mean (SD)	Specificity, mean (SD)	Accuracy, mean (SD)	Balanced accuracy, mean (SD)	AUROC[a], mean (SD)
Using all 914 features (including ICD-10[b])					
Proposed model (DNN[c])	0.8599 (0.0151)	0.8838 (0.0097)	0.8834 (0.0093)	0.8718 (0.0036)	0.9513 (0.0023)
AdaBoost[d]	0.818 (0.0100)	0.9025 (0.0006)	0.9010 (0.0005)	0.8603 (0.0048)	0.9442 (0.0020)
XGBoost[e]	0.8105 (0.0085)	0.8865 (0.0011)	0.8854 (0.0010)	0.8485 (0.0037)	0.9354 (0.0018)
LightGBM[f]	0.8112 (0.0080)	0.8861 (0.0018)	0.8848 (0.0016)	0.8486 (0.0032)	0.9354 (0.0019)
AdaBoost+XGBoost	0.8109 (0.0073)	0.8882 (0.0013)	0.8868 (0.0012)	0.8496 (0.0034)	0.9367 (0.0017)
AdaBoost+LightGBM	0.8118 (0.0081)	0.8875 (0.0014)	0.8862 (0.00130)	0.8497 (0.0035)	0.9367 (0.0018)
XGBoost+LigtGBM	0.8104 (0.0079)	0.8865 (0.0010)	0.8851 (0.0009)	0.8484 (0.0035)	0.9354 (0.0018)
AdaBoost+XGBoost+LightGBM	0.8107 (0.0075)	0.8871 (0.0011)	0.8857 (0.0010)	0.8489 (0.0033)	0.9361 (0.0018)
Using 866 features (ICD-10 only)					
DNN	0.8294 (0.0153)	0.8175 (0.009)	0.8177 (0.0086)	0.8234 (0.0037)	0.8975 (0.0023)
AdaBoost	0.7586 (0.0157)	0.8493 (0.0048)	0.8477 (0.0045)	0.8039 (0.0057)	0.8796 (0.0030)
XGBoost	0.6575 (0.0141)	0.8939 (0.0035)	0.8897 (0.0032)	0.7757 (0.0055)	0.8627 (0.0033)
LightGBM	0.6585 (0.0115)	0.8937 (0.0024)	0.8896 (0.0022)	0.7761 (0.0049)	0.8635 (0.0037)
AdaBoost+XGBoost	0.6637 (0.0065)	0.8922 (0.0017)	0.8882 (0.0016)	0.7780 (0.0027)	0.8785 (0.0029)
AdaBoost+LightGBM	0.6640 (0.0076)	0.8918 (0.0012)	0.8878 (0.0011)	0.7779 (0.0032)	0.8786 (0.0031)
XGBoost+LigtGBM	0.6590 (0.0117)	0.8932 (0.0024)	0.8891 (0.0022)	0.7761 (0.0048)	0.8635 (0.0035)
AdaBoost+XGBoost+LightGBM	0.6624 (0.0070)	0.8924 (0.0017)	0.8883 (0.0016)	0.7774 (0.0029)	0.8784 (0.0028)
Using 48 features (excluding ICD-10)					
DNN	0.8003 (0.0266)	0.9072 (0.0161)	0.9053 (0.0154)	0.8537 (0.0068)	0.9398 (0.003)
AdaBoost	0.8148 (0.0125)	0.8922 (0.0022)	0.8908 (0.0022)	0.8535 (0.0062)	0.9380 (0.0025)
XGBoost	0.8294 (0.0056)	0.863 (0.0033)	0.8623 (0.0032)	0.8462 (0.0018)	0.9328 (0.0022)
LightGBM	0.8323 (0.0044)	0.8619 (0.0032)	0.8614 (0.0032)	0.8471 (0.0018)	0.9328 (0.0021)
AdaBoost+XGBoost	0.8303 (0.0058)	0.8635 (0.0029)	0.8630 (0.0028)	0.8469 (0.0019)	0.9337 (0.0022)
AdaBoost+LightGBM	0.8314 (0.0052)	0.8634 (0.0028)	0.8628 (0.0027)	0.8474 (0.002)	0.9336 (0.0020)
XGBoost+LigtGBM	0.8321 (0.0046)	0.8618 (0.0032)	0.8613 (0.0031)	0.847 (0.0020)	0.9328 (0.0021)
AdaBoost+XGBoost+LightGBM	0.8312 (0.0052)	0.8630 (0.0024)	0.8624 (0.0024)	0.8471 (0.0022)	0.9333 (0.0021)
Traditional methods[g]					
Inclusive SRR[h]	0.8953	0.6823	0.7893	0.7888	0.8266
Exclusive SRR	0.8272	0.7590	0.7936	0.7931	0.8737
EMR-ISS[i]	0.7867	0.7276	0.7572	0.7571	0.6108
KTAS[j]	0.9353	0.1390	0.5495	0.5372	0.1057

[a]AUROC: area under the receiver operating characteristic.

[b]ICD-10: International Classification of Disease 10th Revision.

[c]DNN: deep neural network.

[d]AdaBoost: Adaptive Boosting.

[e]XGBoost: Extreme Gradient Boosting.

[f]LightGBM: light gradient boosting machine.

[g]Only yielded a single value, so no SD is reported.

[h]SRR: survival risk ratio.

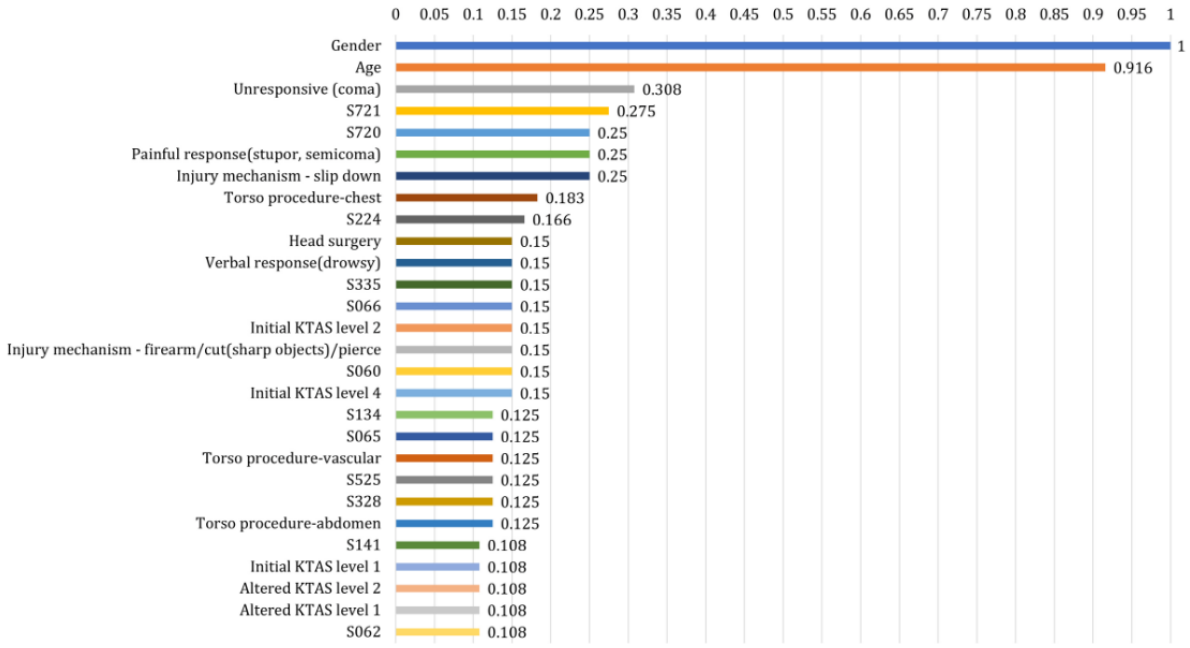
[i]EMR-ISS: Mortality Ratio-adjusted Injury Severity Score.

[j]KTAS: Korean Triage and Acuity Scale.

Ranked Feature Importance: Explainable AI

To analyze the effects of features, we first applied the data to 3 different machine learning algorithms: AdaBoost, XGBoost, and LightGBM. As summarized in Table 3, the AdaBoost model was the best classifier for predicting mortality in trauma patients. We then performed the feature importance analysis (see Figure 3 for ranked normalized feature importance) to confirm the contribution of each feature.

Figure 3. Results of the ranked normalized feature importance from the Adaptive Boosting (AdaBoost) model. KTAS: Korean Triage and Acuity Scale.



Based on the AdaBoost, gender had the highest importance value, followed by age, unresponsive (coma), S721 (pertrochanteric fracture of the femur), S720 (fracture of neck of femur), painful response (stupor, semicoma), injury mechanism-slip down, and torso procedure-chest. Among the 914 features, only 71 (7.77%) features had nonzero values indicating that the other 843 features did not contribute to mortality prediction. Table S4 in Multimedia Appendix 1 shows the complete ranked normalized feature importance values. All features with the highest importance value showed a statistically significant difference between the deceased and surviving group (Table 2 and Table S3 in Multimedia Appendix 1).

Cross-Validation Result of DNNs Using a Different Set of Features According to Importance

We investigated the cross-validation performance from our DNN model with 2 input conditions: (1) the top 71 features having nonzero feature importance values from the AdaBoost, the best among the machine learning models; and (2) all 914 features (Table S5 in Multimedia Appendix 1). The DNN with all 914 features provided a higher balanced accuracy of 0.8718 and AUROC of 0.9513 compared to the DNN with the top 71 features, which had a balanced accuracy of 0.8389 and AUROC of 0.9386. Features with 0 values of feature importance can contribute to mortality prediction. Sensitivity increased by more than 0.1 for the former, whereas specificity decreased by less than 0.05. For the latter, sensitivity increased to 0.8599 from 0.7480, and specificity decreased to 0.8838 from 0.9299. Therefore, we considered all features in our AI model and validated the performance with the isolated testing data.

Testing Data Results

With the testing data set (n=155,623), our proposed ensemble-based 9-layer DNN showed a sensitivity of 0.8768, specificity of 0.8625, accuracy of 0.8628, balanced accuracy of 0.8697, and AUROC of 0.9507. Furthermore, compared with the cross-validation results, the model was neither overfitted nor underfitted, with minimal differences between cross-validation and testing data results: sensitivity of 0.8599 versus 0.8768, specificity of 0.8838 versus 0.8625, accuracy of 0.8834 versus 0.8628, balanced accuracy 0.8718 versus 0.8697, and AUROC of 0.9513 versus 0.9507.

Our proposed ensemble of deep neural networks (EDNN) using all 914 features demonstrated higher values of balanced accuracy and AUROC than any other model (Table 4). Models with 48 features provided the next most accurate prediction results. These results showed the same trend as the cross-validation results. Figure 4 shows the AUROC curves for our model, AdaBoost, XGBoost, and LightGBM, which are plotted according to the following features: all 914 features, 48 features excluding ICD-10, and 866 features with ICD-10 only. Our model outperformed the traditional methods such as inclusive SRR, exclusive SRR, EMR-ISS, and KTAS. Figure 5 shows the AUROC curves for our model and 4 traditional models. The calculated inclusive SRR and exclusive SRR are shown in Table S6 in Multimedia Appendix 1. Finally, the model using the top 71 features from the AdaBoost also provided a lower balanced accuracy of 0.8245 and AUROC of 0.9194, similar to the cross-validation results.

Table 4. Comparison of the prediction performances of the prediction models on the test data set.

Model	Sensitivity	Specificity	Accuracy	Balanced accuracy	AUROC[a]
Using all 914 features (including ICD-10[b])					
Proposed model (DNN[c])	0.8768	0.8625	0.8628	0.8697	0.9507
AdaBoost[d]	0.8619	0.8655	0.8654	0.8637	0.9433
XGBoost[e]	0.8292	0.8660	0.8653	0.8476	0.9331
LightGBM[f]	0.8601	0.8365	0.8369	0.8483	0.9332
Using 866 features (ICD-10 only)					
DNN	0.8365	0.8159	0.8162	0.8262	0.8964
AdaBoost	0.7896	0.8319	0.8312	0.8108	0.8773
XGBoost	0.7660	0.8348	0.8336	0.8004	0.8564
LightGBM	0.7729	0.8285	0.8276	0.8007	0.8565
Using 48 features (excluding ICD-10)					
DNN	0.8347	0.8784	0.8776	0.8565	0.9383
AdaBoost	0.8354	0.8660	0.8655	0.8507	0.9363
XGBoost	0.8339	0.8565	0.8561	0.8452	0.9318
LightGBM	0.8299	0.8597	0.8592	0.8448	0.9318
Traditional methods					
Inclusive SRR[g]	0.8964	0.6831	0.8926	0.7898	0.8699
Exclusive SRR	0.8733	0.7078	0.8703	0.7905	0.8224
EMR-ISS[h]	0.7874	0.7231	0.7863	0.7552	0.6171
KTAS[i]	0.9359	0.0121	0.9178	0.4740	0.1841
Others					
DNN using top 71 features from AdaBoost	0.7129	0.9362	0.9322	0.8245	0.9194

[a]AUROC: area under the receiver operating characteristic.

[b]ICD-10: International Classification of Disease 10th Revision.

[c]DNN: deep neural network.

[d]AdaBoost: Adaptive Boosting.

[e]XGBoost: Extreme Gradient Boosting.

[f]LightGBM: light gradient boosting machine.

[g]SRR: survival risk ratio.

[h]EMR-ISS: Mortality Ratio-adjusted Injury Severity Score.

[i]KTAS: Korean Triage and Acuity Scale.

Figure 4. Area under the receiver operating characteristic (AUROC) curves for our model, Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), and light gradient boosting machine (LightGBM): (left) using all 914 features including International Classification of Diseases 10th Revision (ICD-10), (middle) using 48 features excluding ICD-10, and (right) using 866 features with ICD-10 only. DNN: deep neural network.

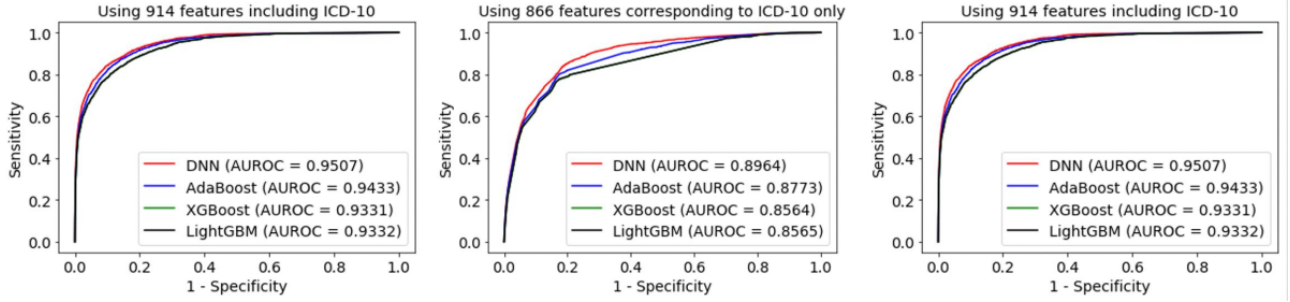
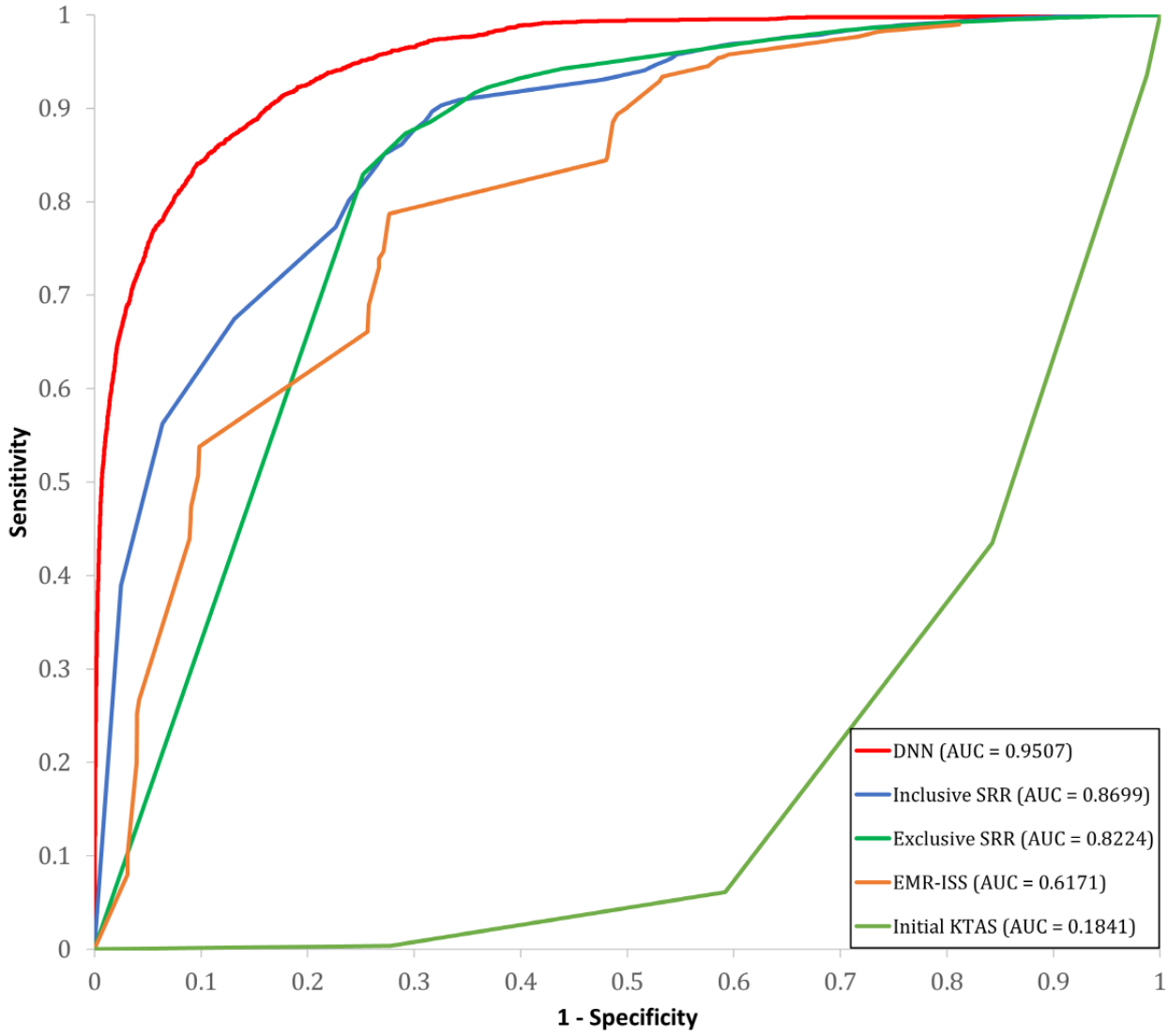


Figure 5. Area under the receiver operating characteristic (AUROC) curves of our model and 4 traditional models. AUC: area under the curve; DNN: deep neural network; EMR-ISS: mortality ratio-adjusted Injury Severity Score; KTAS: Korean Triage and Acuity Scale; SRR: survival risk ratio.



AI-Driven Public Website Development

We deployed our AI on a public website [27] to allow public access to the mortality prediction results in trauma patients (Figure S1 in Multimedia Appendix 1). Figure S1(a) shows a user's web interface to enter information. A user inputs age, gender, intentionality, injury mechanism, emergent symptoms, AVPU scale, initial KTAS, altered KTAS, torso procedures (chest, abdomen, vascular, and heart), head surgery, ECMO, and ICD-10 codes. Especially for ICD codes, a user can input multiple codes with a comma (eg, S072, S224, T083). As shown in Figure S1(b), after entering information in the web application, the user can immediately obtain the mortality results. The prediction results also include the probability of mortality.

Discussion

Principal Findings

Our AI model outperformed traditional ICD-10-based models and KTAS. Traditional methods produced high sensitivity and low specificity, with substantial bias in predicting mortality. Prediction performance was optimal when using all features, including ICD-10, as input features. The similarity between the cross-validation result and the testing data set indicates that overfitting or underfitting was minimal. In terms of ranked normalized feature importance, gender had the highest value, followed by age, coma, femur fracture, stupor, slip down, rib fracture, and head procedure. We used a population-based data set from all types of ED in South Korea, producing more robust and reliable results. To the best of our knowledge, our study is the first to demonstrate an AI model that drastically outperforms conventional ICD-based models and triage scales using a population-based data set. Our future goal is to construct a more comprehensive model incorporating both NEDIS-based and AIS-based AI [17].

Our proposed AI model has several advantages in clinical practice. First, a specialist is not required for AIS coding, so our AI model does not require additional burden. Second, our AI model demonstrates the ability to augment the KTAS provider's decision. Third, the feature importance used may benefit clinical decision-making and future research. Deep learning is generally considered a “black box,” hence the feature

importance analysis based on a machine learning algorithm provides meaningful insight to clinicians and researchers. Finally, we aspire for the global application of our model and have produced a publicly available web application for hospitals to utilize for the benefit of the entire trauma system [28,29].

Currently, ISS and ICISS are the most popular risk estimation models of trauma-related mortality. More complex models containing physiologic and demographic parameters are available [2,4,5,7], but none supersedes ISS or ICISS [1,9]. ISS is simple to use, but AIS coding is time consuming and expensive, whereas ICISS utilizes diagnostic code to claim charges. Therefore, ICISS is more useful for population-based data sets than ISS [8]. The results from ICISS in our study were comparable to those from previous studies [26,30]. We also applied EMR-ISS to the NEDIS data set, which showed good performance in a previous study [11] but poor accuracy here.

Recently, several AI models were proposed to predict trauma-related mortality. Previously, in a multicenter retrospective study in South Korea, we investigated a deep learning model using the AIS code for predicting mortality [17]. We reanalyzed the ISS system and redefine 46 new regions to discriminate the risk among different internal organs. The DNN with 46 features from the 46 new regions produced the highest accuracy. We found that the AI model can augment the performance of the AIS system. Recently, Tran et al [18] reported a machine-learning model that predicted trauma-related mortality using ICD-10. The authors used the NTDB data set and compared machine learning with ISS and TMPM10 [10], an ICD-10-based metric. However, the accuracy of each model was comparable. In this study, our AI model drastically outperformed ICISS and EMR-ISS. Kwon et al [31], in a retrospective observational study using a NEDIS data set including trauma and nontrauma patients, reported a deep learning-based model that showed a higher accuracy than KTAS for predicting in-hospital mortality. To the best of our knowledge, our AI model is the most accurate model and outperforms both diagnostic code-based metrics and triage scales in trauma patients.

Limitations and Future Works

Our study has several limitations. First, this is a retrospective study and may induce substantial selection and survival bias; further prospective trials and validation are needed. Second, we used procedure codes as 1 of the input features. However, they are not practically available during ED admission. Thus, in a prospective study, unconfirmed procedure codes may be used for predicting in-hospitality mortality. Third, in this study, we did not consider physiological signals, such as blood pressure, heart rate, and body temperature. We tried to train and develop an AI model using the information of physiological signals. However, the model's performance was poor because limited physiological signals were recorded in NEDIS; only blood pressure, heart rate, and temperature values at the time of admission were recorded. We believe that time-series physiological signals, such as electrocardiogram, photoplethysmogram, and blood pressure waveform, could improve our proposed model. Fourth, due to the structure of the NEDIS data set, some data, such as age, are collected as categorized data instead of continuous data. Thus, our proposed AI model could enhance the prediction performance with age as a continuous value. Fifth, some categorized input variables in the injury mechanism may appear inappropriate. For instance, the term "traffic accident-pedestrian, train, airplane, ship, etc" is considered 1 variable. However, pedestrians are not associated with an airplane and a ship. In addition, pedestrians have the highest mortality in road traffic collisions. Thus, the term should be separated into multiple variables. In future work, we plan to separate the variable into multiple categories and investigate the impact of each category. Sixth, we could not compare the prediction performances from our AI model with those from AIS code-based approaches such as ISS and NISS, as NEDIS does not provide AIS codes. Recently, we presented an AI model using AIS codes to predict in-hospital mortality [17]. The model outperformed conventional methods such as ISS and NISS for all accuracy metrics of sensitivity, specificity, balanced accuracy, and AUROC. As in the previous study, this study used ICD-10 and several clinical features instead of AIS codes and showed that the AI model outperformed conventional methods. Our goal is to construct a more comprehensive model incorporating both NEDIS-based and AIS-based AI models. Finally, our data did not include other races or data from other countries. Currently, our public website includes the following text: "This AI model was trained and evaluated from Korean trauma patients and may not be applicable to patients in other countries." Thus, future external validation is warranted, wherein we consider using global data to further improve our proposed AI model.

Conclusions

Our proposed AI model shows high accuracy and outperforms traditional diagnostic code-based prediction models and triage scales. We believe that our population-based AI model can facilitaite better understanding and practice in physical trauma care. Moreover, this AI and data-driven prediction model may minimize the bias and workload of humans. However, future external validation and prospective studies are warranted to prove the true effect size.

Acknowledgments

This work was supported by the Korea Medical Device Development Fund grant by the Korean government (KMDF_PR_20200901_0095). It was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF-2020R1A2C1014829) and by the Wonkwang Institute of Clinical Medicine in 2019.

Data Availability

The data are not publicly available due to restrictions from National Emergency Department Information System (NEDIS) policy and belong to the National Emergency Medical Center (NEMC) of Korea. NEMC provides deidentified NEDIS data to researchers for nonprofit academic research. Details for accessing the raw data and guide are available on the NEMC website.

Authors' Contributions

All the authors wrote the manuscript and created the figures. WSK and JL were responsible for the study concept and design. SSL and HK carried out the simulation. SSL, HK, WSK, and JL performed the statistical analysis. SSL, SHS, DWK, WSK, and JL interpreted the data. All authors critically reviewed and agreed to the submission of the final manuscript. SSL, WSK, SHS, and DWK contributed equally to this work and should be considered the first coauthors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables S1 to S6 and Figure S1.
[DOCX File , 463 KB - jmir_v24i12e43757_app1.docx]

References

1. Baker SP, O'Neill B, Haddon W, Long WB. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 1974 Mar;14(3):187-196. [Medline: 4814394]
2. Champion HR, Copes WS, Sacco WJ, Lawnick MM, Bain LW, Gann DS, et al. A new characterization of injury severity. J Trauma 1990 May;30(5):539-45; discussion 545. [doi: 10.1097/00005373-199005000-00003] [Medline: 2342136]
3. Kuo SCH, Kuo P, Chen Y, Chien P, Hsieh H, Hsieh C. Comparison of the new Exponential Injury Severity Score with the Injury Severity Score and the New Injury Severity Score in trauma patients: A cross-sectional study. PLoS One 2017;12(11):e0187871 [FREE Full text] [doi: 10.1371/journal.pone.0187871] [Medline: 29121653]
4. Macleod J, Kobusingye O, Frost C, Lett R. Kampala Trauma Score (KTS): is it a new triage tool? East Cent Afr J Surg 2007 Jan 01:12-297.
5. Moore L, Hanley JA, Turgeon AF, Lavoie A, Eric B. A new method for evaluating trauma centre outcome performance: TRAM-adjusted mortality estimates. Ann Surg 2010 May;251(5):952-958. [doi: 10.1097/SLA.0b013e3181d97589] [Medline: 20395844]
6. Moore L, Lavoie A, Le Sage N, Bergeron E, Emond M, Abdous B. Consensus or data-derived anatomic injury severity scoring? J Trauma 2008 Feb;64(2):420-426. [doi: 10.1097/01.ta.0000241201.34082.d4] [Medline: 18301208]
7. West TA, Rivara FP, Cummings P, Jurkovich GJ, Maier RV. Harborview assessment for risk of mortality: an improved measure of injury severity on the basis of ICD-9-CM. J Trauma 2000 Sep;49(3):530-40; discussion 540. [doi: 10.1097/00005373-200009000-00022] [Medline: 11003333]
8. Gagné M, Moore L, Sirois MJ, Simard M, Beaudoin C, Kuimi BLB. Performance of International Classification of Diseases-based injury severity measures used to predict in-hospital mortality and intensive care admission among traumatic brain-injured patients. J Trauma Acute Care Surg 2017 Feb 27;82(2):374-382 [FREE Full text] [doi: 10.1097/TA.0000000000001319] [Medline: 28107311]
9. Bergeron E, Simons R, Linton C, Yang F, Tallon JM, Stewart TC, et al. Canadian benchmarks in trauma. J Trauma 2007 Feb;62(2):491-497. [doi: 10.1097/01.ta.0000202483.67135.f3] [Medline: 17297340]
10. Osler TM, Glance LG, Cook A, Buzas JS, Hosmer DW. A trauma mortality prediction model based on the ICD-10-CM lexicon: TMPM-ICD10. J Trauma Acute Care Surg 2019 May;86(5):891-895. [doi: 10.1097/TA.0000000000002194] [Medline: 30633101]
11. Kim J, Shin SD, Im TH, Kug JL, Ko SB, Park JO, et al. Development and validation of the Excess Mortality Ratio-adjusted Injury Severity Score using the International Classification of Diseases 10th Edition. Acad Emerg Med 2009 May;16(5):454-464 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00412.x] [Medline: 19388920]
12. Ryu J, Min M, Lee D, Yeom S, Lee S, Wang I, et al. Changes in relative importance of the 5-Level Triage System, Korean Triage and Acuity Scale, for the disposition of emergency patients induced by forced reduction in its level number: a multi-center registry-based retrospective cohort study. J Korean Med Sci 2019 Apr 15;34(14):e114 [FREE Full text] [doi: 10.3346/jkms.2019.34.e114] [Medline: 30977315]
13. Lee JH, Park YS, Park IC, Lee HS, Kim JH, Park JM, et al. Over-triage occurs when considering the patient's pain in Korean Triage and Acuity Scale (KTAS). PLoS One 2019;14(5):e0216519 [FREE Full text] [doi: 10.1371/journal.pone.0216519] [Medline: 31071132]

14. Korndorffer JR, Hawn MT, Spain DA, Knowlton LM, Azagury DE, Nassar AK, et al. Situating artificial intelligence in surgery: a focus on disease severity. Ann Surg 2020 Sep 01;272(3):523-528. [doi: 10.1097/SLA.0000000000004207] [Medline: 33759839]
15. Madani A, Namazi B, Altieri MS, Hashimoto DA, Rivera AM, Pucher PH, et al. Artificial intelligence for intraoperative guidance: using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. Ann Surg 2022 Aug 01;276(2):363-369. [doi: 10.1097/SLA.0000000000004594] [Medline: 33196488]
16. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Ann Surg 2018 Jul;268(1):70-76 [FREE Full text] [doi: 10.1097/SLA.0000000000002693] [Medline: 29389679]
17. Kang WS, Chung H, Ko H, Kim NY, Kim DW, Cho J, et al. Artificial intelligence to predict in-hospital mortality using novel anatomical injury score. Sci Rep 2021 Dec 07;11(1):23534 [FREE Full text] [doi: 10.1038/s41598-021-03024-1] [Medline: 34876644]
18. Tran Z, Zhang W, Verma A, Cook A, Kim D, Burruss S, et al. The derivation of an International Classification of Diseases, Tenth Revision-based trauma-related mortality model using machine learning. J Trauma Acute Care Surg 2022 Mar 01;92(3):561-566. [doi: 10.1097/TA.0000000000003416] [Medline: 34554135]
19. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015 Jan 06;162(1):W1-73 [FREE Full text] [doi: 10.7326/M14-0698] [Medline: 25560730]
20. Kelly CA, Upex A, Bateman DN. Comparison of consciousness level assessment in the poisoned patient using the alert/verbal/painful/unresponsive scale and the Glasgow Coma Scale. Ann Emerg Med 2004 Aug;44(2):108-113. [doi: 10.1016/j.annemergmed.2004.03.028] [Medline: 15278081]
21. McNarry AF, Goldhill DR. Simple bedside assessment of level of consciousness: comparison of two simple assessment scales with the Glasgow Coma scale. Anaesthesia 2004 Jan;59(1):34-37 [FREE Full text] [doi: 10.1111/j.1365-2044.2004.03526.x] [Medline: 14687096]
22. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. JAIR 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
23. Mathanker S, Weckler P, Bowser T, Wang N, Maness N. AdaBoost classifiers for pecan defect classification. Comput Electron Agric 2011 Jun 28;77(1):60-68. [doi: 10.1016/j.compag.2011.03.008]
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Aug 13 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
25. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W. LightGBM: a highly efficient gradient boosting decision tree. 2017 Presented at: 31st International Conference on Neural Information Processing Systems; December 4-9; Long Beach, CA p. 3149-3157. [doi: 10.5555/3294996]
26. Kim Y, Jung KY. Utility of the international classification of diseases injury severity score: detecting preventable deaths and comparing the performance of emergency medical centers. J Trauma 2003 Apr;54(4):775-780. [doi: 10.1097/01.TA.0000028251.15629.1D] [Medline: 12707543]
27. In-hospital mortality prediction of trauma patients. AI Model. URL: http://ai-wm.khu.ac.kr/Trauma_Mortality [accessed 2022-12-04]
28. Lee HS, Sung WY, Lee JY, Lee WS, Seo SW. Comparative evaluation of emergency medical service trauma patient transportation patterns before and after level 1 regional trauma center establishment: a retrospective single-center study. J Trauma Inj 2021 Jun;34(2):87-97. [doi: 10.20408/jti.2020.0021]
29. Park Y, Lee GJ, Lee MA, Choi KK, Gwak J, Hyun SY, et al. Major causes of preventable death in trauma patients. J Trauma Inj 2021 Dec;34(4):225-232. [doi: 10.20408/jti.2020.0074]
30. Berecki-Gisolf J, Tharanga Fernando D, D'Elia A. International Classification of Disease-based Injury Severity Score (ICISS): a data linkage study of hospital and death data in Victoria, Australia. Injury 2022 Mar;53(3):904-911. [doi: 10.1016/j.injury.2022.01.011] [Medline: 35058065]
31. Kwon J, Lee Y, Lee Y, Lee S, Park H, Park J. Validation of deep-learning-based triage and acuity score using a large national dataset. PLoS ONE 2018 Oct 15;13(10):e0205836. [doi: 10.1371/journal.pone.0205836]

Abbreviations

AdaBoost: Adaptive Boosting
AI: artificial intelligence
AIS: Abbreviated Injury Scale
AUROC: area under the receiver operating characteristic
CTAS: Canadian Triage and Acuity Scale
DNN: deep neural network
ECMO: extracorporeal membrane oxygenation
ED: emergency department

EDNN: ensemble of deep neural networks
EMR-ISS: mortality ratio-adjusted Injury Severity Score
FC: fully connected
GCS: Glasgow Coma Scale
ICD: International Classification of Diseases
ICD-10: International Classification of Diseases 10th Revision
ICISS: International Classification of Diseases–based Injury Severity Score
ICU: intensive care unit
ISS: Injury Severity Score
KTAS: Korean Triage and Acuity Scale
LightGBM: light gradient boosting machine
NEDIS: National Emergency Department Information System
NTDB: National Trauma Database
Ps: probability of survival
SMOTE: Synthetic Minority Oversampling Technique
SRR: survival risk ratio
TMPM-ICD10: trauma mortality models using ICD-10
XGBoost: Extreme Gradient Boosting

Edited by G Eysenbach; submitted 23.10.22; peer-reviewed by B Reyes, Y Nam, D Chrimes, M Kapsetaki; comments to author 12.11.22; accepted 30.11.22; published 13.12.22.

Please cite as:
Lee S, Kang WS, Seo S, Kim DW, Ko H, Kim J, Lee S, Lee J
Model for Predicting In-Hospital Mortality of Physical Trauma Patients Using Artificial Intelligence Techniques: Nationwide Population-Based Study in Korea
J Med Internet Res 2022;24(12):e43757
URL: https://www.jmir.org/2022/12/e43757
doi:10.2196/43757
PMID:36512392

Original Paper

Examining the Use of an Artificial Intelligence Model to Diagnose Influenza: Development and Validation Study

Sho Okiyama[1,2], MD; Memori Fukuda[1,3], MD, MPH; Masashi Sode[1], MSc; Wataru Takahashi[1], MSc; Masahiro Ikeda[1,4], MD, PhD; Hiroaki Kato[1,5], MD, MBA; Yusuke Tsugawa[6], MD, PhD; Masao Iwagami[7,8], MD, MPH, MSc, PhD

[1]Aillis, Inc, Tokyo, Japan

[2]Japanese Red Cross Medical Center, Tokyo, Japan

[3]Department of Health Informatics, Graduate School of Medicine, School of Public Health, Kyoto University, Kyoto, Japan

[4]Department of Palliative Therapy, Cancer Institute Hospital of Japanese Foundation for Cancer Research, Tokyo, Japan

[5]Tokyo Medical and Dental University, Tokyo, Japan

[6]Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine at The University of California, Los Angeles, Los Angeles, CA, United States

[7]Department of Health Services Research, University of Tsukuba, Tsukuba, Japan

[8]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding Author:
Sho Okiyama, MD
Aillis, Inc
1-10-1-11F, Yurakucho
Chiyoda-ku
Tokyo, 100-0006
Japan
Phone: 81 3 5218 2374
Fax: 81 3 6800 1453
Email: sho.okiyama@aillis.jp

Abstract

Background: The global burden of influenza is substantial. It is a major disease that causes annual epidemics and occasionally, pandemics. Given that influenza primarily infects the upper respiratory system, it may be possible to diagnose influenza infection by applying deep learning to pharyngeal images.

Objective: We aimed to develop a deep learning model to diagnose influenza infection using pharyngeal images and clinical information.

Methods: We recruited patients who visited clinics and hospitals because of influenza-like symptoms. In the training stage, we developed a diagnostic prediction artificial intelligence (AI) model based on deep learning to predict polymerase chain reaction (PCR)–confirmed influenza from pharyngeal images and clinical information. In the validation stage, we assessed the diagnostic performance of the AI model. In additional analysis, we compared the diagnostic performance of the AI model with that of 3 physicians and interpreted the AI model using importance heat maps.

Results: We enrolled a total of 7831 patients at 64 hospitals between November 1, 2019, and January 21, 2020, in the training stage and 659 patients (including 196 patients with PCR-confirmed influenza) at 11 hospitals between January 25, 2020, and March 13, 2020, in the validation stage. The area under the receiver operating characteristic curve for the AI model was 0.90 (95% CI 0.87-0.93), and its sensitivity and specificity were 76% (70%-82%) and 88% (85%-91%), respectively, outperforming 3 physicians. In the importance heat maps, the AI model often focused on follicles on the posterior pharyngeal wall.

Conclusions: We developed the first AI model that can accurately diagnose influenza from pharyngeal images, which has the potential to help physicians to make a timely diagnosis.

(J Med Internet Res 2022;24(12):e38751) doi:10.2196/38751

KEYWORDS

influenza; physical examination; pharynx; deep learning; diagnostic prediction

Introduction

Background

According to the Global Burden of Disease Study 2016, the global burden of influenza is substantial. In the study, the disease was estimated to cause 39.1 million acute lower respiratory infection episodes and 58,200 deaths [1]. It has been estimated that influenza is responsible for 291,243 to 645,832 seasonal respiratory deaths (4.0-8.8 per 100,000 individuals) annually [2]. Timely and accurate diagnosis of influenza has the potential to prevent widespread transmission of the virus within the population, and subsequent epidemics and pandemics, in addition to the unnecessary prescription of antibiotics in primary care, which is a cause of emerging antibiotic-resistant bacteria. Moreover, early intervention, such as hydration and antiviral drugs, is expected to reduce the mortality risk among high-risk patients, including the older adults and individuals with comorbidities.

The COVID-19 pandemic and surge in the use of telemedicine highlighted the importance of accurately diagnosing influenza infection without increasing the risk of spreading the virus through physical interaction. The gold-standard method for diagnosing influenza infection is the reverse transcription–polymerase chain reaction (RT-PCR) of nasopharyngeal aspirates or swabs [3,4]; however, RT-PCR is not easily performed in primary care, and the result turnaround time could delay timely diagnosis and preventive or treatment interventions. A more commonly used test is the rapid immunochromatographic antigen detection test; however, its validity is modest compared with RT-PCR and varies across studies [5,6]. Neither of these tests can be performed through telemedicine, and the sensitivity and specificity of diagnosing influenza using clinical information only are suboptimal [7,8]. Given the recent increase in the number of patients being diagnosed through telemedicine, an alternative influenza test that can be conducted through telemedicine is warranted.

Objectives

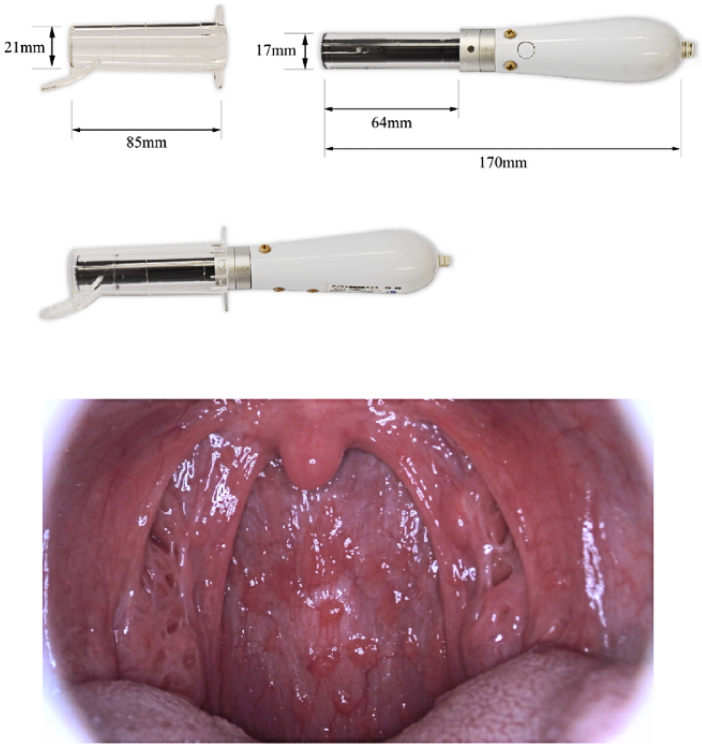
To address this important knowledge gap, we developed a deep learning model to diagnose influenza infection using pharyngeal images and clinical information. We tested the performance of the artificial intelligence (AI) model for diagnostic prediction using data from the real-world patient population and compared it with the diagnostic performance of 3 physicians. We also investigated the regions of the pharynx on which the AI model focused to differentiate between individuals with and without influenza infection.

Methods

Pilot Study to Develop a Medical Camera to Capture Standardized Pharyngeal Images

For our pilot study, we recruited 4765 patients aged 6 to 90 years with influenza-like symptoms, and they visited 37 clinics or hospitals between November 28, 2018, and February 4, 2019 (registered as jRCTs032180041). To capture images of the pharynx in a standardized manner, we developed a pharyngeal camera with a light-emitting diode light source and a disposable clear camera cover to hold down the tongue of patients (Figure 1). In this pilot study, we adjusted the size of the pharyngeal camera and tongue depressors to make them suitable for many patients. The device contained a full high-definition digital camera and was connected via Wi-Fi to a cloud service for the analysis of pharyngeal images, together with clinical information. During this pilot study, we improved the image quality of the camera in terms of resolution, brightness, and contrast. Specifically, we reduced the view angle appropriately to reduce distortion and improve the resolution because the angle was excessive. We also placed an imaging sensor near the tip of the camera to avoid light attenuation and ensure image brightness. In addition, we improved the image contrast by coating the clear camera cover with antifogging material to prevent fogging caused by exhalation. We used a rapid continuous shooting function to obtain high-quality pharyngeal images in a short time while avoiding motion blur. The camera could capture an image every 0.3 seconds, and 30 sequential images were captured per shot.

Figure 1. Presentation of the artificial intelligence–assisted camera and a representative pharyngeal image of a patient with polymerase chain reaction–confirmed influenza infection.



Study Design and Participants

This study included a training stage (registered as jRCTs032190120) and a validation stage (registered as Pharmaceuticals and Medical Devices Agency clinical trial identification code AI-02-01). We enrolled patients with influenza-like symptoms who visited clinics or hospitals and satisfied the following inclusion and exclusion criteria at 64 hospitals between November 1, 2019, and January 21, 2020, in the training stage, and 11 hospitals between January 25, 2020, and March 13, 2020, in the validation stage. A list of study sites is provided in Table S1 in Multimedia Appendix 1.

The inclusion criteria were as follows: (1) patients who provided written consent by themselves or their parents (if they were aged <18 years) to participate in the study, (2) those aged ≥6 years, and (3) those who satisfied at least one of the following 4 conditions in the training stage and at least 2 in the validation stage: first, body temperature of ≥37.0 °C; second, systematic influenza-like symptoms, such as joint pain, muscle pain, headache, tiredness, and appetite loss; third, respiratory symptoms, such as cough, sore throat, and nasal discharge or congestion; and fourth, an episode of close contact with patients with influenza or influenza-like symptoms within 3 days, or in any other scenario in which the consulting physician suspected influenza infection. The exclusion criteria included the following: (1) patients with fluctuating teeth; (2) those with severe oral lesions; (3) those with severe nausea; (4) those with difficulty in opening the mouth sufficiently for the use of the camera (eg, small mouth, temporomandibular joint pain, incompatibility of dentures, disturbed consciousness, or respiratory failure); (5) those who had participated in another clinical trial within 7 days before this study, those who were scheduled to participate in another clinical trial (excluding postmarketing surveillance), or those with difficulty in follow-up owing to mental, family, social, geographic, or other reasons; (6) pediatric patients who clearly did not agree to participate in the study; and (7) those judged to be inappropriate to participate in the study by the responsible physician at each site. In addition, we excluded patients with only poor-quality images from the analysis.

In the training stage, we aimed to collect clinical information and pharyngeal images from patients with RT-PCR–confirmed influenza-positive and influenza-negative results in a ratio of approximately 1:1 to enable the most efficient supervised learning of the AI model. There is no consensus on the size of the samples (ie, number of patients) that should be used to train an AI model; thus, we arbitrarily set the size to 7500 patients, including 3750 patients with RT-PCR–positive results and 3750 patients with RT-PCR–negative results. In the validation stage, we aimed to determine the lower bound of the 95% one-sided CI of sensitivity to achieve ≥70% and that of specificity to achieve ≥85%. With a 1-sided *P* value of 5% and power of 85%, assuming an actual sensitivity of 80% and specificity of 90% as suggested by our training stage, we calculated the required sample sizes to be 137 for patients with RT-PCR–positive results and 323 for RT-PCR–negative results. Therefore, we planned to stop the recruitment of study participants on the day when 150 patients with positive results and 350 patients with negative results were obtained.

In Japan, the first case of SARS-CoV-2 infection (COVID-19) was reported on January 15, 2020, and the first wave of the pandemic occurred in late March 2020. During the study period, in the validation stage, we asked the participating clinics and hospitals to report any suspected cases of COVID-19 in the study participants. There were no such reports from any study

site throughout the study, which suggests that our study sample was not affected by the COVID-19 pandemic.

Collection of Pharyngeal Images, Clinical Information, and Nasopharyngeal Specimens

In addition to the pharyngeal images of the study participants, the following clinical information was obtained using a standardized case report form based on electronic data capture: age; sex; time (hours) from symptom onset; highest body temperature before study site visit; episode of close contact; status and date of the most recent influenza vaccination; use of antipyretics; subjective symptoms, including tiredness, appetite loss, chill, sweating, joint pain, muscle pain, headache, nasal discharge or congestion, cough, sore throat, and digestive symptoms; and objective findings by the consulting physicians at study sites, including body temperature, pulse rate, and tonsillar findings (tonsillitis, white moss, and redness).

Furthermore, nasopharyngeal swabbing was conducted to obtain nasopharyngeal specimens from the participants, which were sent to the central clinical laboratory (LSI Medience Corporation) for RT-PCR testing, which is the gold standard (reference standard) for the diagnosis of influenza infection. We standardized the process of collecting the nasopharyngeal specimens among the study sites using our own Japanese manual (not publicly available).

Development of the AI Model to Predict RT-PCR–Confirmed Influenza

We developed an ensemble AI model (version FLU2021.06) to predict the probability of RT-PCR–confirmed influenza using pharyngeal images and clinical information (Figure S1 in Multimedia Appendix 1). This model consists of 3 main machine learning models: a multiview convolutional neural network (MV-CNN), a multimodal convolutional neural network (MM-CNN), and boosting models. In the training stage, we trained these 3 types of machine learning models and integrated them using ridge regression [9] into the ensemble AI model.

First, we trained the MV-CNN using SE-ResNext-50 as an image feature extractor, which was pretrained on ImageNet [10,11]. The MV-CNN architecture used several pharyngeal images that contained views from various angles [12]. On pharyngeal imaging, the tongue and uvula often overlap with the posterior pharyngeal wall. The MV-CNN addressed this issue by gathering information from various image angles. From 30 (or more if several shots were taken) sequential images, 1 to 5 of the most appropriate images per patient were selected as the input to the MV-CNN using an automatic image quality evaluation system. We determined the number of input images by considering the MV-CNN performance and the memory size limitation of the graphics processor units. Although, in general, the MV-CNN performs better with more input images, the memory size of the graphics processor unit constrains the number of images. If the number of selected images was <5, we padded them with uninformative images filled with zeros, similar to zero padding in the boundary region of an image. To quantify the visual image quality criteria, we trained the image quality evaluation system that used a lightweight CNN model [13] in the training stage using human-annotated visual image quality criteria (eg, visibility of the posterior pharyngeal wall, brightness, focus, motion blur, and exhalation fog) defined by one of the authors (MF) who is a physician. The input images for the MV-CNN were resized and then augmented (eg, flipped, rotated, blurred, and contrast-changed) to improve the accuracy and generalization performance. To prevent overfitting, we used well-established training strategies, including batch normalization, learning rate decay, and cross-validation. To manage various pharyngeal magnification rates, we trained the MV-CNNs with multiple image sizes and combined their scores by averaging them.

Second, we developed the MM-CNN based on the MV-CNN to process both multiview pharyngeal images and clinical information as input data [14,15]. In detail, we extended the final classification layer of the MV-CNN and connected it to the neural network to manage clinical information. The image feature extractor of the MM-CNN was initialized using trained MV-CNN weights. Then, we applied the same training and ensemble strategies as those used for the MV-CNN.

Third, we trained boosting models based on the prediction results of MV-CNN and clinical information. We selected LightGBM and CatBoost as the boosting models [16,17]. Finally, the probability of influenza was obtained by integrating each prediction from the MV-CNN, MM-CNN, and boosting models using ridge regression. We trained the ridge regression weights using cross-validation. The probability of influenza was obtained by averaging all the folds of the ridge regression model predictions.

Statistical Analysis

In the training stage, we compared the clinical characteristics of the study participants according to the RT-PCR test results (positive or negative) using *t* tests (2 tailed) for continuous variables with a normal distribution (age, highest body temperature before the study site visit, body temperature at visit, and pulse rate), Mann-Whitney *U* test for continuous variables with a nonnormal distribution (time from symptom onset), and chi-square tests for categorical variables. We repeated these analyses in the validation stage.

In the training stage, using a 5-fold cross-validation method, we conducted a receiver operating characteristic (ROC) curve analysis to measure the discrimination ability of (1) the probability score of the MV-CNN, which uses only pharyngeal images in the prediction; (2) the probability score of the clinical information AI, which is an AI model that uses all the aforementioned clinical information, except for the pharyngeal images, in the prediction; and (3) the probability score of the ensemble AI model using both the pharyngeal images and clinical information. We also measured the reclassification ability of the pharyngeal images by comparing the clinical information AI model and the ensemble AI model by calculating the continuous reclassification improvement and integrated discrimination improvement [18].

In the validation stage, we also conducted ROC analysis and calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for influenza infection, according to a selected cutoff point.

We performed statistical analysis using R software (version 4.1.1; R Foundation) and Python software (version 3.8.5; Python Software Foundation). *P* values of <.05 were considered statistically significant. A third-party organization (Statcom Co Ltd, Tokyo, Japan) performed the sample size estimation and calculation of the area under the ROC curve (AUROC) and validity (sensitivity, specificity, PPV, and NPV) in the validation stage. To avoid the post hoc adjustment of the developed AI model to fit the validation data in the regulatory approval process, the authors were prohibited from directly touching the validation data or conducting additional analyses in the validation stage. Therefore, any other analyses (eg, the calculation of AUROC for pharyngeal images and clinical information independently or for the MV-CNN, MM-CNN, and boosting model separately) were not possible in the validation stage.

Additional Analysis

We conducted 4 types of additional analyses. First, we compared the performance of the AI-assisted diagnostic camera with that of the 3 physicians. For this analysis, we used the existing data (pharyngeal images and clinical information) of 200 patients (100 patients with RT-PCR–positive results and 100 with RT-PCR–negative results), which were randomly selected from study participants in the training stage. A total of 3 physicians among the authors (SO, MF, and M Ikeda), who were blinded to the patients' identifiers and their RT-PCR test results, assessed the data to assign an influenza prediction score between 0 and 1 (ie, between 0% and 100%). As there is generally no established practice or criteria for physicians to diagnose influenza from pharyngeal images and clinical information, the 3 physicians were asked to guess the probability of influenza infection for each patient, as they usually do in their actual clinical practice. We applied the diagnostic prediction AI model to the existing data and compared the AUROC of the diagnostic prediction AI model with that of each physician and the average prediction score of the 3 physicians. We recalculated the AUROC of the AI model for the 200 patients for a fair comparison.

Second, we attempted to interpret the mechanisms of the MV-CNN prediction to differentiate between influenza cases and noninfluenza cases using pharyngeal images. We modified the guided gradient-weighted class activation mapping for the MV-CNN to visualize the importance heat maps. The aim was to determine the focus area of MV-CNN when differentiating between patients with RT-PCR–positive and RT-PCR–negative results. We used the same data set of 200 patients (100 patients with RT-PCR–positive and 100 with RT-PCR–negative results) that we used in the first additional analysis. To quantify and interpret the importance heat maps, 2 physicians among the authors (MF and M Ikeda) independently determined whether the MV-CNN highlighted each part of the pharynx (classified into 5 parts: lateral pharyngeal bands, posterior pharyngeal wall, palatal arch, tonsils, and follicles) for each patient. When the 2 physicians made different judgments (ie, presence vs absence of highlighting by the MV-CNN), a consensus was reached through discussion between them. Consequently, for each part of the pharynx, we calculated the proportion of patients with images highlighted by the MV-CNN among the 100 patients with RT-PCR–positive results and 100 RT-PCR–negative results and compared the groups using chi-square tests.

Third, as a post hoc experiment, using the 200 samples, we compared the performance of our final model (ie, the ridge regression ensemble model) with the performance of each of the component models: the MV-CNN, MM-CNN, and boosting models.

Finally, as another post hoc experiment, we compared the performance of the MV-CNN model with the proposed backbone (SE-ResNext-50) and that of various CNN backbones, that is, ResNet-50, ResNeXt-50 (32×4d), EfficientNet-B0, and DenseNet-121, which were available at the time of our model development.

Ethics Approval

The ethics committee of Hattori Clinic approved the pilot study and the training study, and the validation study was approved by the ethics committee of Takahashi Clinic, Kobori Central Clinical, and Haradoi Hospital.

Results

Training Stage

Figure S2 in Multimedia Appendix 1 shows the flowchart of patient selection during the training stage. We obtained informed consent from 9029 patients with influenza-like symptoms who visited one of 64 clinics or hospitals between November 1, 2019, and January 21, 2020. Among them, 199 patients (2.20%) experienced nausea during the examination when pharyngeal images were being captured, including 1 (0.01%) patient with severe nausea and 14 (0.16%) patients who vomited. We did not complete the image-capturing procedure for these 15 patients (0.17%). Among the remaining 9014 patients, we selected 7831 patients (mean age 33.8, SD 18.4 years; female patients: 3901/7831, 50%) with 25,168 high-quality images (out of approximately 300,000 images), which consisted of 3733 (47.67%) patients with influenza RT-PCR–positive results with 12,154 (48.29%) pharyngeal images and 4098 (52.33%) patients with RT-PCR–negative results with 13,014 (51.71%) pharyngeal images. Table 1 compares the clinical characteristics of the patients based on the RT-PCR test results. Compared with the patients with RT-PCR–negative results, the patients with RT-PCR–positive results yielded the following: the average age was slightly lower; the time from symptom onset to the study site visit was shorter; the proportion of close contact, use of antipyretics, and most subjective symptoms were higher; and the temperature and pulse rate were higher, whereas the proportion of recent influenza vaccinations, digestive symptoms, and tonsillar findings were lower. There was no difference in the proportions of sex and sore throat between the groups.

Using the training data set, we established the ensemble AI model to estimate the probability of influenza in individual patients. The feature importance of each variable in the LightGBM and CatBoost models is shown in Figures S3 and S4 in Multimedia Appendix 1, which suggest that pharyngeal images were the most important variable in the diagnostic prediction AI model, followed by body temperature and cough.

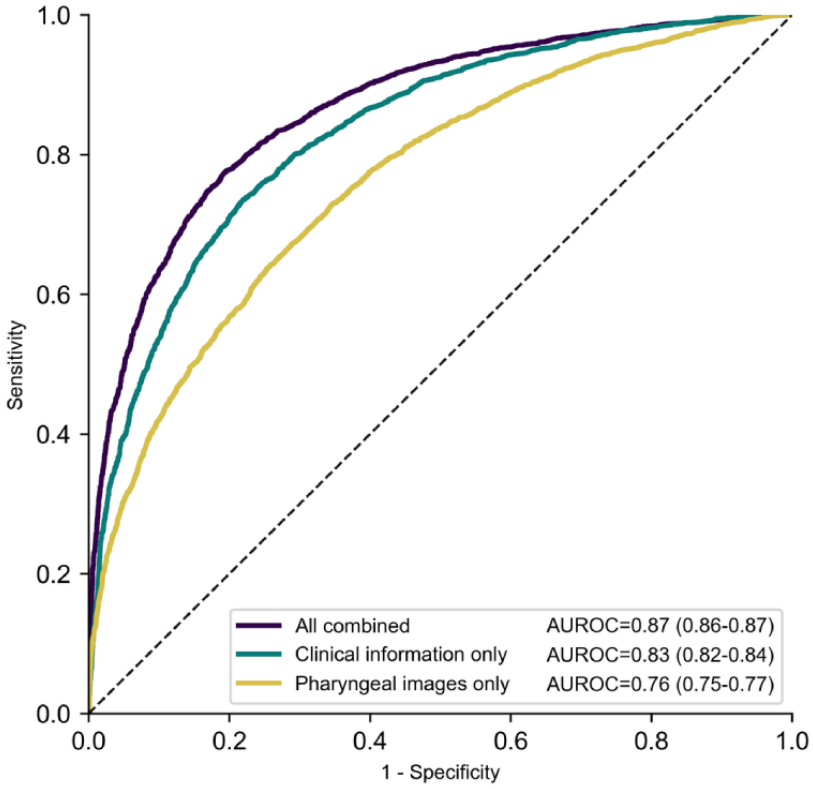
In the 5-fold cross-validation, the AUROC of the MV-CNN probability score for pharyngeal images was 0.76 (95% CI 0.75-0.77) and that of the AI model with clinical information (ie, all the clinical information in Table 1) was 0.83 (95% CI 0.82-0.84; Figure 2). The AUROC of the diagnostic prediction AI model with pharyngeal images and clinical information was 0.87 (95% CI 0.86-0.87), which means that the AUROC significantly increased because of the addition of pharyngeal images to the AI model with clinical information (*P*<.001). Regarding reclassification ability, the continuous reclassification improvement was 0.25 (95% CI 0.22-0.29) among patients with RT-PCR–positive results and 0.33 (95% CI 0.30-0.36) among patients with RT-PCR–negative results and the integrated discrimination improvement was 0.08 (95% CI 0.07-0.08), which also indicate that the accuracy of the diagnostic prediction AI model significantly improved because of the addition of pharyngeal images to the AI model with clinical information.

Table 1. Characteristics of the study participants with or without reverse transcription–polymerase chain reaction (RT-PCR)–confirmed influenza.

Characteristics	Participants in the training stage				Participants in the validation stage			
	All (n=7831)	RT-PCR test result			All (n=659)	RT-PCR test result		
		Positive (n=3733)	Negative (n=4098)	*P* value		Positive (n=196)	Negative (n=463)	*P* value
Age (years), mean (SD)	33.8 (18.4)	33.0 (18.5)	34.5 (18.4)	<.001	33.3 (17.6)	30.4 (18.6)	34.5 (17.0)	.008
Sex , n (%)				.54				.54
Male	3930 (50.2)	1887 (50.5)	2043 (49.9)		318 (48.3)	91 (46.4)	227 (49.0)	
Female	3901 (49.8)	1846 (49.5)	2055 (50.1)		341 (51.7)	105 (53.6)	236 (51.0)	
Time from onset (hours), mean (SD)	31.2 (25.3)	28.3 (20.6)	33.8 (28.6)	<.001	27.5 (31.2)	24.6 (10.8)	28.7 (36.5)	.67
Highest BT[a] before visit (°C), mean (SD)	38.2 (0.9)	38.6 (0.8)	38.0 (0.9)	<.001	38.2 (0.8)	38.6 (0.7)	38.0 (0.8)	<.001
Close contact, n (%)	2520 (32.2)	1687 (45.2)	833 (20.3)	<.001	208 (31.6)	120 (61.2)	88 (19.0)	<.001
Recent influenza vaccination, n (%)	2873 (36.7)	1248 (33.4)	1625 (39.7)	<.001	278 (42.2)	73 (37.2)	205 (44.3)	.09
Use of antipyretics, n (%)	2975 (38)	1530 (41)	1445 (35.3)	<.001	297 (45.1)	95 (48.5)	202 (43.6)	.25
Subjective symptoms , n (%)								
Tiredness	5937 (75.8)	3010 (80.6)	2927 (71.4)	<.001	506 (76.8)	159 (81.1)	347 (74.9)	.09
Appetite loss	3361 (42.9)	1823 (48.8)	1538 (37.5)	<.001	259 (39.3)	96 (49)	163 (35.2)	<.001
Chill	4215 (53.8)	2231 (59.8)	1984 (48.4)	<.001	338 (51.3)	115 (58.7)	223 (48.2)	.01
Sweating	2188 (27.9)	1128 (30.2)	1060 (25.9)	<.001	206 (31.3)	60 (30.6)	146 (31.5)	.82
Joint pain	3735 (47.7)	1992 (53.4)	1743 (42.5)	<.001	316 (48)	103 (52.6)	213 (46)	.12
Muscle pain	2362 (30.2)	1276 (34.2)	1086 (26.5)	<.001	192 (29.1)	62 (31.6)	130 (28.1)	.36
Headache	4725 (60.3)	2414 (64.7)	2311 (56.4)	<.001	403 (61.2)	126 (64.3)	277 (59.8)	.28
Nasal discharge or congestion	4472 (57.1)	2202 (59)	2270 (55.4)	.001	410 (62.2)	134 (68.4)	276 (59.6)	.03
Cough	5219 (66.6)	3053 (81.8)	2166 (52.9)	<.001	384 (58.3)	161 (82.1)	223 (48.2)	<.001
Sore throat	4928 (62.9)	2353 (63)	2575 (62.8)	.86	440 (66.8)	126 (64.3)	314 (67.8)	.38
Digestive symptoms	1298 (16.6)	558 (14.9)	740 (18.1)	<.001	127 (19.3)	30 (15.3)	97 (21)	.09
Objective findings								
BT at visit (°C), mean (SD)	37.6 (0.9)	38.0 (0.9)	37.3 (0.8)	<.001	37.5 (0.9)	37.9 (0.9)	37.3 (0.8)	<.001
Pulse rate, mean (SD)	95.0 (17.8)	100.2 (17.7)	90.3 (16.6)	<.001	93.8 (17.7)	100.8 (18.6)	90.9 (16.4)	<.001
Tonsillitis, n (%)	1238 (15.8)	529 (14.2)	709 (17.3)	<.001	63 (9.6)	8 (4.1)	55 (11.9)	.002
Tonsillar white moss, n (%)	126 (1.6)	17 (0.5)	109 (2.7)	<.001	23 (3.5)	1 (0.5)	22 (4.8)	.007
Tonsillar redness, n (%)	1292 (16.5)	540 (14.5)	752 (18.4)	<.001	69 (10.5)	13 (6.6)	56 (12.1)	.04

[a]BT: body temperature.

Figure 2. Receiver operating characteristic curves of the diagnostic prediction models in the 5-fold cross-validation of the training data set. In the figure, all combined indicates ensemble artificial intelligence (AI) model using pharyngeal images and clinical information; pharyngeal images only indicates multiview convolutional neural network using multiple pharyngeal images; clinical information only indicates ensemble AI model without pharyngeal image information. AUROC: area under the receiver operating characteristic.



Validation Stage

Figure S5 in Multimedia Appendix 1 shows the flowchart of patient selection during the validation stage. In the validation stage, we obtained informed consent from 706 patients with influenza-like symptoms who visited one of 11 clinics or hospitals between January 25, 2020, and March 13, 2020, which comprised a safety analysis set. Of the 706 patients, 12 (1.7%) felt nauseous during the examination when the pharyngeal images were being captured, including 1 patient (0.1%) with severe nausea for whom we did not complete the image-taking procedure. In addition, 33 patients (4.7%) did not satisfy the predefined criteria of the protocol for the full analysis set, mostly because of the difficulties in saving pharyngeal images at the study sites. Furthermore, 13 (1.8%) patients were excluded from the automated image quality evaluation system that removed low-quality pharyngeal images. Thus, we used the pharyngeal images and clinical information of the remaining 659 patients (mean age 33.3 years, SD 17.6 years; female patients: 341/659, 51.7%) for the validation stage analysis. Similar to the training stage, compared with noncases, the RT-PCR–confirmed cases yielded the following results: the average age was slightly lower; the proportion of close contact and several subjective symptoms (tiredness, chills, nasal discharge or obstruction, and cough) was higher; and the temperature (both before the clinic or hospital visit and on site) and pulse rate were higher, whereas the proportion of tonsillar findings was lower (Table 1).

In the validation stage, the AUROC of the diagnostic prediction AI model was 0.90 (95% CI 0.87-0.93). At a selected cutoff point on the ROC curve (Figure 3), the sensitivity and specificity were 76% (95% CI 70%-82%) and 88% (95% CI 85%-91%), respectively, and the PPV and NPV were 73% (95% CI 69%-79%) and 90% (95% CI 87%-92%), respectively (Table 2).

Figure 3. Receiver operating characteristic curve for the diagnostic prediction model in the validation data set. AI: artificial intelligence; AUROC: area under the receiver operating characteristic curve.

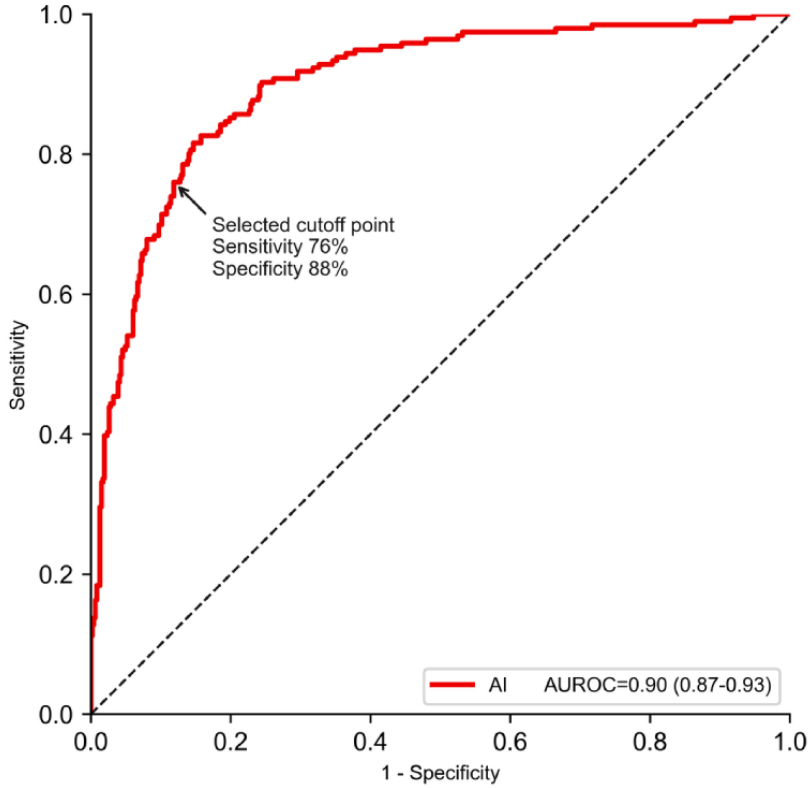


Table 2. Validity of the artificial intelligence (AI)–assisted device compared with the gold-standard diagnosis of influenza virus infection based on reverse transcription–polymerase chain reaction (RT-PCR).

	Influenza virus infection based on RT-PCR		Total, n	Values, % (95% CI)	
	True positive	True negative		PPV[a], % (95% CI)	NPV[b], % (95% CI)
Prediction by the AI-assisted device[c], n					
Positive	149	55	204	73 (67-79)	N/A[d]
Negative	47	408	455	N/A	90 (87-92)
Total, n	196	463	659	N/A	N/A
Sensitivity, % (95% CI)	76 (70-82)	N/A	N/A	N/A	N/A
Specificity, % (95% CI)	N/A	88 (85-91)	N/A	N/A	N/A

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]According to the selected cutoff point on the receiver operating characteristic curve of the diagnostic prediction model of the AI-assisted device shown in Figure 3.

[d]N/A: not applicable.

Additional Analysis

In our additional analysis, among the 200 randomly selected patients (100 patients with RT-PCR–positive results and 100 with RT-PCR–negative results), the AUROC of the diagnostic prediction AI model was 0.89 (95% CI 0.84-0.93), which was higher than that of each of the 3 physicians (0.76, 0.73, and 0.74). It was also higher than that of the average prediction score of the 3 physicians (0.79, 95% CI 0.73-0.85; Figure 4).

Figure S6 in Multimedia Appendix 1 and Figure 5 show examples of the pharyngeal images and those highlighted using the importance heat maps. An assessment of the importance heat maps for the 200 patients (100 patients with RT-PCR–positive results and 100 with RT-PCR–negative results) conducted by 2 physicians showed that the proportion of patients with AI model–highlighted images of follicles on the posterior pharyngeal wall was significantly different between the patients with RT-PCR–positive and RT-PCR–negative results (73% vs 38%; $P<.001$), which suggests that the AI model often focused on these parts (Figure S7 in Multimedia Appendix 1).

Finally, our post hoc experiments suggested that the performance of our final model (ie, the ridge regression ensemble model) was superior or similar to (at least not inferior to) the performance of each component model (Table S2 in Multimedia Appendix 1). In addition, the backbone model proposed in our final model was superior to various CNN backbones (Table S3 in Multimedia Appendix 1).

Figure 4. Receiver operating characteristic curves for the diagnostic prediction artificial intelligence model and 3 physicians. In the figure, AI indicates ensemble AI model using pharyngeal images and clinical information. The AI model was the same as that used in the validation stage. However, the AUROC was slightly different because of the small sample size used in the additional analysis. AI: artificial intelligence; AUROC: area under the receiver operating characteristic curve. 3 physicians: average prediction score of the 3 physicians.

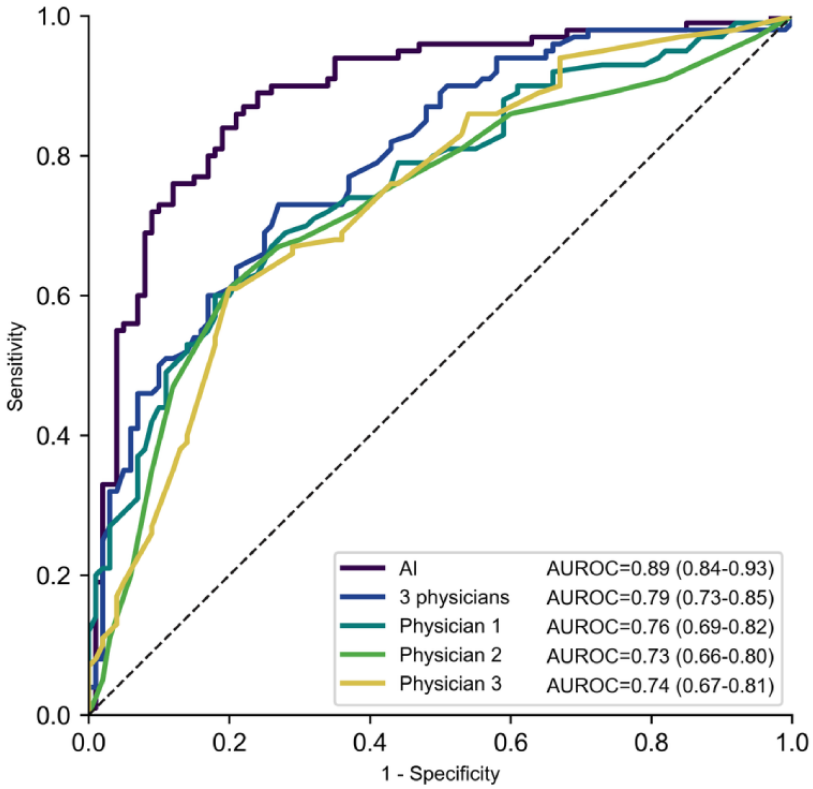
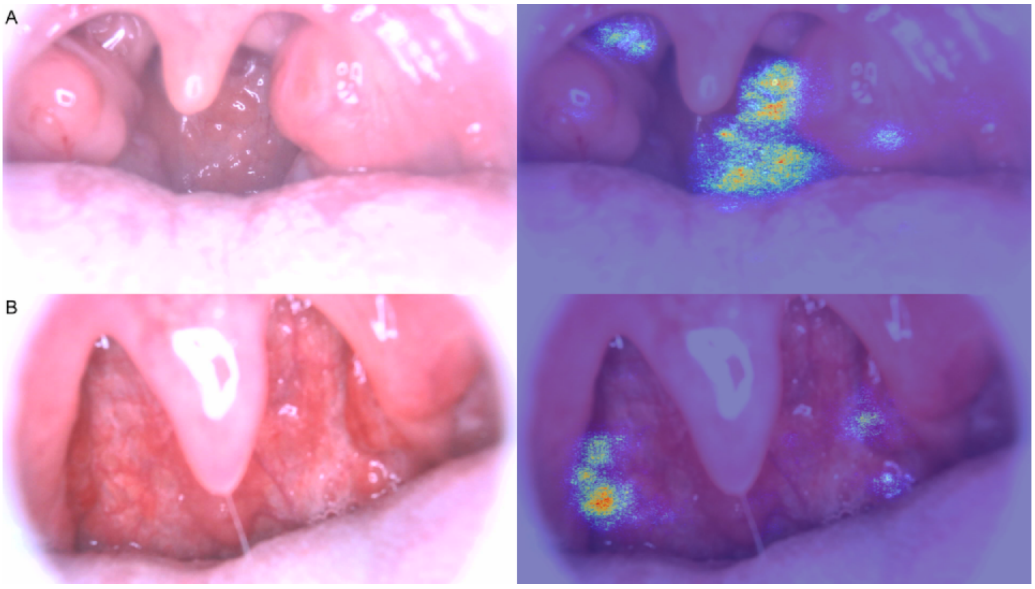


Figure 5. Examples of pharyngeal images (left) and those highlighted using the importance heatmaps (right). These importance heat maps show areas in which the artificial intelligence (AI) model focused on differentiating between reverse transcriptase–polymerase chain reaction (RT-PCR)–positive cases and RT-PCR–negative cases. In example A, the AI model focused on follicles. In example B, the AI model focused on the lateral pharyngeal bands.



Discussion

Principal Findings

In this study, we developed an AI-assisted diagnostic camera using a diagnostic prediction model for influenza (Multimedia Appendix 2). In the training stage, we found that the pharyngeal images contributed significantly to the improvement of the diagnostic prediction AI model compared with the clinical information AI model. In the validation stage, the AUROC of the diagnostic prediction AI model was 0.90 (95% CI 0.87-0.93), with a sensitivity and specificity of 76% (95% CI 70%-82%) and 88% (95% CI 85%-91%), respectively. In our additional analysis, the AI-assisted camera performed better than the 3 physicians in predicting influenza. Furthermore, in the importance heat maps, we found that the AI model often focused on follicles to differentiate between patients with RT-PCR–positive and RT-PCR–negative results.

Clinical characteristics associated with RT-PCR–confirmed influenza infection among people with influenza-like symptoms were examined in 2 previous studies [7,8]. Both studies concluded that fever and cough were the best predictors of influenza diagnosis. However, the sensitivity and specificity of the combination of these 2 factors were suboptimal, at 78% and 55% in one study [7] and 64% and 67% in another study, respectively [8]. In our study, considering the feature importance of each variable in the LightGBM and CatBoost models (Figures S3 and S4 in Multimedia Appendix 1), body temperature and cough were highly ranked among clinical information, whereas the feature importance of pharyngeal images was even larger than the highly ranked clinical information.

Recently, several AI-assisted diagnostic prediction models have been proposed for influenza diagnosis [19-22]. In a single-center study from Japan, researchers reported a machine learning–based infection screening system that incorporates a random tree algorithm that uses vital signs [19]. The researchers reported a sensitivity of 81% to 96% and NPV of 81% to 96% in their training data sets (they did not report specificity and PPV); however, they did not validate the performance of the model outside the center. Researchers at the University of Pittsburgh Medical Center Health System reported machine learning classifiers for influenza detection from free-text reports of the emergency department [20,21]. Among the 31,268 emergency department reports from 4 hospitals, the AUROCs of the 7 machine learning classifiers for influenza detection ranged from 0.88 to 0.93 [21], which was better than an expert-built Bayesian model [20]. These studies were also limited because performance outside the health care system of the University of Pittsburgh was unknown. More recently, a Korean study reported an influenza screening system based on deep learning using a combination of epidemiological and patient-generated health data from a mobile health app [22]. However, the gold standard in the study was the clinical diagnosis of influenza at a clinic reported by app users instead of laboratory-confirmed influenza. Notably, none of the previous studies included an assessment of pharyngeal images in their diagnostic prediction models [19-22]. The novelty of our study is that we have developed the first AI-assisted diagnostic camera for influenza and prospectively validated its performance through a Good Clinical Practice-based clinical trial process.

We showed that pharyngeal images significantly improved the discrimination and reclassification abilities of the diagnostic prediction AI model. In addition, we considered the mechanisms by which the AI model differentiated between true influenza cases and noninfluenza cases using pharyngeal images. To the best of our knowledge, there is no established approach to quantitatively scale the regions of images on which the AI model focuses. Indeed, most previous studies on AI-assisted diagnostic cameras showed only representative images highlighted using gradient-weighted class activation mapping or saliency maps to speculate on the possible mechanisms of AI classification [23-25]. In our study, we attempted to quantify these regions by calculating the proportion of patients with images highlighted by the AI model for each part of the pharynx among the patients with RT-PCR–positive and RT-PCR–negative results. Consequently, we found that the AI model mainly focused on follicles on the posterior pharyngeal wall. Notably, this finding is in line with previous case reports and case series that suggest that the follicles on the posterior pharyngeal wall are specific to influenza infection and are useful for the diagnosis of influenza [26-29]. Physical examination, including visual inspection of the pharynx, generally requires the experience of individual physicians, and physical examination skills may vary widely among physicians. Our study suggests that AI could minimize the variation and may help to standardize physical examination skills among physicians. In addition, when attempting to discriminate between diseases, doctors may be able to learn where to focus on their visual examination using AI systems.

Limitations

Our study has some limitations. First, we recruited participants with influenza-like symptoms from a large number of clinics and hospitals in Japan to increase the generalizability of our study. However, there may be a country or cultural difference in terms of people with influenza-like symptoms seeking medical care from health care providers. In Japan, with its universal health care coverage, people have relatively easy and timely access to clinics and hospitals compared with those in other countries. Therefore, generalizing our findings to different clinical care settings in different countries requires caution and independent assessment. Second, our additional analysis of the comparison between the AI-assisted diagnostic camera and the 3 physicians was not planned in the study protocols (jRCTs032190120 and Pharmaceuticals and Medical Devices Agency clinical trial identification code AI-02-01); however, these physicians were blinded to the patients' identifiers and their RT-PCR results. Finally, in addition to pharyngeal images, we collected as many relevant clinical variables (suggested in previous large studies [7,8]) as possible to establish an accurate diagnostic prediction AI model. However, there may be other useful variables for the prediction of true influenza diagnosis that we did not collect in our study. For example, in some studies, researchers have suggested that the population-level trend of influenza outbreaks in an area is useful for predicting an individual patient's influenza infection [22]. Further improvement of the AI-assisted diagnostic camera by including

additional variables, as well as an improvement of the AI models to analyze pharyngeal images, is justified.

Conclusions

In conclusion, we developed the first AI-assisted diagnostic camera for influenza and prospectively validated its high performance. We found that the AI model often focused on follicles, which confirmed previous case reports and series suggesting that visual inspection of the pharynx would help in the diagnosis of influenza infection.

Acknowledgments

The authors would like to thank all the study participants, medical staff at the research facilities, and physicians, particularly Dr Jun Fukuda. They would also like to thank Shuhei Fujimoto and Tetsuro Oda for their statistical advice and Dr Akihiko Miyamoto and Dr Shigeyuki Watanabe for their clinical advice. The authors trained the artificial intelligence model on the AI Bridging Cloud Infrastructure of the National Institute of Advanced Industrial Science and Technology, Japan. The authors would also like to thank A2 Healthcare Corporation, Tokyo, Japan, for assisting with the clinical study operation, data collection process, auditing, and quality control and assurance, and Statcom Co Ltd, Tokyo, Japan, for performing statistical analysis in the validation stage. The development of the camera was led by members of the hardware team of Aillis; the development of the AI model was led by members of the AI team; and the clinical studies were led by the medical team. The authors also acknowledge all the members of Aillis, including, but not limited to, the aforementioned teams, and their dedicated efforts that have made this study possible. The member list is presented in Table S4 in Multimedia Appendix 1. This study was funded by the New Energy and Industrial Technology Development Organization, Japan (grant 30STS713) and Aillis Inc. The New Energy and Industrial Technology Development Organization had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The role of Aillis Inc in the pilot, training, and validation stages of the study was to provide the research equipment and testing costs. Aillis Inc was not involved in the clinical study operation, data collection process, auditing, and quality control and assurance of the study. Employees of Aillis Inc designed and conducted the study; managed, analyzed, and interpreted the data; prepared, reviewed, and approved the study; and were involved in the decision to submit the study.

Conflicts of Interest

SO is the chief executive officer and HK is the chief strategy officer of Aillis Inc, and they hold stock in the company. MF, MS, WT, M Ikeda, and HK are employees of Aillis Inc. YT and M Iwagami received consultant fees from the company to supervise the study and to draft the manuscript.

Multimedia Appendix 1
Supplementary Figures 1-7 and Tables 1-3.
[DOCX File , 6952 KB - jmir_v24i12e38751_app1.docx]

Multimedia Appendix 2
Concept video of pharyngeal examination and artificial intelligence (AI) diagnosis using our pharyngeal AI camera system “nodoca”.
[MP4 File (MP4 Video), 73341 KB - jmir_v24i12e38751_app2.mp4]

References

1. Troeger C, Blacker B, Khalil IA, Rao PC, Cao J, Zimsen SR, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect Dis 2018 Nov;18(11):1191-1210 [FREE Full text] [doi: 10.1016/S1473-3099(18)30310-4] [Medline: 30243584]
2. Iuliano AD, Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, Global Seasonal Influenza-associated Mortality Collaborator Network. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet 2018 Mar 31;391(10127):1285-1300 [FREE Full text] [doi: 10.1016/S0140-6736(17)33293-2] [Medline: 29248255]
3. Harper S, Bradley J, Englund J, File TM, Gravenstein S, Hayden FG, Expert Panel of the Infectious Diseases Society of America. Seasonal influenza in adults and children--diagnosis, treatment, chemoprophylaxis, and institutional outbreak management: clinical practice guidelines of the infectious diseases society of America. Clin Infect Dis 2009 Apr 15;48(8):1003-1032 [FREE Full text] [doi: 10.1086/598513] [Medline: 19281331]
4. Uyeki T, Bernstein H, Bradley J, Englund JA, File TM, Fry AM, et al. Clinical practice guidelines by the infectious diseases society of america: 2018 update on diagnosis, treatment, chemoprophylaxis, and institutional outbreak management of seasonal influenzaa. Clin Infect Dis 2019 Mar 05;68(6):e1-47 [FREE Full text] [doi: 10.1093/cid/ciy866] [Medline: 30566567]

5. Dunn JJ, Ginocchio CC. Point-counterpoint: can newly developed, rapid immunochromatographic antigen detection tests be reliably used for the laboratory diagnosis of influenza virus infections? J Clin Microbiol 2015 Jun;53(6):1790-1796. [doi: 10.1128/jcm.02739-14]
6. Chartrand C, Leeflang MM, Minion J, Brewer T, Pai M. Accuracy of rapid influenza diagnostic tests: a meta-analysis. Ann Intern Med 2012 Apr 03;156(7):500-511 [FREE Full text] [doi: 10.7326/0003-4819-156-7-201204030-00403] [Medline: 22371850]
7. Boivin G, Hardy I, Tellier G, Maziade J. Predicting influenza infections during epidemics with use of a clinical case definition. Clin Infect Dis 2000 Nov 15;31(5):1166-1169. [doi: 10.1086/317425] [Medline: 11073747]
8. Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. Arch Intern Med 2000 Nov 27;160(21):3243-3247. [doi: 10.1001/archinte.160.21.3243] [Medline: 11088084]
9. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 1970 Feb;12(1):55-67. [doi: 10.1080/00401706.1970.10488634]
10. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018 Presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; Jun 18-23, 2018; Salt Lake City, UT, USA URL: https://ieeexplore.ieee.org/document/8578843/ [doi: 10.1109/cvpr.2018.00745]
11. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015 Apr 11;115(3):211-252. [doi: 10.1007/s11263-015-0816-y]
12. Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3D shape recognition. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). 2015 Presented at: 2015 IEEE International Conference on Computer Vision (ICCV); Dec 07-13, 2015; Santiago, Chile. [doi: 10.1109/iccv.2015.114]
13. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; Jun 18-23, 2018; Salt Lake City, UT, USA URL: https://ieeexplore.ieee.org/document/8578572/ [doi: 10.1109/cvpr.2018.00474]
14. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY. Multimodal deep learning. In: Proceedings of the 28th International Conference on International Conference on Machine Learning. 2011 Presented at: 28th International Conference on International Conference on Machine Learning; Jun 28-Jul 2, 2011; Bellevue Washington USA.
15. Guo W, Wang J, Wang S. Deep multimodal representation learning: a survey. IEEE Access 2019;7:63373-63394. [doi: 10.1109/access.2019.2916887]
16. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4 - 9, 2017; Long Beach California USA.
17. Prokhorenkova L, Gusev G, Vorobev A, Dorogush A, Gulin A. CatBoost: unbiased boosting with categorical features. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018 Presented at: NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems; Dec 3 - 8, 2018; Montréal Canada. [doi: 10.5555/3327757.3327770]
18. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies. Ann Intern Med 2014 Jan 21;160(2):122-131. [doi: 10.7326/m13-1522]
19. Dagdanpurev S, Abe S, Sun G, Nishimura H, Choimaa L, Hakozaki Y, et al. A novel machine-learning-based infection screening system via 2013-2017 seasonal influenza patients' vital signs as training datasets. J Infect 2019 May;78(5):409-421. [doi: 10.1016/j.jinf.2019.02.008] [Medline: 30797793]
20. Ye Y, Tsui F, Wagner M, Espino JU, Li Q. Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. J Am Med Inform Assoc 2014 Sep 01;21(5):815-823 [FREE Full text] [doi: 10.1136/amiajnl-2013-001934] [Medline: 24406261]
21. López Pineda A, Ye Y, Visweswaran S, Cooper GF, Wagner MM, Tsui F. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. J Biomed Inform 2015 Dec;58:60-69 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.019] [Medline: 26385375]
22. Choo H, Kim M, Choi J, Shin J, Shin S. Influenza screening via deep learning using a combination of epidemiological and patient-generated health data: development and validation study. J Med Internet Res 2020 Oct 29;22(10):e21369 [FREE Full text] [doi: 10.2196/21369] [Medline: 33118941]
23. Sjoding MW, Taylor D, Motyka J, Lee E, Co I, Claar D, et al. Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. Lancet Digit Health 2021 Jun;3(6):e340-e348 [FREE Full text] [doi: 10.1016/S2589-7500(21)00056-X] [Medline: 33893070]
24. Hiremath A, Shiradkar R, Fu P, Mahran A, Rastinehad AR, Tewari A, et al. An integrated nomogram combining deep learning, Prostate Imaging–Reporting and Data System (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: a retrospective multicentre study. Lancet Digit Health 2021 Jul;3(7):e445-e454. [doi: 10.1016/s2589-7500(21)00082-0]

25. Rim TH, Lee CJ, Tham Y, Cheung N, Yu M, Lee G, et al. Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. Lancet Digit Health 2021 May;3(5):e306-e316. [doi: 10.1016/s2589-7500(21)00043-1]
26. Tomoda Y, Toya M, Kagawa S, Tanaka K. Influenza follicles. Intern Med 2019 Aug 01;58(15):2269 [FREE Full text] [doi: 10.2169/internalmedicine.2573-18] [Medline: 30996184]
27. Miyamoto A, Watanabe S. Influenza follicles and their buds as early diagnostic markers of influenza: typical images. Postgrad Med J 2016 Sep 27;92(1091):560-561 [FREE Full text] [doi: 10.1136/postgradmedj-2016-134271] [Medline: 27466411]
28. Kenzaka T. Influenza follicles in the posterior pharyngeal wall. Postgrad Med J 2015 Aug 04;91(1078):472. [doi: 10.1136/postgradmedj-2015-133344] [Medline: 26243881]
29. Miyamoto A, Watanabe S. Posterior pharyngeal wall follicles as a diagnostic marker of influenza during physical examination: considering their meaning and value. J Nihon Uni Med Assoc 2013;72(1):11-18 [FREE Full text] [doi: 10.4264/numa.72.11]

Abbreviations

AI: artificial intelligence
AUROC: area under the ROC curve
MM-CNN: multimodal convolutional neural network
MV-CNN: multiview convolutional neural network
NPV: negative predictive value
PPV: positive predictive value
ROC: receiver operating characteristic
RT-PCR: reverse transcription–polymerase chain reaction

Edited by T Leung; submitted 19.04.22; peer-reviewed by S You, G Lim; comments to author 02.08.22; revised version received 26.09.22; accepted 31.10.22; published 23.12.22.

Please cite as:
Okiyama S, Fukuda M, Sode M, Takahashi W, Ikeda M, Kato H, Tsugawa Y, Iwagami M
Examining the Use of an Artificial Intelligence Model to Diagnose Influenza: Development and Validation Study
J Med Internet Res 2022;24(12):e38751
URL: https://www.jmir.org/2022/12/e38751
doi:10.2196/38751
PMID:36374004

Original Paper

The Feasibility and Acceptability of an mHealth Conversational Agent Designed to Support HIV Self-testing in South Africa: Cross-sectional Study

Xolani Ntinga[1*], MSocSci; Franco Musiello[1*], MSocSci; Alfred Kipyegon Keter[1,2,3*], MSc; Ruanne Barnabas[4,5], DPhil; Alastair van Heerden[1,6*], PhD

[1]Centre for Community Based Research, Human Sciences Research Council, Pietermaritzburg, South Africa

[2]Department of Clinical Sciences, Institute of Tropical Medicine Antwerp, Antwerp, Belgium

[3]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

[4]Department of Medicine, Harvard Medical School, Boston, MA, United States

[5]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, United States

[6]South African Medical Research Council/WITS Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[*]these authors contributed equally

Corresponding Author:
Xolani Ntinga, MSocSci
Centre for Community Based Research
Human Sciences Research Council
Old Bus Depot Mbubu Road
Sweetwaters
Pietermaritzburg, 3201
South Africa
Phone: 27 333245000 ext 5018
Email: xntinga@hsrc.ac.za

Abstract

Background: HIV testing rates in sub-Saharan Africa remain below the targeted threshold, and primary care facilities struggle to provide adequate services. Innovative approaches that leverage digital technologies could improve HIV testing and access to treatment.

Objective: This study aimed to examine the feasibility and acceptability of *Nolwazi_bot*. It is an isiZulu-speaking conversational agent designed to support HIV self-testing (HIVST) in KwaZulu-Natal, South Africa.

Methods: Nolwazi_bot was designed with 4 different personalities that users could choose when selecting a counselor for their HIVST session. We recruited a convenience sample of 120 consenting adults and invited them to undertake an HIV self-test facilitated by the Nolwazi_bot. After testing, participants completed an interviewer-led posttest structured survey to assess their experience with the chatbot-supported HIVST.

Results: Participants (N=120) ranged in age from 18 to 47 years, with half of them being men (61/120, 50.8%). Of the 120 participants, 111 (92.5%) had tested with a human counselor more than once. Of the 120 participants, 45 (37.5%) chose to be counseled by the female Nolwazi_bot personality aged between 18 and 25 years. Approximately one-fifth (21/120, 17.5%) of the participants who underwent an HIV self-test guided by the chatbot tested positive. Most participants (95/120, 79.2%) indicated that their HIV testing experience with a chatbot was much better than that with a human counselor. Many participants (93/120, 77.5%) reported that they felt as if they were talking to a real person, stating that the response tone and word choice of Nolwazi_bot reminded them of how they speak in daily conversations.

Conclusions: The study provides insights into the potential of digital technology interventions to support HIVST in low-income and middle-income countries. Although we wait to see the full benefits of mobile health, technological interventions including conversational agents or chatbots provide us with an excellent opportunity to improve HIVST by addressing the barriers associated with clinic-based HIV testing.

(J Med Internet Res 2022;24(12):e39816) doi:10.2196/39816

KEYWORDS

HIV; HIV self-testing; HIVST; chatbot; conversational agents; mobile health; mHealth; mobile phone

Introduction

Background

Identifying patients with undiagnosed HIV and preventing new HIV infections remain critical public health issues. To reduce HIV incidence, global strategies emphasize early diagnosis, immediate treatment, and ongoing viral suppression for those living with HIV [1]. Despite the expansion of HIV testing services (HTSs) in sub-Saharan Africa, one-fifth of those aged between 15 and 64 years remain undiagnosed. Men, adolescents aged between 15 and 19 years, and adults aged ≥40 years continue to be infected with HIV owing to HIV testing gaps, which contribute to poor health outcomes and continued HIV transmission [2,3].

South Africa's health system is characterized by a quadruple burden of communicable, noncommunicable, maternal and child health, and injury-related disorders [4-7]. Primary health care (PHC) facilities in South Africa grapple with screening, initiating, and treating people with HIV, coupled with high incidence of tuberculosis, high maternal and child mortality levels, and growing burden of noncommunicable diseases [4,6]. As a result of the COVID-19 pandemic, facility-based HTS face new barriers in South Africa and elsewhere [8,9]. Currently, several approaches are promising in relieving some of the strain PHC facilities face, including HIV self-testing (HIVST), Chronic Medicines Dispensing and Distribution, community-based adherence clubs, and quick pharmacy pickups [7,10-12].

Despite these critically important and valuable initiatives, we suggest that, rather than incremental improvements within the existing framework of primary care, what is required is a reimagination of primary care that places digital services at the entry point of the health system instead of relying exclusively on human resources. Innovative approaches that leverage digital technologies could benefit populations that are not currently served by existing approaches. Although many of the benefits of mobile health (mHealth) have not yet materialized as hoped, the ubiquity of mobile phones, increasing availability of point-of-care health devices and screening tests, ability to collect and availability of large amounts of data about human behavior, and advances in machine intelligence make this proposed approach a possibility in the near future. HIV self-screening is an excellent test case for nonhuman intervention in the PHC system. Several issues with the current model of care deter people from getting tested [9]. Routine HTS in South Africa primarily uses a provider-based approach [13]. In this approach, individuals must visit an HIV testing location, such as a hospital or community center [9].

Although there have been gains in increasing access to HTS, barriers to the uptake of facility-based HTS include stigmatizing norms, discrimination from health care workers, distance to health facilities, and direct and indirect service use costs [9,14-17]. Innovative strategies to overcome these barriers will be critical to achieving the Joint United Nations Programme on HIV/AIDS *95–95–95* goals.

HIVST is a relatively new approach that provides an opportunity to reach, test, and diagnose or prevent infection among populations previously considered to be unreachable, even during the COVID-19 pandemic, owing to people's ability to self-test at home [15]. The World Health Organization recommends HIVST as an additional approach to provide HTS to help close this testing gap by increasing access and acceptability for HIV testing [13]. HIVST presents a private, convenient, and confidential approach to providing HTS that removes some of the barriers to routine HTS by allowing people to collect their samples and receive their results in the privacy of their homes, without interacting with a health care professional [9]. HIVST also reduces costs and saves time for the health delivery system and end user by triaging out the patients who are HIV-negative [18-20]. To support the use of HIVST in South Africa, guidelines for HIVST implementation were issued by the National Department of Health in February 2018 [9]. However, there are several concerns related to HIVST, such as lack of a formal pipeline for users to self-report their results or be linked to care following the self-test, potential of mental health risks associated with testing positive without counseling support, potential inability of testers to cope with their result, and that patients who undergo HIVST are less likely to access care [21,22]. Strong mobile phone penetration in low-income and middle-income countries has led to the development of various mHealth interventions to complement HIVST [23,24]. These include telephone hotlines, SMS text messaging interventions, internet-based platforms, and mobile apps. South Africa's mobile phone penetration and access to the internet is strong, with 89 (40%) of the households nationally having a mobile phone. In KwaZulu-Natal, 87.5% of households own mobile phones. Moreover, the national proportion of households with internet access was 74.1% in 2020. In the same year, 72.3% of households in KwaZulu-Natal had access to the internet [25].

Exploration of conversational agents in a health care setting suggests that users accept [21,22] and can form a working alliance with [21] embodied conversational agents. Examples of such agents include Florence, which was developed in the United Kingdom by the National Health Services as a digital solution for patient self-management and adherence through user-friendly, intelligent messaging that improves health outcomes—freeing up time and resources for clinicians and the health care system [26]. In addition, there is Molly, an empathy-based conversational platform developed by Sensely Corporation for linking people to care and managing chronic conditions, which is currently available in Japan, the Philippines, the United Kingdom, and the United States [27]. In addition, KOKO (developed by KoKo Incorporated) is a platform used to manage mental health, which is currently available in the United States [28]. In South Africa, mHealth apps developed

to support HIVST and reporting include the AspectTM HIVST app designed by SystemOne LLC [22] and the Ithaka mobile app developed by Aviro Health [24]. These conversational agents use a natural language understanding (NLU) engine to understand and respond to human interaction. NLU makes it possible to identify underlying user intents and enables the extraction of context, meanings, and domain-specific entities in users' utterances. NLU typically identifies three aspects in a sentence: (1) intent, which is done by mapping users' utterances to a specific class that allows digital web-based assistants to decide a response or action; (2) entities that illustrate important information such as date, times, and locations; and (3) contexts, which correspond to the context of the object the user is referring to [29]. A chatbot's accurate response to users' input requires combining these intents, entities, and contexts. Although there is increasing interest in the use of NLU-driven conversational agents in the health care context, the extent to which people find them acceptable for different uses needs further evaluation.

Objective

In this study, we examined the feasibility and acceptability of *Nolwazi_bot*, an isiZulu-speaking conversational agent designed to support HIVST in South Africa. The work on Nolwazi_bot began with the pilot study in 2017 [30], followed by a grant application that was successful in 2020. Upon commencement of data collection, there was already a trend in the development of digital interventions to support HIV, which is continuing to grow [31-35] owing to their agility and scalability because of low implementation, long-term recurring costs, and opportunity to reduce stigma and confidentiality concerns even among hard-to-reach populations [36]. Some of the recent studies on digital HIV interventions in South Africa [22,24] have been conducted in inner-city Johannesburg, which is South Africa's largest city. This study was conducted using a community-based approach in the rural Vulindlela subdistrict in KwaZulu-Natal province. In South Africa, KwaZulu-Natal is the province with the highest HIV prevalence, with evidence that uMgungundlovu is one of the districts with the highest prevalence (30%) in the country [37]. Given the different population of this chatbot feasibility study compared with the others conducted in South Africa, this study is the first to test a chatbot in a rural setting that has high HIV prevalence in South Africa. This study is also aiming to add to the literature that supports the idea that digital innovations are highly acceptable across diverse settings.

Methods

Study Design

This cross-sectional pilot study was conducted from December 2020 to April 2021. A convenience sample of 120 consenting adults were recruited from the Vulindlela subdistrict (uMgungundlovu district) in KwaZulu-Natal, South Africa. Recruitment was undertaken by a trained community outreach team who spoke to the public about this study. Those interested were screened against inclusion and exclusion criteria and then brought to the Human Sciences Research Council's Sweetwaters office to provide consent and complete the study. Participants were included if they had previously tested for HIV with a human counselor at any time in their life, were aged ≥18 years, resided in Vulindlela or a neighboring community, could use a smartphone to chat with the Nolwazi_bot (chatbot), and were able to provide written informed consent. Participants were excluded if they did not meet the inclusion criteria, had any condition that may have interfered with the testing process (such as intoxication or poor vision), or reported being HIV-positive. App feasibility is often assessed in a variety of ways. This study assessed feasibility by considering the following three criteria: (1) participants' acceptance of using the app (the chatbot in this study), (2) the participant's ability to complete the task on the app, and (3) the ability of the app to perform the required task [38]. For this study, these criteria were operationalized as follows: (1) the participants' willingness to undergo an HIVST guided by a chatbot, (2) the participants' ability to interact with the chatbot and follow the instructions of testing, and (3) the chatbot's ability to guide the participants to conduct HIVST and interpret their results. We also explored the socioeconomic status using the assets (electric stove, tap water, and car) available at participants' homes. We sorted individuals by the asset index and established cutoff values for percentiles of the population. Then, we assign households to a group based on their value on the index. For expository convenience, we refer to the bottom 40% as *poor*, the next 40% as *middle*, and the top 20% as *rich* [39].

Ethics Approval

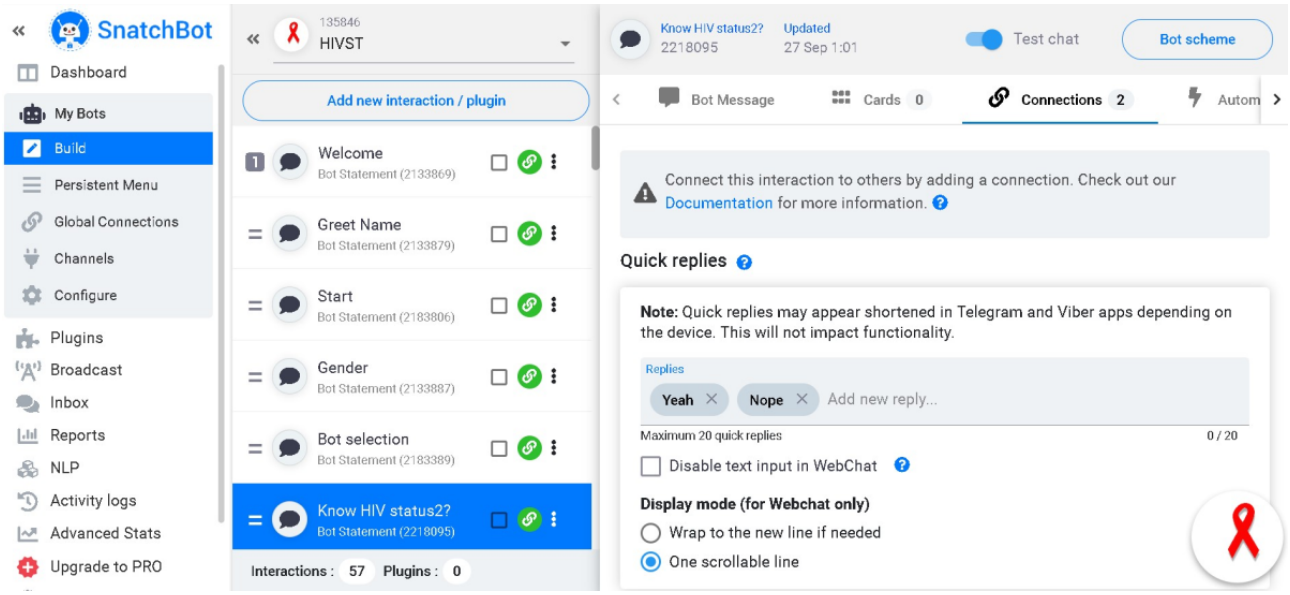
Ethics approval was obtained from the Human Sciences Research Council Research Ethics Committee (reference number 13/22/11/17). Trained study staff obtained written informed consent from all study participants using an information sheet and informed consent document approved by the Human Sciences Research Council Research Ethics Committee. The informed consent form was available in English and isiZulu. Participants were given reimbursement of R150 (equivalent to US $8.68) for their participation in the study.

Agent Development

The authors adapted a previous chatbot they had created using dialogue flow. Several commercially available websites that offered Natural Language Processing as a service were reviewed for suitability. Criteria used in the evaluation of the offerings were (1) integration with chat clients popular in South Africa (such as WhatsApp, WeChat, and Telegram), (2) graphical user interface for chatbot design and training, (3) ability to export model for publication, (4) protection of personal information regulatory compliance, and (5) offering webhook integration to interface with other services and allow the addition of anticipated future functionality. SnatchBot was selected, and Nolwazi_bot was designed and built on SnatchBot according to the Center for Disease Control's guidelines for the provision of HIV counseling and testing in a nonclinical setting [40]. The content of the chatbot was reviewed by a nurse working at the Human Sciences Research Council in English and isiZulu to assess its compliance with the self-testing guidelines described in the South African National HIV Testing Services Policy [41]. The content was also reviewed by bilingual translators at the Human Sciences Research Council to ensure that the language was culturally appropriate to use and easy to understand. The development of Nolwazi_bot followed the principles of the

SnatchBot platform (drag and drop and code-free design). SnatchBot provides an in-built editor that can be used to develop a simple or complex conversation with action buttons and translations. In addition, SnatchBot allows designers to create many interactions in relation with activities of the chatbot. During development, we created interactions that describe the predefined response patterns from the chatbot (including messages, videos, graphs, etc) after a user has said something. For example, in building an interaction between the chatbot and the user to introduce themselves, we created an interaction called *welcome*; if a person greets the chatbot it will reply with predefined response, that is, *welcome*. In a similar way, many other interactions were created, including name, bot selection (selecting a counselor), goodbye, known HIV status, HIV test results, and linkage to care, among others. Using SnatchBot, we created a design consisting of interactions and subjects. Figure 1 shows some part of the building scheme of Nolwazi_bot.

Figure 1. The building scheme of Nolwazi_bot.



Nolwazi_bot was designed to have 4 personalities that users could choose to be their counselor during the session. Of the 4 personalities, 2 were aged between 35 and 50 years (a middle-aged man and a middle-aged woman), and the other 2 personalities were aged between 18 and 25 years (a man and a woman in their youth). During the session, the older personalities spoke more formally in isiZulu, whereas the younger ones used a mixture of English and isiZulu.

Testing Procedure

Overview

The trained researcher obtained voluntary written informed consent from the participants in a private room. Each participant was temporarily provided with a Samsung J4 mobile phone running Android 8.1.0, with the Telegram messaging app preinstalled on the phone and an accompanying HIVST kit. The sealed test kit contained an English brochure with instructions for use as part of the standard packaging; however, the participant was requested to perform HIVST by following the isiZulu instructions on Nolwazi_bot and only use the instructions on the HIVST kit if requested by the chatbot. Once the participant was alone in the room, they opened the Telegram messaging app, searched for Nolwazi_bot, and opened the chatbot. Then, it greeted them and told them that they could choose to speak to a human at any point during the conversation by typing *help*, *I need help*, or *please help*. Then, the chatbot introduced 4 people, one of whom they could choose as their counselor for the session. Overall, 4 images were presented by the chatbot, which included 2 young individuals (aged 18-25 years; 1 man and 1 woman) and 2 older individuals (aged 35-50 years; 1 man and 1 woman). Stock photos were used to represent the personalities. For the young personalities, the language used by the bot was more colloquial and a mixture of isiZulu and English to represent how the young demographic group speaks in the study community, and it is typical of what would be used by young HIV counselors. For the older personalities, the chatbot converses in professional isiZulu, which did not include any English words.

The chatbot guided them through the conversations and emphasized that they could choose not to test if they did not feel comfortable or were not ready to test. Once they were ready to test, the chatbot provided them with a link to a video about using the kit and interpreting the result. The BioSure HIVST kit (BioSure Ltd) was used for the study, as it is already available and used in South Africa. Obtaining the sample takes 2 to 3 minutes, followed by a 15-minute waiting period for results to be produced. Then, participants would interpret their results alone and provide the app with their test results. A status-neutral approach was used for testing, with participants receiving a negative test result being asked if they wanted to learn more about pre-exposure prophylaxis (PrEP) options in the area. Participants who are HIV-positive were asked if they needed assistance with disclosing the result to family, friends, or their partner or if they required assistance in linking to care. Participants were informed that they still needed to visit the clinic to get tested again and confirm their positive result, according to South African HIVST guidelines. Participants were invited by Nolwazi_bot to take a picture of their test kit that

confirmed their HIV-positive results and present it to the clinic when they link to care. Figure 2 shows an example of the interaction between the tester and Nolwazi_bot in the Telegram messaging app. Following the test, the researcher asked the participants to participate in a posttest survey.

Figure 2. Interaction between the tester and Nolwazi_bot in the Telegram messaging app.

Posttest Survey

A face-to-face, interviewer-led, posttest structured survey (Multimedia Appendix 1) was conducted to obtain user feedback on their experience with Nolwazi_bot. The posttest survey was paper-based, and the pragmatic reason for having a paper-based survey was that the researchers would have had to create a separate survey bot, as the current bot was created for HIV testing. The investigators had planned to use the system usability scale by Brooke [42], which is a standardized scale for system or product usability assessment. However, after reviewing the scale, it was decided that considerable adaptation of the items would be required for this study. Therefore, we decided to use an adapted, unvalidated scale to assess usability. Unfortunately, resources were not available to perform cognitive interviews or other validation procedures before its use.

The posttest survey questions in this study were designed to support the understanding of the user experience of HIVST guided by a chatbot compared with a human counselor. Several questions were dichotomous *yes* and *no* questions, one was a Likert scale question, one was a scale question, and some were asking participants to choose between 2 different options. Participants were also asked open-ended questions regarding their perception of how human the conversation felt, why they preferred having a conversation with a particular sex, and the advantages and disadvantages of the chatbot. Open-ended questions were included in the study owing to the exploratory nature of the study. All participants (120/120, 100%) were invited to participate in the poststudy survey. Participants were eligible to participate if they provided consent and had completed the HIVST using the chatbot. All participants (120/120, 100%) participated in the poststudy survey. After the test and posttest survey were completed, the participants returned the phone to the researcher, and the used HIVST kits were disposed according to guidelines.

Data Analysis

All data extracted from the survey questionnaire (paper-based) and downloaded from the counseling chatbot were entered into the SPSS (version 25.0; IBM Corp) software. Exploratory descriptive statistics including frequencies and proportions were generated for demographic information, questions about HIV testing, chosen counselor, and HIV status, and PrEP information was described using frequencies and percentages. The open-ended questions were coded into ATLAS.ti and analyzed thematically. The qualitative data provided categories that supported the quantitative responses and allowed for better understanding of the participants' quantitative responses. For example, there was a question that asked participants "did it feel like a real person was replying?" We produced counts of the number of people who reported that the chatbot was similar to a real person; then, a follow-up qualitative question asked why it felt similar to a real person or why it did not feel similar to a real person. These qualitative responses were then used to provide themes to support the *realness* of the chatbot. A chi-square test was used to assess associations between categorical variables.

Results

Overview

Between December 2, 2020, and April 9, 2021, we screened 126 participants for the study, of whom 4 (3.2%) were not eligible owing to reported known HIV-positive status and 2 (1.6%) were not able to use a smartphone. A final sample of 95.2% (120/126) of participants aged 18 to 47 years, with median age of 24 (IQR 21.8-28) years, performed HIV self-test using the Nolwazi_bot. The sample was approximately equally divided across sexes. Overall, two-thirds of the participants (81/120, 67.5%) had secondary or high school education, and 92.5% (111/120) had tested more than once with a human counselor before testing with the chatbot. Up to 37.5% (45/120) of the participants chose a woman aged between 18 and 25 years to have a counseling session with. Participants conducted HIVST guided by the chatbot, and 17.5% (21/120) of them tested positive. These participants were provided with a referral to visit their preferred nearest clinic for a confirmatory test and linked to care. Thereafter, the community outreach team at the Human Sciences Research Council conducted follow-ups to check their linkage to care. Participants who tested negative were offered the option to learn more about PrEP, and 82.8% (82/99) of them wanted to know more about PrEP. Many participants (49/120, 40.8%) were from middle socioeconomic background, followed by poor socioeconomic background (48/120, 40%). There were no differences in responses by participants across variables when comparing HIV status. Further demographic data are presented in Table 1.

Table 1. Demographics characteristics of participants (N=120).

Characteristics	Value, n (%)
Sex	
Male	61 (50.8)
Female	59 (49.2)
Education	
Primary school	2 (1.7)
Secondary or high school	81 (67.5)
Tertiary institution	33 (27.5)
None	4 (3.3)
Times tested for HIV	
1	9 (7.5)
2-5	47 (39.2)
5-10	28 (23.3)
>10	36 (30)
Chosen counselor (age [years]; sex)	
18-25; male	26 (21.7)
18-25; female	45 (37.5)
35-50; male	24 (20)
35-50; female	25 (20.8)
HIV test outcome	
Negative	99 (82.5)
Positive	21 (17.5)
SES[a]	
Poor (lower 40%)	48 (40)
Middle (middle 40%)	49 (40.8)
Rich (upper 20%)	23 (19.2)
Response to question about whether they would like to know about PrEP[b] (n=99)	
Yes	82 (83)
No	17 (17)

[a]SES: socioeconomic status.

[b]PrEP: pre-exposure prophylaxis.

Chatbot Experience

After completing HIVST guided by the chatbot, participants were asked to assess the experience of undergoing HIV testing with the assistance of a chatbot compared with that of a human counselor. Of the 120 participants, most participants (n=95, 79.2%) indicated that their HIV testing experience with a chatbot was much better than that with a human counselor, 14 (11.7%) felt that the experience was approximately the same, and 7 (5.8%) felt that the experience was slightly better. Overall, 1.7% (2/120) of the participants felt that the experience was much worse than that with a human counselor.

Realness of Chatbot Conversations

Participants were asked whether they felt the counseling support they received during their testing to be similar to that obtained while talking to a real person. Of the 120 participants, 93 (77.5%) participants reported that they felt as if they were talking to a real person because the responses were in a tone that they would normally experience when talking to a person and the choice of words was similar to what they use in daily conversations. Other participants felt that the answers were correct and followed the order of pretest and posttest counseling that they usually participate in at clinics and hospitals. Of the 120 participants, 15 (12.5%) participants said that the counseling session did not feel as if they were chatting with a real person, citing that the chatbot replies were quicker than how people would respond, they did not have an opportunity to ask other questions and had to stick to the conversation, and they had to read all the chatbot responses with no option to use voice to record their own responses. Overall, 10% (12/120) of the participants did not respond to the question.

Advantages of Chatbot-Supported HIVST Compared With Testing With a Human Counselor

Participants were asked to provide the advantages of chatbot-supported HIVST compared with testing with a human counselor, if they felt they were any. Participants provided many responses; the responses were grouped into 5 broad categories that captured the responses. Of the 120 participants, 28 (23.3%) said that the chatbot provided them with a safe space. Participants mentioned that they do not feel vulnerable and exposed, which makes it easy to communicate with confidence and honesty, without fearing the counselor. The chatbot allows testers to answer questions in comfort without having to think about the other person (as there is no other person), and it allows testers to carefully answer in their own time, without having to worry about wasting a (human) counselor's time. Participants also mentioned that the bot does not criticize or judge them based on their sexual activity, which makes them feel safe, and they have time to adjust to the counseling session without the nurse being in a hurry to see the next client.

Of the 120 participants, 27 (22.5%) participants reported that an advantage of the chatbot is that it offered HIV testing and counseling (HTC) that is confidential, because there would not be any unintended disclosure as only they will know their HIV test outcome, whereas at the clinic, it is possible for a nurse to talk to someone about a person's status. Of the 120 participants, 15 (12.5%) participants said that the advantage of the chatbot was its functionality; they mentioned that the chatbot asked the right questions and the counseling was conducted in an empathetic and polite manner and educated them about things they did not know, such as acute HIV infection, PrEP, and information about viral suppression for individuals who are HIV-positive. Of the 120 participants, 12 (10%) participants said that the chatbot was efficient. Participants mentioned that it is fast and saves a lot of time, given that it can work at a fast pace if desired, unlike a human counselor, who will decide the pace of the HTC. In addition, participants indicated that they do not have to spend the (usual) entire day at a clinic to know their results; instead, in 30 minutes, they can know their status and take the next steps. Of the 120 participants, 9 (7.5%) participants felt that the chatbot was easy to use, and they indicated the following:

> *It is very easy to use, less stressful very understandable, it is the best and very advanced product one could ever wish for.*

Participants also indicated that chatbot-assisted HIVST is a good approach to HTC owing to the prevalent high use of cell phones and, in particular, social media. Of note, 24.2% (29/120) of the participants did not have any advantages to provide when asked about their chatbot experience.

Disadvantages of Chatbot-Supported HIVST Compared With Testing With a Human Counselor

Participants were also asked to provide the disadvantages they noticed when using the chatbot in comparison with performing HTC with a human. Of the 120 participants, 11 (9.2%) participants said that the chatbot lacked empathy in comparison with a human counselor. Furthermore, participants said that if they test HIV-positive, they could kill themselves because they will not receive the same comfort as provided by a real counselor. Moreover, participants indicated that if they exhibit suicidal ideations, the chatbot will not be able to intervene, and they would, in this instance, need to talk to a human counselor as the chatbot would not be able to provide verbal comfort, show feelings, or provide physical comfort such as a hug. Another disadvantage mentioned by 5% (6/120) of the participants was that the conversation with the chatbot was unidirectional. These participants indicated that they felt they could not ask questions during the counseling session. It is worth mentioning that none of the participants (0/120, 0%) elected to speak to a human counselor despite being informed that they could do so at any point during their session. Another point made by participants regarding the unidirectional conversation was that the chatbot will not change its response, even if the response is deemed to be unsatisfactory. Of the 120 participants, 4 (3.3%) participants reported the disadvantage of it being easy to make a mistake when chatting with a chatbot, as some may not be able to follow the instructions correctly and, consequently, make a mistake with the interpretation of their HIV results. Interestingly, 82.5% (99/120) of the participants did not have any disadvantages to provide when asked about their chatbot experience.

Counselor Preference

Chatbot-supported HTC compared with human HTC was evaluated on a scale of 1 to 10, with *1* being terrible and *10* being brilliant. Of the 120 participants, 12 (10%) participants did not respond to this question. Among those who responded, the average score was 9.32 (SD 1), with minimum score of 6 and maximum score of 10. Preference for the counselor among the participants was assessed. The participants were asked whether they would prefer a male or female conversational agent and the reason for their choice (Table 2). Of the 120 participants, 45 (37.5%) chose a young female counselor. Stratified by sex, the results reveal that a low proportion of male participants chose a female counselor aged between 35 and 50 years and a low proportion of female participants chose a male counselor. A chi-square test of association shows some evidence of association between the participant's sex and the sex of the chosen conversational agent (P=.01). This finding revealed that both male and female participants were more likely to select a counselor who was young and of the same sex.

Table 2. Preference for counselor among the participants (N=120)[a,b].

Counselor chosen (age [years]; sex)	Participant sex, n (%)		Total, n (%)
	Male	Female	
18-25; male	16 (13.3)	10 (8.3)	26 (21.7)
18-25; female	18 (15)	27 (22.5)	45 (37.5)
35-45; male	18 (15)	6 (5)	24 (20)
35-50; female	9 (7.5)	16 (13.3)	25 (20.8)

[a]χ^2_3=11.1.

[b]P=.01.

Discussion

Principal Findings

Acceptance of an HIV self-test using the Nolwazi_bot was assessed in 120 participants. This entailed an assessment of participants acceptance of performing HIVST guided by a chatbot, participants' ability to interact with the chatbot and follow the instructions of testing, and the chatbots' ability to guide the participants to conduct HIVST and interpret their results.

This study is one of the first investigations of an mHealth chatbot in South Africa to self-report HIVST results as an outcome outside of a clinical setting. The findings from this study have established that participants showed high acceptability of the chatbot, while also identifying challenges that can be targeted for improvement. The results suggest that some strengths of the chatbot are that it removed time constraints (which is common with a human counselor) and it was empathetic, polite, and educational. Weaknesses of the study include that the almost instantaneous responses of the chatbot were a reminder that it was not human; however, this speed was also noted as an advantage by some participants as it saved time, and 77.5% (93/120) of the participants reported that they felt as if they were talking to a real person. Moreover, the fact that the chatbot was not human was mentioned as an advantage, as the process was viewed as nonjudgmental.

Strengths of the HIVST Chatbot

Regarding the strengths of the chatbot, an advantage that was mentioned by participants was that they did not feel pressured for time while interacting with the agent, as it allowed testers to carefully answer in their own time and comfort, given that they did not have to worry about wasting another person's time, especially that of a busy health care worker. In South Africa, a contributing barrier to HIV testing may be rushed HIV counseling services, owing to high patient loads and inadequate facilities [43]. The chatbot may offer an acceptable alternative, which may encourage individuals to conduct HIV testing, who may otherwise not have tested owing to rushed interactions with a health care worker.

Further strengths of the chatbot were illustrated by some participants (15/120, 12.5%) indicating that an advantage of the chatbot was its functionality. These participants highlighted that the chatbot asked the right questions and the counseling was conducted in an empathetic and polite manner and educated them about things they did not know, such as acute HIV infection, PrEP, and information about viral suppression for individuals who are HIV-positive. This finding is of particular importance when considering that 68.3% (82/120) of the participants who tested negative wanted to know more about PrEP.

Another strength that was mentioned was the fast pace of the chatbot agent, as several participants (12/120, 10%) noted that the process can take a mere 30 minutes in comparison with what could take a full day at a PHC clinic. Moreover, another strength of the chatbot that was highlighted by the participants was the perceived safety of the interaction, given that a chatbot would not make any judgments based on their sexual activity. Therefore, chatbots could offer a suitable alternative solution to PHC testing, given that waiting times and issues of privacy have been reported to be barriers to HIV testing in men, who are known to be less-frequent users of public health facilities [44].

Perhaps related to privacy and judgment regarding sexual activity, of the 4 options that participants had for a counselor (man aged 18-25 years, woman aged 18-25 years, man aged 35-50 years, and woman aged 35-50 years), most participants (45/120, 37.5%) selected the woman aged between 18 and 25 years. Interestingly, 40.8% (49/120) of the sample selected an

older counselor in comparison with the 59.2% (71/120) of participants who selected a young HIV counselor. Given that 97.5% (117/120) of the participants were aged <35 years, it was hypothesized that most participants would prefer a young counselor. Some participants reported that a reason for the selection of an older counselor was the perception that with increased age, there is increased wisdom, which in turn would benefit the recipient of counseling. Given the limited sample size, no concrete conclusions can be drawn beyond these simple observations. Of significance for this study is that most participants (93/120, 77.5%) reported that their HIV testing experience with a chatbot was much better than that with a human counselor. Considering the abovementioned human resource limitations in the South African health care system, using mHealth tools in HIVST could contribute to alleviating the current demands on the health care system. Although further studies and development are still required to understand the potential uses, legal implications, and impact of conversational agents in health care, the data suggest that with improvement, chatbots may be able to provide public health screening not only for chronic infectious diseases (such as HIV) but also for noncommunicable diseases (such as diabetes and hypertension) in low-resource settings.

Weaknesses of the HIVST Chatbot

If one looks at the weaknesses of the chatbot, a disadvantage that was illustrated in the results is that the chatbot replied very fast, and therefore, the responses were said to be not human-like. This was mentioned as a disadvantage by 12.5% (15/120) of the participants. The same finding was noted in a similar study using the less-advanced Nolwazi_bot mHealth counseling agent [30]. With further programming, this perceived disadvantage can be overcome. For example, the chat agent responses could be made to be more human-like by delaying how quickly a response is sent and simulating slow (human-like) typing speeds. However, it may be worthwhile to give participants the option at the beginning of the conversation to indicate whether they are in a rush, as several participants (12/120, 10%) indicated that an advantage of the HIVST process with the chatbot was that it was a quicker process than that ordinarily done with a human counselor, and a contributing factor to this speed would be the almost immediate response time from the chatbot. Furthermore, the perceived disadvantage of rapid responses was only noted by a minority of the participants, and it should be noted together with the finding that 77.5% (93/120) of the participants reported that they felt as if they were talking to a real person.

Despite the apparent advantages of the HIVST chatbot in comparison with a human counselor, it is necessary to be cognizant that artificial intelligence is limited in its ability to weigh competing personal values and to be conscientious [45]. Parviainen and Rantala [46] argue that the intelligence of chatbots cannot assess emergency health situations and may cause harm owing to the lack of knowledge of personal factors associated with specific patients. This is particularly pertinent when one considers the range of emotions that one may be experiencing upon receiving an HIV-positive result. Nevertheless, when one considers that an individual may not go to a clinic for an HIV test owing to fear related to stigma, long waiting cues, or other concerns, it would be more beneficial for someone to perform a home HIV test with an mHealth conversational chatbot than to not test and not be aware of their (potential) HIV-infected status.

Limitations and Strengths

The study has several limitations. First, the generalizability of the findings is limited because convenience sampling was used, and participants were recruited from 1 subdistrict of KwaZulu-Natal. It would be beneficial for similar studies conducted in the future to have a large sampling frame to improve the generalizability of the findings. Second, another limitation to generalizability is that only 1 HIVST kit was used; therefore, these results cannot be generalized across all HIVST kits. Future studies would benefit from using several HIVST kits. Third, most participants (103/120, 85.8%) were aged <30 years, which may have increased the acceptability and ease of navigation of the chatbot, as young individuals are likely to have more technological skills than older age groups. Fourth, the chatbot was only tested on a Samsung phone and Telegram. This may not reflect the usability of the chatbot on other phones and platforms such as WhatsApp, which has a higher download rate on Google Play in South Africa than Telegram. Fifth, recent studies have developed validated data collection methods to determine the usability of mHealth tools [47]. The last limitation of the study is that the authors used an adapted, unvalidated scale to assess usability. Unfortunately, resources were not available to perform cognitive interview or other validation procedures before its use. The results from the study-specific questions of this study may not be replicable in similar settings, as (to the best of our knowledge) there are no validated data collection tools to assess the acceptability and feasibility of mHealth for HIVST.

The study had some strengths. First, 92.5% (111/120) of the participants have completed ≥2 HIV tests with a human counselor, which can be argued to contribute to the face validity of the abovementioned results, as the participants can be said to be well placed to compare the chatbot with a human counselor and therefore determine the strengths and weaknesses of the chatbot in relation to a human counselor. Second, the sample included a good representation of participants with various levels of education, including tertiary level, secondary or high school level, primary school level, and even no education. Third, the sample had a good representation of both men and women with the distribution being approximately balanced.

Conclusions

Although we wait to see the full benefits of mHealth, technological interventions including conversational agents or chatbots provide us with a good opportunity to improve HIVST, by addressing some of the barriers faced by both facilities and patients. The study provides insights into the potential of digital technological interventions to support health to improve HIVST, by addressing the barriers associated with clinic-based HIV testing.

Acknowledgments

The authors thank the people who participated in this study for their motivation and dedication.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

XN, AVH, and RVB were involved in conceptualizing or designing the study. XN was involved in data collection. XN, AVH, RVB, and AKK were involved in data analysis and interpretation. XN, AVH, and FM were responsible for drafting the manuscript. XN, AVH, and FM provided critical revisions to the manuscript. All the authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Posttest structured survey.
[PDF File (Adobe PDF File), 632 KB - jmir_v24i12e39816_app1.pdf]

References

1. Global AIDS Strategy 2021-2026 - End Inequalities. End AIDS. UNAIDS. URL: https://www.unaids.org/en/Global-AIDS-Strategy-2021-2026 [accessed 2022-03-14]
2. Indravudh PP, Fielding K, Kumwenda MK, Nzawa R, Chilongosi R, Desmond N, et al. Community-led delivery of HIV self-testing to improve HIV testing, ART initiation and broader social outcomes in rural Malawi: study protocol for a cluster-randomised trial. BMC Infect Dis 2019 Sep 18;19(1):814 [FREE Full text] [doi: 10.1186/s12879-019-4430-4] [Medline: 31533646]
3. Hlongwa M, Mashamba-Thompson T, Makhunga S, Hlongwana K. Barriers to HIV testing uptake among men in sub-Saharan Africa: a scoping review. Afr J AIDS Res 2020 Mar 15;19(1):13-23. [doi: 10.2989/16085906.2020.1725071] [Medline: 32174231]
4. WHO country cooperation strategy at a glance: South Africa. World Health Organization. 2018 May 1. URL: https://www.who.int/publications/i/item/WHO-CCU-18.02-South-Africa [accessed 2022-03-14]
5. Achoki T, Sartorius B, Watkins D, Glenn SD, Kengne AP, Oni T, et al. Health trends, inequalities and opportunities in South Africa's provinces, 1990-2019: findings from the Global Burden of Disease 2019 Study. J Epidemiol Community Health 2022 Jan 19;76(5):471-481 [FREE Full text] [doi: 10.1136/jech-2021-217480] [Medline: 35046100]
6. Samodien E, Abrahams Y, Muller C, Louw J, Chellan N. Non-communicable diseases – a catastrophe for South Africa. South African J Sci 2021 May 28;117(5/6):32-37. [doi: 10.17159/sajs.2021/8638]
7. Meyer JC, Schellack N, Stokes J, Lancaster R, Zeeman H, Defty D, et al. Ongoing initiatives to improve the quality and efficiency of medicine use within the public healthcare system in South Africa; a preliminary study. Front Pharmacol 2017 Nov 09;8:751 [FREE Full text] [doi: 10.3389/fphar.2017.00751] [Medline: 29163151]
8. Mhango M, Chitungo I, Dzinamarira T. COVID-19 lockdowns: impact on facility-based HIV testing and the case for the scaling up of home-based testing services in Sub-Saharan Africa. AIDS Behav 2020 Nov;24(11):3014-3016 [FREE Full text] [doi: 10.1007/s10461-020-02939-6] [Medline: 32488551]
9. Majam M, Conserve DF, Zishiri V, Haile ZT, Tembo A, Phiri J, et al. Implementation of different HIV self-testing models with implications for HIV testing services during the COVID-19 pandemic: study protocol for secondary data analysis of the STAR Initiative in South Africa. BMJ Open 2021 May 18;11(5):e048585 [FREE Full text] [doi: 10.1136/bmjopen-2020-048585] [Medline: 34006558]
10. Fatti G, Ngorima-Mabhena N, Tiam A, Tukei BB, Kasu T, Muzenda T, et al. Community-based differentiated service delivery models incorporating multi-month dispensing of antiretroviral treatment for newly stable people living with HIV receiving single annual clinical visits: a pooled analysis of two cluster-randomized trials in southern Africa. J Int AIDS Soc 2021 Oct;24 Suppl 6:e25819 [FREE Full text] [doi: 10.1002/jia2.25819] [Medline: 34713614]
11. Liu L, Christie S, Munsamy M, Roberts P, Pillay M, Shenoi SV, et al. Correction to: expansion of a national differentiated service delivery model to support people living with HIV and other chronic conditions in South Africa: a descriptive analysis. BMC Health Serv Res 2021 Jun 04;21(1):549 [FREE Full text] [doi: 10.1186/s12913-021-06561-7] [Medline: 34088315]
12. Long L, Kuchukhidze S, Pascoe S, Nichols B, Cele R, Govathson C, et al. Differentiated models of service delivery for antiretroviral treatment of HIV in sub-Saharan Africa: a rapid review protocol. Syst Rev 2019 Dec 06;8(1):314 [FREE Full text] [doi: 10.1186/s13643-019-1210-6] [Medline: 31810482]
13. WHO recommends HIV self-testing – evidence update and considerations for success. World Health Organization. 2019 Nov 27. URL: https://www.who.int/publications/i/item/WHO-CDS-HIV-19.36 [accessed 2022-03-15]

14. Bwalya C, Simwinga M, Hensen B, Gwanu L, Hang'andu A, Mulubwa C, HPTN 071 (PopART) study team. Social response to the delivery of HIV self-testing in households: experiences from four Zambian HPTN 071 (PopART) urban communities. AIDS Res Ther 2020 Jun 11;17(1):32 [FREE Full text] [doi: 10.1186/s12981-020-00287-y] [Medline: 32527261]
15. Chimoyi L, Chikovore J, Musenge E, Mabuto T, Chetty-Makkan CM, Munyai R, et al. Understanding factors influencing utilization of HIV prevention and treatment services among patients and providers in a heterogeneous setting: a qualitative study from South Africa. PLOS Glob Public Health 2022 Feb 3;2(2):e0000132 [FREE Full text] [doi: 10.1371/journal.pgph.0000132]
16. Mohlabane N, Tutshana B, Peltzer K, Mwisongo A. Barriers and facilitators associated with HIV testing uptake in South African health facilities offering HIV Counselling and Testing. Health SA Gesondheid 2016 Dec;21:86-95. [doi: 10.1016/j.hsag.2015.11.001]
17. Venter F, Majam M, Jankelowitz L, Adams S, Moorhouse M, Carmona S, et al. South African HIV self-testing policy and guidance considerations. South Afr J HIV Med 2017 Oct 26;18(1):775 [FREE Full text] [doi: 10.4102/sajhivmed.v18i1.775] [Medline: 29568643]
18. World Health Organization. Guidelines on HIV Self-testing and Partner Notification: Supplement to Consolidated Guidelines on HIV Testing Services. Geneva: World Health Organization; 2016.
19. Ingold H, Mwerinde O, Ross AL, Leach R, Corbett EL, Hatzold K, et al. The Self-Testing AfRica (STAR) Initiative: accelerating global access and scale-up of HIV self-testing. J Int AIDS Soc 2019 Mar 25;22 Suppl 1(S1):e25249 [FREE Full text] [doi: 10.1002/jia2.25249] [Medline: 30907517]
20. Harichund C, Moshabela M. Acceptability of HIV self-testing in Sub-Saharan Africa: scoping study. AIDS Behav 2018 Feb 11;22(2):560-568 [FREE Full text] [doi: 10.1007/s10461-017-1848-9] [Medline: 28699017]
21. Hamilton A, Thompson N, Choko AT, Hlongwa M, Jolly P, Korte JE, et al. HIV self-testing uptake and intervention strategies among men in Sub-Saharan Africa: a systematic review. Front Public Health 2021 Feb 19;9:594298 [FREE Full text] [doi: 10.3389/fpubh.2021.594298] [Medline: 33681120]
22. Gous N, Fischer AE, Rhagnath N, Phatsoane M, Majam M, Lalla-Edward ST. Evaluation of a mobile application to support HIV self-testing in Johannesburg, South Africa. South Afr J HIV Med 2020 Jun 30;21(1):1088 [FREE Full text] [doi: 10.4102/sajhivmed.v21i1.1088] [Medline: 32670629]
23. Nwaozuru U, Obiezu-Umeh C, Shato T, Uzoaru F, Mason S, Carter V, et al. Mobile health interventions for HIV/STI prevention among youth in low- and middle-income countries (LMICs): a systematic review of studies reporting implementation outcomes. Implement Sci Commun 2021 Nov 06;2(1):126. [doi: 10.1186/s43058-021-00230-w] [Medline: 34742357]
24. Fischer AE, Phatsoane M, Majam M, Shankland L, Abrahams M, Rhagnath N, et al. Uptake of the Ithaka mobile application in Johannesburg, South Africa, for human immunodeficiency virus self-testing result reporting. South Afr J HIV Med 2021 Feb 22;22(1):1197 [FREE Full text] [doi: 10.4102/sajhivmed.v22i1.1197] [Medline: 33824737]
25. State of the ICT sector in South Africa - 2022 report. Independent Communications Authority of South Africa. URL: https://www.icasa.org.za/legislation-and-regulations/state-of-the-ict-sector-in-south-africa-2022-report [accessed 2022-09-23]
26. Florence Shaping the future of behavioural change by design. Generated Health. URL: https://getflorence.net/ [accessed 2022-04-15]
27. Increasing Access. Lowering Costs. Improving Health. Sensely. URL: https://sensely.com/ [accessed 2022-03-17]
28. Koko homepage. Koko. URL: https://www.kokocares.org/ [accessed 2022-04-16]
29. Inupakutika D, Nadim M, Gunnam GR, Kaghyan S, Akopian D, Chalela P, et al. Integration of NLP and speech-to-text applications with chatbots. Electron Imaging 2021 Jun 18;33(3):35-1-35-6. [doi: 10.2352/issn.2470-1173.2021.3.mobmu-035]
30. van Heerden A, Ntinga X, Vilakazi K. The potential of conversational agents to provide a rapid HIV counseling and testing services. In: Proceedings of the 2017 International Conference on the Frontiers and Advances in Data Science (FADS). 2017 Presented at: 2017 International Conference on the Frontiers and Advances in Data Science (FADS); Oct 23-25, 2017; Xi'an, China. [doi: 10.1109/FADS.2017.8253198]
31. John SA, Starks TJ, Rendina HJ, Parsons JT, Grov C. High willingness to use novel HIV and bacterial sexually transmitted infection partner notification, testing, and treatment strategies among gay and bisexual men. Sex Transm Infect 2020 May 12;96(3):173-176 [FREE Full text] [doi: 10.1136/sextrans-2019-053974] [Medline: 31189548]
32. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond "implementation": digital health innovation and service design. NPJ Digit Med 2018 Sep 20;1(1):48 [FREE Full text] [doi: 10.1038/s41746-018-0059-8] [Medline: 31304327]
33. De Boni RB, Veloso VG, Fernandes NM, Lessa F, Corrêa RG, Lima RD, et al. An internet-based HIV self-testing program to increase HIV testing uptake among men who have sex with men in Brazil: descriptive cross-sectional analysis. J Med Internet Res 2019 Aug 01;21(8):e14145 [FREE Full text] [doi: 10.2196/14145] [Medline: 31373276]
34. Yun K, Chu Z, Zhang J, Geng W, Jiang Y, Dong W, et al. Mobile phone intervention based on an HIV risk prediction tool for HIV prevention among men who have sex with men in China: randomized controlled trial. JMIR Mhealth Uhealth 2021 Apr 13;9(4):e19511 [FREE Full text] [doi: 10.2196/19511] [Medline: 33847597]
35. Chan PS, Chidgey A, Lau J, Ip M, Lau JT, Wang Z. Effectiveness of a novel HIV self-testing service with online real-time counseling support (HIVST-online) in increasing HIV testing rate and repeated HIV testing among men who have sex with

men in Hong Kong: results of a pilot implementation project. Int J Environ Res Public Health 2021 Jan 15;18(2):729 [FREE Full text] [doi: 10.3390/ijerph18020729] [Medline: 33467770]
36. Qin Y, Han L, Babbitt A, Walker JS, Liu F, Thirumurthy H, et al. Experiences using and organizing HIV self-testing. AIDS 2018 Jan 28;32(3):371-381 [FREE Full text] [doi: 10.1097/QAD.0000000000001705] [Medline: 29194120]
37. Dwyer-Lindgren L, Cork MA, Sligar A, Steuben KM, Wilson KF, Provost NR, et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. Nature 2019 Jun 15;570(7760):189-193 [FREE Full text] [doi: 10.1038/s41586-019-1200-9] [Medline: 31092927]
38. Pant Pai N, Bhargava M, Joseph L, Sharma J, Pillay S, Balram B, et al. Will an unsupervised self-testing strategy be feasible to operationalize in Canada? Results from a pilot study in students of a large Canadian university. AIDS Res Treat 2014;2014:747619 [FREE Full text] [doi: 10.1155/2014/747619] [Medline: 24511392]
39. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data-or tears: an application to educational enrollments in states of India. Demography 2001 Feb;38(1):115. [doi: 10.2307/3088292]
40. Implementing HIV testing in nonclinical settings a guide for HIV testing providers. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/pdf/testing/CDC_HIV_Implementing_HIV_Testing_in_Nonclinical_Settings.pdf?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fhiv%2Feffective-interventions%2Flibrary%2Fredirects%2Fimplementation-materials%2Fredirect1i.html [accessed 2022-11-13]
41. National HIV testing services: policy 2016. Department of Health Republic of South Africa. 2016. URL: https://www.uj.ac.za/wp-content/uploads/2021/10/2016-hts-policy-doc-19-jan-3.pdf [accessed 2022-09-23]
42. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996.
43. Haffejee S, Groeneveld I, Fine D, Patel R, Bowman B. An assessment of counselling and support services for people living with HIV in Gauteng, South Africa: findings of a baseline study. Afr J AIDS Res 2010 Dec;9(4):367-372. [doi: 10.2989/16085906.2010.545642] [Medline: 25875885]
44. Sharma M, Ying R, Tarr G, Barnabas R. Systematic review and meta-analysis of community and facility-based HIV testing to address linkage to care gaps in sub-Saharan Africa. Nature 2015 Dec 03;528(7580):S77-S85 [FREE Full text] [doi: 10.1038/nature16044] [Medline: 26633769]
45. Powell J. Trust me, I'm a chatbot: how artificial intelligence in health care fails the turing test. J Med Internet Res 2019 Oct 28;21(10):e16222 [FREE Full text] [doi: 10.2196/16222] [Medline: 31661083]
46. Parviainen J, Rantala J. Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. Med Health Care Philos 2022 Mar 04;25(1):61-71 [FREE Full text] [doi: 10.1007/s11019-021-10049-w] [Medline: 34480711]
47. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]

Abbreviations

HIVST: HIV self-testing
HTC: HIV testing and counseling
HTS: HIV testing service
mHealth: mobile health
NLU: natural language understanding
PHC: primary health care
PrEP: pre-exposure prophylaxis

Edited by T Leung; submitted 24.05.22; peer-reviewed by L Chang, K Clouse, A Krishnan; comments to author 06.09.22; revised version received 28.09.22; accepted 28.10.22; published 12.12.22.

Please cite as:
Ntinga X, Musiello F, Keter AK, Barnabas R, van Heerden A
The Feasibility and Acceptability of an mHealth Conversational Agent Designed to Support HIV Self-testing in South Africa: Cross-sectional Study
J Med Internet Res 2022;24(12):e39816
URL: https://www.jmir.org/2022/12/e39816
doi:10.2196/39816
PMID:36508248

Original Paper

Perspectives of Rare Disease Social Media Group Participants on Engaging With Genetic Counselors: Mixed Methods Study

Megan Yabumoto[1], MSc; Emily Miller[2], BA; Anoushka Rao[2]; Holly K Tabor[2,3], PhD; Kelly E Ormond[1,2,4], MSc; Meghan C Halley[2], MPH, PhD

[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA, United States

[2]Stanford Center for Biomedical Ethics, Stanford University School of Medicine, Stanford, CA, United States

[3]Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

[4]Health Ethics and Policy Lab, Department of Health Sciences and Technology, Swiss Federal Institute of Technology (Eidgenössische Technische Hochschule Zurich), Zurich, Switzerland

Corresponding Author:
Meghan C Halley, MPH, PhD
Stanford Center for Biomedical Ethics
Stanford University School of Medicine
300 Pasteur Drive
Stanford, CA, 94305
United States
Phone: 1 414 745 9381
Email: mhalley@stanford.edu

Abstract

Background: Social media provides a potential avenue for genetic counselors to address gaps in access to reliable genetics information for rare disease communities. However, only limited research has examined patient and family attitudes toward engaging with genetic counselors through social media.

Objective: Our study assessed the attitudes of members of rare disease social media groups toward engaging with genetic counselors through social media, characteristics associated with greater interest, and the benefits and potential pitfalls of various approaches to such engagement.

Methods: We conducted a mixed methods survey of patients and family members recruited from a systematic sample of rare disease Facebook groups. Patient characteristics and their associations with interest in engagement with genetic counselors were evaluated using univariate and bivariate statistics. Responses to open-ended questions were analyzed using thematic content analysis.

Results: In total, 1053 individuals from 103 rare disease groups participated. The median overall interest in engaging with genetic counselors on social media was moderately high at 7.0 (IQR 4.0-9.0, range 0-10). No past experience with a genetic counselor was associated with greater interest in engaging with one through social media (μ=6.5 vs 6.0, *P*=.04). Participants expressed greatest interest (median 9.0, IQR 5.0-10.0) in engagement models allowing direct communication with genetic counselors, which was corroborated by the majority (n=399, 61.3%) of individuals who responded to open-ended questions explicitly stating their interest in 1-on-1 interactions. When asked what forms of support they would request from genetic counselors through social media, participants desired individualized support and information about how to access services. However, participants also expressed concerns regarding privacy and confidentiality.

Conclusions: Patients and family members in rare disease social media groups appear interested in engaging with genetic counselors through social media, particularly for individualized support. This form of engagement on social media is not meant to replace the current structure and content of genetic counseling (GC) services, but genetic counselors could more actively use social media as a communication tool to address gaps in knowledge and awareness about genetics services and gaps in accessible patient information. Although encouraging, concerns regarding privacy and feasibility require further consideration, pointing to the need for professional guidelines in this area.

(J Med Internet Res 2022;24(12):e42084) doi:10.2196/42084

KEYWORDS

social media; rare disease; genetic counseling; genetics; genomics; delivery of health care

Introduction

Over 10,000 different rare diseases collectively affect an estimated 300 million people worldwide [1]. Approximately 70% of these rare diseases are genetic in etiology, and 80% have symptoms that emerge in childhood [2]. Patients with rare diseases typically experience long delays in diagnosis due to providers' lack of familiarity with these conditions and limited access to diagnostic testing [3]. Although genomic technologies, such as exome and genome sequencing, continue to identify both new and existing rare genetic disorders, these advances have rapidly outpaced the availability of genetic testing (GT) services [4,5]. Additionally, even after diagnosis, patients report significant challenges in accessing reliable and patient-friendly information about their disease, including prognosis, natural history, and management [6]. As the field of genomics continues to expand, new strategies will be needed to increase access to genetics services for the large and heterogenous population of patients with rare diseases.

Genetic counselors are particularly well positioned to address the informational, social, and emotional needs of patients with rare genetic diseases and their family members [7,8]. Trained in both the clinical implications of genetics and patient- and family-centered communication, genetic counselors are in an ideal position to disseminate accessible information about rare genetic diseases to patients, families, and patient communities [9,10]. However, limited availability of genetic counselors is an ongoing challenge. Although there are an estimated 7000 genetic counselors currently practicing worldwide, over 60% of these counselors practice in North America [11]. Even within the United States, the demand for genetic counselors is placing immense pressure on the workforce [12,13]. Although the field is working to train new genetic counselors to meet the rapidly growing global need, responding to this demand will also require creative and efficient service delivery models [14].

One potential strategy for disseminating information to large numbers of genetics patients on a global scale is through social media [7]. Social media provides an accessible tool for individuals to connect with one another and share information and support, including around health and illness [15-17]. Individuals impacted by rare diseases and their family members are particularly active on these platforms for multiple reasons, including for social and emotional support from those experiencing similar conditions worldwide and to fill in gaps in information about their rare disease due to local providers' limited exposure to their condition or limited available research in general [10,18,19]. Further, a recent study suggested that patients and family members may be interested in using social media to receive general information about GC and genetics services. However, they also suggest concerns about maintaining privacy and confidentiality in the group environment [9]. Additional information is needed to understand attitudes toward engaging with genetic counselors through social media in the broader rare disease community, how to structure such interactions to balance patient preferences regarding privacy and access, and who might benefit from interactions with genetic counselors on social media platforms.

To address this gap, we conducted a survey of patients with rare genetic diseases and their family members using a systematic sampling structure to include a broad range of rare diseases. We intentionally focused recruitment on current social media users to better understand the benefits and barriers specific to genetic counselor interactions in this context. Here, we report our findings on participants' attitudes toward engaging with genetic counselors through social media, individual characteristics associated with greater interest in engagement with genetic counselors, and the perceived benefits and drawbacks of various approaches to engaging with genetic counselors in this context.

Methods

Study Design

We conducted an online survey from October to December 2021 of patients with rare diseases and their family members participating in social media support groups identified from a systematic sample of rare diseases.

Ethical Considerations

All study procedures were approved by the Stanford University School of Medicine Institutional Review Board (IRB protocol no. 61783).

Sampling and Participant Recruitment

Studies of patients with rare diseases often include only a small subset of the more common rare diseases (eg, cystic fibrosis, amyotrophic lateral sclerosis, and Huntington disease) [20]. To address this limitation of the current literature, we used a systematic approach to identifying and recruiting participants from a broad range of rare diseases.

Identifying Rare Diseases

To recruit patients with rare diseases and their family members, we selected a random sample of rare diseases from the Orphanet database, stratified by disease prevalence [21,22]. Nearly 85% of rare diseases listed in Orphanet are "ultra-rare" (defined as having a prevalence of <1 in 1,000,000), but an estimated 80% of the population burden of rare diseases is attributable to only 4% of rare diseases that are more "common-rare" diseases (defined as having a prevalence of 1-9 in 1,000,000 or greater). To ensure inclusion of both common-rare and ultra-rare diseases, we oversampled for common-rare diseases from Orphanet. Additional parameters were based on estimates that approximately 70% of Orphanet diseases are genetic in etiology and that 30% of ultra-rare diseases and 70% of common-rare diseases are expected to have a Facebook group [2,23]. Based on these estimates, we selected a stratified random sample of 1200 rare diseases with the expectation of identifying a Facebook group for at least 400 different rare diseases and enrolling participants from 100 of the identified groups. After selecting this sample of 1200 rare diseases, we screened each

disease on rare disease databases to only include rare diseases with a known or suspected genetic etiology based on information provided by organizations, such as the National Organization of Rare Disorders [24] and GeneReviews [25].

Identifying Social Media Groups

To identify social media groups for our list of rare genetic diseases, we used Facebook, the largest social media platform available and on which rare disease groups are known to be active [23]. Using a dedicated, study-specific Facebook account for the study's principal investigator (author MCH), we searched each identified disease in our sample using both the disease's primary name and up to 5 alternative names listed in Orphanet. Eligibility criteria for groups included (1) categorized as a group on Facebook and (2) explicitly focused on an eligible rare disease per the public group description. If more than 1 group was identified for a single disease, only the group with the largest number of members was included. For ultra-rare diseases with multiple subtypes, umbrella groups covering more than 1 subtype were reviewed and included if the specific subtype was named in the group description (eg, autosomal dominant optic atrophy [ADOA] as the umbrella group for the diseases ADOA Kjer-type and ADOA-plus type).

Participants and Procedures

To recruit participants, a member of the study team contacted up to 3 moderators or administrators of each identified Facebook group via private message. If moderators and administrators agreed to post the survey link to the group, they were provided with the IRB-approved recruitment language and survey link to share with their group members. We attempted to contact each Facebook group up to 3 times over a 6-week period, and all groups that agreed to post the survey had access to an active survey link for a minimum of 3 weeks. Participant eligibility criteria included (1) aged 18 years or older, (2) able to read and write in English, and (3) self-identified as either a patient with a rare disease or a family member of a patient with a rare disease. All included groups were associated with a disease with a known genetic component. However, a subset of individuals with a recognized genetic disease may be diagnosed clinically, without a molecular diagnosis. All patients and their family members within the identified groups were eligible for participation regardless of whether they had a confirmed molecular diagnosis.

Measures

The survey instrument was developed through an iterative process. Structured questions were drawn from previously published studies whenever possible and included additional measures not represented in the analysis later (see Multimedia Appendix 1). All new items were developed based on the existing literature [9,10] and were pretested with patients with rare diseases prior to dissemination. The measures in the analysis included (1) sociodemographic characteristics [26,27], (2) self-reported interest in engaging with genetic counselors through social media [9], (3) prior access to and experience with GC and GT [9,28,29], and (4) frequency of social media use and self-reported perceived social connectedness [30]. In addition, we presented 4 proposed models for how engagement with genetic counselors through social media could be structured and asked participants to rate their interest in each on a scale of 0-10, with 10 indicating maximum interest. These models were developed to explore participants' attitudes toward varying approaches to engagement with increasing levels of direct access to a genetic counselor. The model "minimal engagement" involved the genetic counselor sending information and resources through the moderator, with no direct interaction with other group members. The model "moderate engagement" involved bidirectional communication between the genetic counselor and the moderator only. "Enhanced moderate engagement" involved the genetic counselor joining the group directly but providing only information and resources. Finally, "maximum engagement" involved the genetic counselor joining the group and engaging in bidirectional communication with all group members. Four open-ended questions also were included to elicit participant perspectives on the benefits and drawbacks of these different models. The survey was distributed using Qualtrics software (Qualtrics).

Finally, data on rare disease social media group characteristics were extracted from publicly available Facebook information (eg, size of group, activity within group), and additional data on each rare disease represented in the final sample were extracted from Orphanet (eg, disease classification, inheritance pattern, age of onset). We integrated the social media and rare disease characteristics into individual participant-level data.

Data Analysis

We used R version 4.1.1 (R Core Team and the R Foundation for Statistical Computing) [31] for quantitative analyses and Microsoft Excel (Microsoft Corporation) for qualitative analysis. Descriptive analyses were performed for all variables using means and SDs to describe normally distributed variables and medians with IQRs for all nonnormally distributed variables. Participants' reported connectedness to their social media group was summarized as a mean social media connectedness score based on 3 questions representing how connected participants felt to their rare disease group on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Our primary comparative analysis was designed to test a series of hypotheses to investigate whether certain participant, social media group, and rare disease characteristics were associated with overall interest in engaging online (primary outcome variable). To do this, we first performed bivariate analyses (Welch 2-sample *t* tests and ANOVA) to examine the relationship between each hypothesized predictor and the primary outcome variable. We planned to conduct a multivariable analysis (linear regression) if more than 1 independent variable was associated with the primary outcome variable with $P<.10$. However, bivariate analyses resulted in only 1 potential predictor of increased interest in engagement, so we did not perform the planned multivariate analysis.

Responses to open-ended survey questions were analyzed using a thematic content analysis approach. Two team members reviewed the data and developed a draft codebook based on themes the team determined to be most prevalent in the data. We conducted multiple rounds of codebook revision to ensure high interrater reliability (>90% agreement) and then applied

codes to the full data set (Supplementary Table 1 in Multimedia Appendix 2) [32]. We calculated frequencies for each code by question and identified exemplary quotes for each code.

Results

Social Media Group Member Characteristics

A total of 1053 eligible individuals from 103 Facebook groups responded and completed at least 1 survey question following screening (Figure 1). Of note, our final sample included participants with common-rare diseases at rates proportional to population estimates (n=820, 77.8% in our sample vs 80% population estimate). Over half (n=660, 62.7%) of participants self-identified as an adult patient with a rare disease and 37.3% (n=393) as the family member of a patient with rare diseases. Participants had a median age of 43 years (IQR 35-52) and were predominantly non-Hispanic (n=982, 93.3%), White (n=957, 90.9%), and female (n=868, 82.4%). Approximately one-quarter (n=287, 27.3%) of participants lived outside the United States. Additional individual participant characteristics are provided in Table 1 and Supplementary Table 2 in Multimedia Appendix 2; additional participant, rare disease, and social media group characteristics are provided in Supplementary Table 1 in Multimedia Appendix 2.

Figure 1. Creation of a systematic random sample of rare diseases included in this study. Note: One group was not contacted because the term "rare disease" was inconsistent with how members of the community identified, as described on the Facebook group's public description.

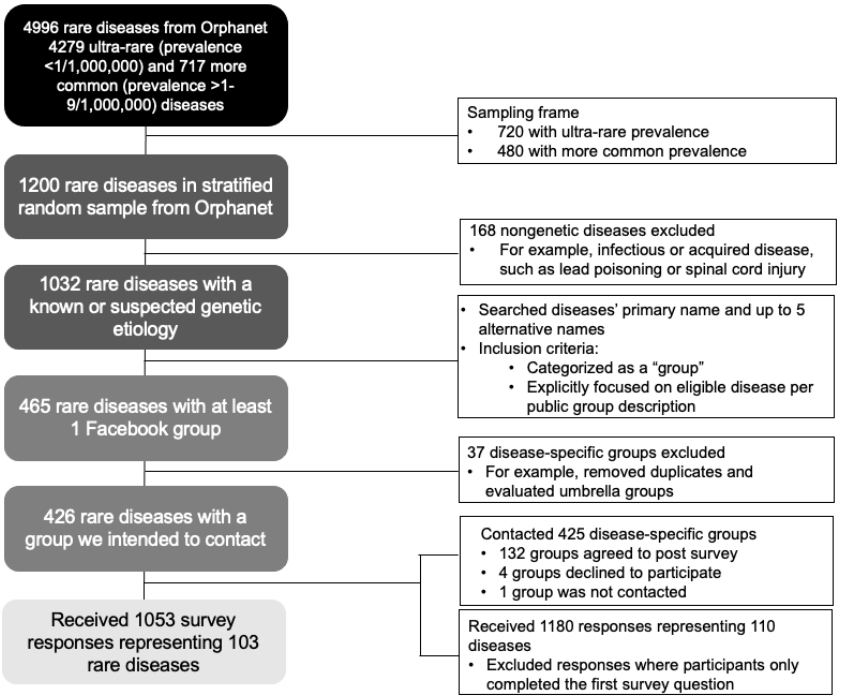


Table 1. Select social media group member demographics (N=1035).

Characteristics	Participants
Group, n (%)	
Patient	660 (62.7)
Family member	393 (37.3)
Age (years), median (IQR)	43 (35-52)
Gender, n (%)	
Female	868 (82.4)
Male	153 (14.5)
Other	13 (1.3)
Missing	19 (1.8)
Hispanic[a], n (%)	
No	982 (93.3)
Yes	59 (5.6)
Missing	12 (1.1)
Race[a], n (%)	
White	957 (90.9)
Black or African American	28 (2.7)
Asian or Asian American	60 (5.7)
American Indian or Alaskan Native	18 (1.7)
Native Hawaiian or other Pacific Islander	2 (0.2)
Some other race	27 (2.6)
Missing	15 (1.4)
Location, n (%)	
United States	615 (58.4)
Outside the United States	287 (27.3)
Missing	151 (14.3)
Highest level of education, n (%)	
Less than high school	11 (1.0)
High school or General Educational Development (GED)	129 (12.3)
Some college or associate degree	265 (25.2)
Bachelor’s degree	300 (28.5)
Advanced or graduate-level coursework or degree	299 (28.4)
Missing	49 (4.7)
Household income[b] (US $), n (%)	
≤25,000	80 (7.6)
25,001-50,000	170 (16.1)
50,001-100,000	252 (23.9)
100,001-200,000	388 (36.8)
Prefer not to say/don’t know	114 (10.8)
Missing	49 (4.7)
Disease prevalence[b], n (%)	
Unknown	155 (14.8)

Characteristics	Participants
<1 in 1,000,000	78 (7.4)
1-9 in 1,000,000	172 (16.3)
1-9 in 100,000	418 (39.7)
1-9 in 10,000	217 (20.6)
>1 in 1000	13 (1.2)
Facebook group disease specification, n (%)	
Specific-to-rare disease	1021 (97.0)
Umbrella rare disease	32 (3.0)
Size of group, median (IQR)[b]	1400 (765-2800)
Number of new posts per month, median (IQR)[b]	36 (19-120)
Number of new members per week, median (IQR)[b]	4 (1-9)

[a]Participants can select more than 1 response.

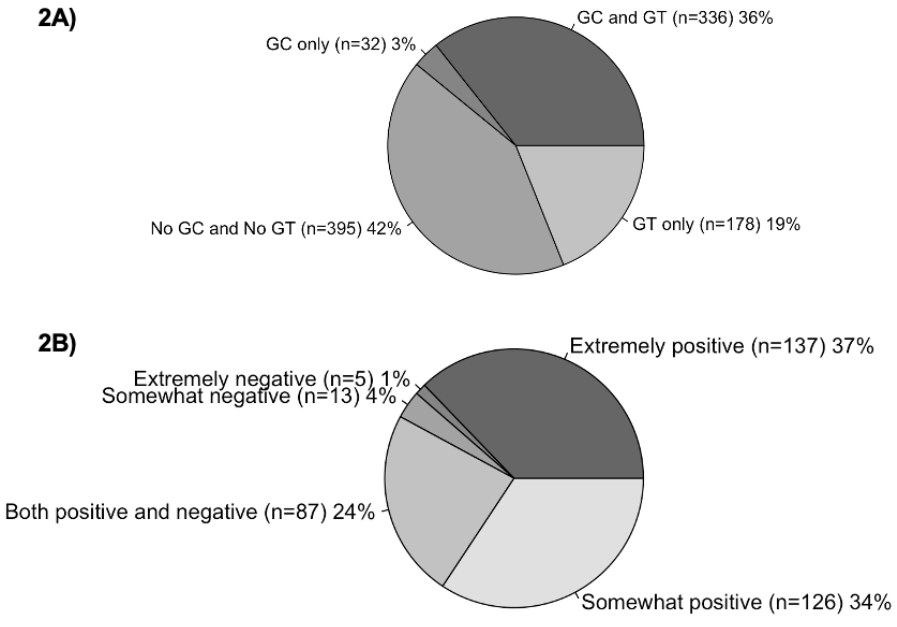
[b]Information about the rare disease and social media group was integrated into the participant-level data.

The rare diseases represented in our final sample included 17 distinct disease classifications and varied widely in the reported age of onset and inheritance pattern (Supplementary Table 3 in Multimedia Appendix 2). Social media groups included had a median group size of 1400 members (IQR 765-2800), 36 new posts per month (IQR 19-120), and 4 new members per week (IQR 1-9). Additional data on rare diseases and social media groups at both the individual participant and group levels are provided in Supplementary Tables 2 and 3 in Multimedia Appendix 2, respectively.

Access and Experience with Genetic Counseling and Testing

Participants varied widely in their previous access to GC and GT (Figure 2). Across the sample, 35.7% (n=336) of the participants reported receiving both GC and GT prior to the study, with an additional 18.9% (n=178) reporting only GT and 3.4% (n=32) only GC. The remaining 42.0% (n=396) of the participants had neither met with a genetic counselor nor received GT in the past. Among those who met with a genetic counselor in the past (n=368, 34.9%), the majority reported having somewhat or extremely positive experiences (n=263, 71.5%). The proportion of participants who reported knowing the specific genetic variant that caused their rare disease (n=422, 40.1%) was similar to those without a molecular diagnosis (n=429, 40.7%); see Supplementary Table 2 in Multimedia Appendix 2. The remaining participants reported having only a partial diagnosis or variants of uncertain significance (n=69, 6.6%), designated “other” on the survey item and elaborated further (n=34, 3.2%), or did not respond (n=99, 9.4%).

Figure 2. Prior experience with GC and GT. (A) Responses to questions about prior experience with GC or GT services (n=941, 89.4%). The excluded participants either chose not to respond or only responded to 1 question and not the other. (B) Respondents who indicated that they have met with a genetic counselor in the past (n=368, 34.9%) were prompted to describe their previous interactions with genetic counselors on a 5-point Likert scale. GC: genetic counseling; GT: genetic testing.

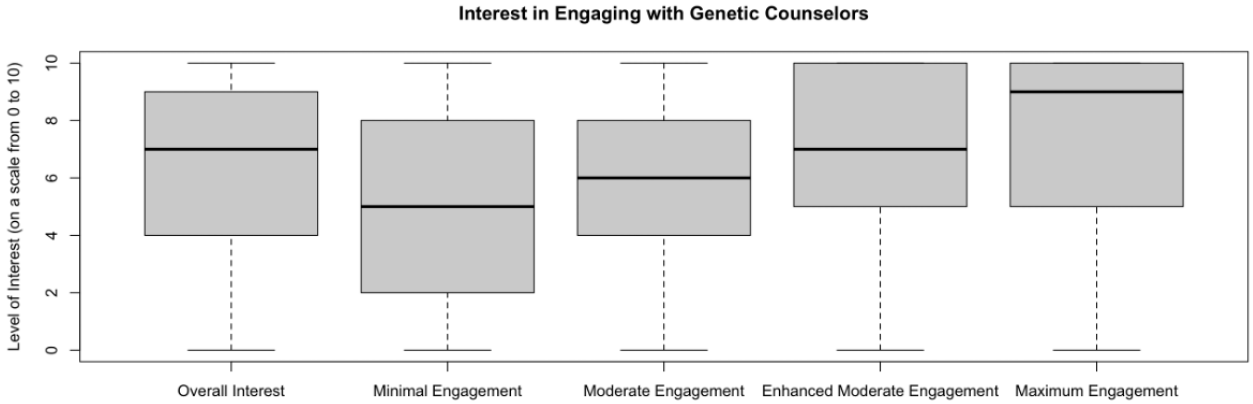


Interest in Engaging with Genetic Counselors Online

Participants reported moderately high connectedness with their Facebook support group, with a mean social media connectedness score of 3.74 out of 5 (SD 0.83); see Table 1. Although a subset of participants reported primarily seeking either social and emotional support (n=211, 20%) or informational support related to medical management (n=132, 12.5%), nearly half (n=489, 46.4%) emphasized the equal value of both types of support from social media groups.

Overall interest in engaging with genetic counselors on social media was also high, with a median score of 7 out of 10 (IQR 4-9); see Figure 3. When asked for their interest in 4 different models of engagement with varying degrees of access to genetic counselors, the participants' level of interest increased as the extent of direct access to a genetic counselor increased. This was reflected in a median interest score of 9 out of 10 (IQR 5-10) for the model with maximum engagement compared to a score of 5 out of 10 (IQR 2-8) for the model with the most minimal engagement.

Figure 3. Interest in engaging with genetic counselors. Respondents were asked to indicate how interested they were in engaging with genetic counselors overall and in varying levels of engagement on an 11-point scale from 0 (not at all interested) to 10 (extremely interested). Overall interest: I am interested in interacting with a genetic counselor on social media. Minimal engagement: The genetic counselor sends information and resources to the moderator of the group but is unavailable to answer specific questions. Moderate engagement: The genetic counselor communicates directly with the moderator of the group and addresses questions that the moderator requests the genetic counselor's input on. Enhanced moderate engagement: The genetic counselor is a member of the group and can post information and resources to the group as they see fit. Maximum engagement: The genetic counselor is a member of the group and can answer questions directly from group members.



Bivariate analysis results of individual participant, group, and rare disease characteristics hypothesized to correlate with higher interest in engaging on social media are summarized in Supplementary Table 4 in Multimedia Appendix 2. Most notably, participants who had not previously met with a genetic counselor expressed greater interest in engaging with genetic counselors online (μ=6.47) than those who did have prior experience (μ=6.01, t_{837}=2.09, *P*=.04). Additionally, participants who lived outside the United States expressed greater interest in engaging with genetic counselors online (μ=6.68) than those who lived within the United States (μ=6.19, t_{743}=1.96, *P*=.05). All other associations assessed were not statistically significant (*P*>.05).

Perceived Benefits and Drawbacks of Engaging with Genetic Counselors Through Social Media

A total of 732 (69.5%) participants contributed a written response to at least 1 open-ended question asking how they wanted to interact with genetic counselors on social media, what they were looking for in these interactions, and why they wanted to engage. When asked about the forms of engagement they most desired, 1-on-1 individual meetings were most preferred (n=399, 61.3%), followed by group-based interactions via social media (n=243, 37.3%); see Table 2. Primary resources the participants hoped to access through engaging with genetic counselors included answers to questions about their specific disease, such as inquiries about the impact of the disease on their family (n=83, 12.5%), available treatment (n=74, 11.1%), and GT (n=123, 18.5%). The primary benefit of social media engagement with genetic counselors was increased accessibility of information (n=189, 28.4%), followed by increased reliability of available information (n=184, 27.7%). The primary concern raised regarding engagement with genetic counselors on social media was the lack of a personal relationship between the patient and the genetic counselor (n=202, 40.2%), with concerns about privacy and confidentiality (n=90, 17.9%) and lack of trust (n=71, 14.1%) also frequently cited. Additional subthemes are illustrated in Table 2.

Table 2. Thematic analysis of open-ended questions.

Themes and subthemes	Participants, n (%)	Illustrative quotation
How do social media group members want to engage with genetic counselors? (n=651)		
1-on-1 interactions	399 (61.3)	• "Individual sessions, through messages or call (voice or video)" [participant (P)193]
Group-based interactions	243 (37.3)	• "Closed Facebook group or a more secure location for group if possible while being user friendly" [P289]
Information only	47 (7.2)	• "Provide resources but not directly answering questions" [P1067]
Through moderator	7 (1.1)	• "Best is for moderator to pass along information and let members contact GCs if they would like" [P9]
Do not bother	23 (3.5)	• "I don't, I think it's a private matter that should be discussed in an office setting" [P466]
What type of support are social media group members looking for? (n=665)		
Available to answer questions	278 (41.8)	• "To be able to answer questions when needed" [P1042]
Available to answer questions: inquiries about family members	83 (12.5)	• "Would like info I could share with my offspring and extended (family)" [P293]
Available to answer questions: inquiries about prognosis	34 (5.1)	• "Information about what to expect" [P1146]
Available to answer questions: inquiries about treatment	74 (11.1)	• "Information to get latest treatment options" [P1051]
Access to services	123 (18.5)	• "Logistical support about how and where to get testing done (and how to pay for it)" [P325]
Research/clinical trials	94 (14.1)	• "Connecting us with up to date information on genetic studies about the disease, opportunities to participate in research studies" [P1070]
Unsure	106 (15.9)	• "I'm not sure what a GC knows or has to offer me" [P830]
What are benefits of engaging with genetic counselors on social media? (n=665)		
Accessible/convenient	189 (28.4)	• "Benefits would be access to the knowledge or advice easier than waiting for your yearly appointment with the GCs" [P957]
Reliable information	184 (27.7)	• "To have someone who is an expert" [P843]
Psychosocial support	100 (15.0)	• "Better understanding and less anxiety" [P748]
No benefits	40 (6.0)	• "None! We don't need counselors, we need genetic testing!" [P751]
What are drawbacks of engaging with genetic counselors on social media? (n=503)		
Lack of personal relationship	202 (40.2)	• "There may be some disconnect between the patient & counselor due to them not being in the same location. It can be difficult to pick up on all the silent communication cues when online - even with video conferencing" [P73]
No drawbacks	93 (18.5)	• "Really can't think of any drawbacks" [P99]
Privacy/confidentiality	90 (17.9)	• "Confidentiality, I enjoy having nonmedicalized spaces to discuss my condition and having a provider there re-medicalizes it" [P128]
Lack of trust	71 (14.1)	• "Lack of trust in someone you can't see face to face, general mistrust of giving info to an unknown internet contact" [P353]

Themes and subthemes	Participants, n (%)	Illustrative quotation
Irrelevant information	39 (7.8)	• "Wrong info could be given for what's accessible where you live" [P145]
Frightening information	21 (4.2)	• "Fear of what I might learn about my future" [P11]
Other	26 (5.2)	• "Feeling like you don't need to go to appointments or the doctor because you found info online" [P483] • "Too much asked of the counselors" [P444]

Discussion

Principal Findings

Our findings from a large survey of patients with rare diseases and their family members across 103 rare disease social media support groups showed high interest in engaging with genetic counselors through social media. Participants who had never met with a genetic counselor in the past expressed greater overall interest in communicating with one on social media than those who had met with a genetic counselor. Moreover, participants preferred models in which genetic counselors were engaged and interactive. Those who elaborated on their interest expressed a desire for 1-on-1, personalized support from a genetic counselor, who they perceived to be a reliable source of information about their rare disease.

Our results suggest higher levels of interest in engaging on social media with genetic counselors than have previously been reported, though this may be attributable to our focus on the perspectives of current social media users [9]. The qualitative data we obtained further reinforced participants' interest in patient–genetic counselor interactions to find answers to their specific questions from reliable sources. Our findings also highlight patient concerns about privacy and confidentiality that may continue to discourage some from engaging with genetic counselors on social media [7,9,33].

Our findings also provide insights regarding access to GC and GT across rare diseases. Many participants within our cohort received GT but had never met with a genetic counselor to explain the test results or to obtain additional information at the time of results disclosure. This suggests that the global scaling-up of GC services is not occurring fast enough to match the expanding implementation of genetic and genomic technologies in the clinical setting. Given that GC services tend to be delivered less systematically in low- and middle-income countries due to costs [34], this gap in care is likely to disproportionately impact patients with rare diseases who are already underserved within the health care system globally [6].

Our findings also have implications for both clinical practice and policy. Many physicians, health care organizations, and nonprofits are already using social media to disseminate health information directly to patient communities for free or at minimal cost [35,36]. For example, providers in the fields of hematology and oncology use social media to provide medical education, rapidly disseminate new information, and encourage patients to engage in their health care [37,38]. Within the clinical context, a genetic counselor can provide patients with accessible information about a given condition, whenever available, and evaluate a patient's or family's response to the information. Participants' interest in high levels of engagement points to the informational support a genetic counselor may uniquely be able to provide in an online setting. Genetic counselors are often employed by academic medical centers, private and public hospitals, and diagnostic laboratories, and their services are charged to health insurance payers for eligible individuals. The type of support participants requested in their written responses accurately underlined roles and responsibilities that fall within the genetic counselors' scope of practice (eg, information about recurrence risk, prognosis, treatment, access to GT). However, the extent to which individualized support is desired would likely require more time commitment than a genetic counselor could feasibly provide outside of work hours. Within the context of this particular study, the proposed models of engagement with genetic counselors on social media implied that a genetic counselor would be available to provide support at no cost to the social media support group. However, the requested 1-on-1 interactions are essentially the equivalent of a clinical consultation and may not be sustainable without appropriate compensation for this form of service. There is little legal precedent to inform recommendations for engaging potential patients on social media, though genetic counselors should be aware of state and federal legal requirements in place that may prohibit such engagement [39]. By raising awareness of GC, some individuals may be more able to seek out these clinical services than others, given discrepancies in access by geographic location [11].

Further, genetic counselors engaging with social media groups would need to be careful to avoid providing medical advice outside of their scope of practice. A genetic counselor interested in engaging with these social media groups could potentially manage providing informational support, but genetic counselors are unlikely to be able to provide the higher level of engagement and dynamic dialogue patients and family members desire. Current technological advances, such as artificial intelligence, are being investigated as a potential means for delivering services in both health care broadly and within the field of genomics, which could bolster different approaches to addressing the needs of rare disease communities [40,41]. This may be more achievable from a time and labor standpoint, but the informational gaps may not be as amenable to this type of support as more well-established and researched diagnoses, such as cancer [42].

Although engagement through social media may be able to fill some of the gaps in knowledge that arise when patients obtain

a diagnosis but are unable to meet with a genetic counselor, a professional may need to actively contribute to the platform's knowledge base to effectively fill these gaps. Studies examining the impact of these efforts in other medical specialties suggest the need to establish best-practice guidelines that address both the provider's motivations and their ability to set boundaries [8,41,42]. Common guidelines for the use of social media by health care providers also highlight the responsibility to only share information from credible sites, refute any inaccurate information encountered, and use the most secure privacy settings available [43]. This points to the practical and logistical concerns that have already arisen in this setting, including balancing providing broad medical information that might help inform decision-making and avoiding the direct provision of medical advice.

Current guidelines, to the extent that they exist within the genetic counselor profession, encourage genetic counselors to be aware of concerns for patient and provider privacy on social media platforms and establish ethical and professional boundaries for themselves. However, there are no guidelines for genetic counselors that demonstrate what this might look like in practice [7,9,10]. Professional organizations, such as the National Society of Genetic Counselors (NSGC) and the American College of Medical Genetics (ACMG), can play a key role in leading the discussion to provide this support. Guidelines will need to include recommendations regarding the types of information to be communicated, the extent of the engagement, and how providers should address ethical concerns that may arise while they are acting as a liaison for the health care system online. It also will be critical to involve patient stakeholders in the creation of these guidelines to determine the best step forward. Although our study suggests high interest in relationships with genetic counselors on social media among the patient and family communities, clearer guidance is needed to address the systems-level issues and concerns genetic counselors may have.

There are clearly many unanswered questions that must be explored in greater depth before patient–genetic counselor interactions through social media are pursued by both interested parties. Further exploration is necessary to consider the goals and outcomes of having a genetic counselor engage with these groups and how success could be measured. Although there is a dire need for alternative approaches to providing patients with rare diseases with reliable sources of information, patient interest alone does not serve as an indication that this is a feasible option for genetics professionals. These interactions are not meant to replace the current structure and content of GC services, but genetic counselors could use social media as a communication tool for addressing gaps in knowledge and awareness about genetics services and gaps in accessible patient information on a global scale.

Limitations

Our study has several limitations. First, we chose to focus on current social media users, and therefore, our data cannot inform our understanding of perspectives of those who are not currently using social media. Second, we were unable to accurately calculate the response rate because we were unaware of how many people viewed the survey throughout the duration of recruitment. Although we collected the number of members in each group at the time of recruitment, this gives little to no indication of the number of active users of these groups who might have seen the study but opted not to respond. Third, we cannot be certain whether all participants have access to GC services in their respective countries. We did not collect the specific countries in which our participants reside in order to understand who undergoes GC and how these services are used in different countries. It also is possible (or even likely) that participants' interest in engaging with genetic counselors through social media also reflects their interest in accessing a genetic counselor in any context.

Using social media as a tool for recruitment is also known to result in a lack of gender, ethnic, racial, and socioeconomic diversity, which is reflected in the sociodemographic characteristics of our sample [20,44]. Compared to the population of Facebook users, a greater proportion of participants in our sample identified as White (67% vs 91%) and female (77% vs 82%), while a smaller proportion reported having household incomes over US $75,000 (73% vs 60%) or a college degree (73% vs 57%) [44]. Additional research is needed to ensure inclusion of diverse perspectives, including of those who do not participate in social media support groups. Furthermore, the heterogeneity and lack of systematic characterization of the rare disease community at large also made it difficult to assess the extent to which our sample captured and reflected key points of variation across the rare diseases. A complete analysis of nonresponders should be performed to further investigate the social media group and rare disease characteristics not represented here in this study.

Conclusion

The results of this study demonstrate that patients with rare diseases and their family members are interested in engaging with genetic counselors on social media as a tool to bridge the current gaps in access to genetics resources. However, the extent to which they desire 1-on-1 interactions raises privacy and confidentiality concerns, as well questions of the scope of practice associated with patient-provider interactions on social media. The data presented in this study therefore illustrate the need for guidelines to facilitate these interactions and to advance the conversation within the genetics community about the use of social media as an opportunity for engagement and information dissemination to meet the variegated, evolving, and complex needs of patients with rare diseases.

Acknowledgments

This study was completed in partial fulfillment of the requirements for the first author's master of science degree in human genetics and genetic counseling from Stanford University. We would like to thank the patients, family members, and communities for their participation in our study. We thank Troy Evans, Sarah Marshall, and Claire Richmond for pretesting our novel questions.

We also thank Ying Jin and Anav Sood, PhD candidates in the Stanford University Department of Statistics, and Katharine Rendle, PhD, MSW, MPH, from the University of Pennsylvania Department of Family Medicine and Community Health and Department of Biostatistics, Epidemiology, and Informatics for their help with statistical analysis. We would also like to thank Charis O Tang for assistance in qualitative data analysis and the attendees of the Stanford Center for Biomedical Ethics Works in Progress session, especially Jennifer L Young, PhD, and Daphne O Martschenko, PhD, for their thoughtful input.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete survey instrument.
[DOCX File , 39 KB - jmir_v24i12e42084_app1.docx]

Multimedia Appendix 2
Supplementary tables.
[DOCX File , 34 KB - jmir_v24i12e42084_app2.docx]

References

1. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, Rehm H, et al. How many rare diseases are there? Nat Rev Drug Discov 2020 Feb;19(2):77-78 [FREE Full text] [doi: 10.1038/d41573-019-00180-y] [Medline: 32020066]
2. Nguengang Wakap S, Lambert D, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet 2020 Feb;28(2):165-173 [FREE Full text] [doi: 10.1038/s41431-019-0508-0] [Medline: 31527858]
3. Sawyer S, Hartley T, Dyment DA, Beaulieu CL, Schwartzentruber J, Smith A, FORGE Canada Consortium, Care4Rare Canada Consortium, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. Clin Genet 2016 Mar;89(3):275-284 [FREE Full text] [doi: 10.1111/cge.12654] [Medline: 26283276]
4. Grant P, Langlois S, Lynd LD, GenCOUNSEL Study, Austin JC, Elliott AM. Out-of-pocket and private pay in clinical genetic testing: a scoping review. Clin Genet 2021 Nov;100(5):504-521 [FREE Full text] [doi: 10.1111/cge.14006] [Medline: 34080181]
5. Phillips K, Douglas M, Wordsworth S, Buchanan J, Marshall DA. Availability and funding of clinical genomic sequencing globally. BMJ Glob Health 2021 Feb;6(2):e004415 [FREE Full text] [doi: 10.1136/bmjgh-2020-004415] [Medline: 33574068]
6. Stoller JK. The challenge of rare diseases. Chest 2018 Jun;153(6):1309-1314 [FREE Full text] [doi: 10.1016/j.chest.2017.12.018] [Medline: 29325986]
7. Gallagher L, McCuaig J, Benoit L, Davies C. It's time for the genetic counseling profession to embrace social media. J Genet Couns 2016 Dec;25(6):1338-1341 [FREE Full text] [doi: 10.1007/s10897-016-9950-0] [Medline: 27026235]
8. Patch C, Middleton A. Genetic counselling in the era of genomic medicine. Br Med Bull 2018 Jun 01;126(1):27-36 [FREE Full text] [doi: 10.1093/bmb/ldy008] [Medline: 29617718]
9. Moore R, Matthews A, Cohen L. How might the genetics profession better utilize social media. J Genet Couns 2018;27:493-504 [FREE Full text] [doi: 10.1007/s10897-018-0215-y]
10. Rocha H, Savatt J, Riggs ER, Wagner JK, Faucett WA, Martin CL. Incorporating social media into your support tool box: points to consider from genetics-based communities. J Genet Couns 2018 Apr;27(2):470-480 [FREE Full text] [doi: 10.1007/s10897-017-0170-z] [Medline: 29130143]
11. Abacan M, Alsubaie L, Barlow-Stewart K, Caanen B, Cordier C, Courtney E, et al. The global state of the genetic counseling profession. Eur J Hum Genet 2019 Feb;27(2):183-197 [FREE Full text] [doi: 10.1038/s41431-018-0252-x] [Medline: 30291341]
12. Hoskovec J, Bennett R, Carey ME, DaVanzo JE, Dougherty M, Hahn SE, et al. Projecting the supply and demand for certified genetic counselors: a workforce study. J Genet Couns 2018 Feb;27(1):16-20 [FREE Full text] [doi: 10.1007/s10897-017-0158-8] [Medline: 29052810]
13. Stoll K, Kubendran S, Cohen SA. The past, present and future of service delivery in genetic counseling: keeping up in the era of precision medicine. Am J Med Genet C Semin Med Genet 2018 Mar;178(1):24-37 [FREE Full text] [doi: 10.1002/ajmg.c.31602] [Medline: 29512888]
14. Amendola L, Golden-Grant K, Scollon S. Scaling genetic counseling in the genomics era. Annu Rev Genomics Hum Genet 2021 Aug 31;22:339-355 [FREE Full text] [doi: 10.1146/annurev-genom-110320-121752] [Medline: 33722076]
15. Pelentsov L, Laws T, Esterman AJ. The supportive care needs of parents caring for a child with a rare disease: a scoping review. Disabil Health J 2015 Oct;8(4):475-491 [FREE Full text] [doi: 10.1016/j.dhjo.2015.03.009] [Medline: 25959710]
16. Fergie G, Hilton S, Hunt K. Young adults' experiences of seeking online information about diabetes and mental health in the age of social media. Health Expect 2016 Dec;19(6):1324-1335 [FREE Full text] [doi: 10.1111/hex.12430] [Medline: 26647109]

17. Nicholl H, Tracey C, Begley T, King C, Lynch AM. Internet use by parents of children with rare conditions: findings from a study on parents' web information needs. J Med Internet Res 2017 Feb 28;19(2):e51 [FREE Full text] [doi: 10.2196/jmir.5834] [Medline: 28246072]
18. Jaffe A, Zurynski Y, Beville L, Elliott E. Call for a national plan for rare diseases. J Paediatr Child Health 2010 Jan;46(1-2):2-4 [FREE Full text] [doi: 10.1111/j.1440-1754.2009.01608.x] [Medline: 19943870]
19. Deuitch N, Beckman E, Halley MC, Young JL, Reuter CM, Kohler J, Undiagnosed Diseases Network, et al. "Doctors can read about it, they can know about it, but they've never lived with it": how parents use social media throughout the diagnostic odyssey. J Genet Couns 2021 Dec;30(6):1707-1718 [FREE Full text] [doi: 10.1002/jgc4.1438] [Medline: 34096130]
20. Miller E, Woodward A, Flinchum G, Young JL, Tabor HK, Halley MC. Opportunities and pitfalls of social media research in rare genetic diseases: a systematic review. Genet Med 2021 Dec;23(12):2250-2259 [FREE Full text] [doi: 10.1038/s41436-021-01273-z] [Medline: 34282302]
21. Orphanet. The Portal for Rare Diseases and Orphan Drugs. URL: http://www.orpha.net/consor/www/cgi-bin/index.php?lng=EN [accessed 2022-12-06]
22. Austin C, Cutillo C, Lau LPL, Jonker AH, Rath A, Julkowska D, International Rare Diseases Research Consortium (IRDiRC). Future of rare diseases research 2017-2027: an IRDiRC perspective. Clin Transl Sci 2018 Jan;11(1):21-27 [FREE Full text] [doi: 10.1111/cts.12500] [Medline: 28796445]
23. Titgemeyer S, Schaaf CP. Facebook support groups for rare pediatric diseases: quantitative analysis. JMIR Pediatr Parent 2020 Nov 19;3(2):e21694 [FREE Full text] [doi: 10.2196/21694] [Medline: 33211019]
24. NORD. What's Happening at NORD. URL: https://rarediseases.org/ [accessed 2022-12-06]
25. Adam MP, Ardinger HH, Pagon RA. GeneReviews®. Seattle, WA: University of Washington; 1993.
26. Halley M, Stanley T, Maturi J, Goldenberg AJ, Bernstein JA, Wheeler MT, et al. "It seems like COVID-19 now is the only disease present on Earth": living with a rare or undiagnosed disease during the COVID-19 pandemic. Genet Med 2021 May;23(5):837-844 [FREE Full text] [doi: 10.1038/s41436-020-01069-7] [Medline: 33420343]
27. Halley M, Mathews K, Diamond LC, Linos E, Sarkar U, Mangurian C, et al. The intersection of work and home challenges faced by physician mothers during the coronavirus disease 2019 pandemic: a mixed-methods analysis. J Womens Health (Larchmt) 2021 Apr;30(4):514-524 [FREE Full text] [doi: 10.1089/jwh.2020.8964] [Medline: 33761277]
28. Centers for Disease Control and Prevention (CDC). Genetic Counseling. URL: https://www.cdc.gov/genomics/gtesting/genetic_counseling.htm [accessed 2022-07-12]
29. Centers for Disease Control and Prevention (CDC). Genetic Testing. URL: https://www.cdc.gov/genomics/gtesting/genetic_testing.htm [accessed 2022-07-12]
30. Linos E, Lasky-Fink J, Halley M, Sarkar U, Mangurian C, Sabry H, et al. Impact of sexual harassment and social support on burnout in physician mothers. J Womens Health (Larchmt) 2022 Jul;31(7):932-940 [FREE Full text] [doi: 10.1089/jwh.2021.0487] [Medline: 35730998]
31. The R Foundation. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2022-12-06]
32. Miles MB, Huberman AM, Saldana J. Qualitative Data Analysis: A Methods Sourcebook, Fourth Edition. Thousand Oaks, CA: Sage; 2014.
33. Omaggio N, Baker M, Conway LJ. Have you ever googled a patient or been friended by a patient? Social media intersects the practice of genetic counseling. J Genet Couns 2018 Apr;27(2):481-492 [FREE Full text] [doi: 10.1007/s10897-017-0206-4] [Medline: 29333562]
34. Zhong A, Darren B, Loiseau B, He LQB, Chang T, Hill J, et al. Ethical, social, and cultural issues related to clinical genetic testing and counseling in low- and middle-income countries: a systematic review. Genet Med 2021 Dec;23(12):2270-2280 [FREE Full text] [doi: 10.1038/s41436-018-0090-9] [Medline: 30072741]
35. Moorhead S, Hazlett D, 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]
36. Farsi D. Social media and health care, part I: literature review of social media use by health care providers. J Med Internet Res 2021 Apr 05;23(4):e23205 [FREE Full text] [doi: 10.2196/23205] [Medline: 33664014]
37. Dizon D, Graham D, Thompson M, Johnson L, Johnston C, Fisch M, et al. Practical guidance: the use of social media in oncology practice. J Oncol Pract 2012 Sep;8(5):e114-e124 [FREE Full text] [doi: 10.1200/jop.2012.000610]
38. Thompson M, Majhail N, Wood WA, Perales MA, Chaboissier M. Social media and the practicing hematologist: Twitter 101 for the busy healthcare provider. Curr Hematol Malig Rep 2015 Dec;10(4):405-412 [FREE Full text] [doi: 10.1007/s11899-015-0286-x] [Medline: 26449718]
39. Farnan J, Snyder Sulmasy L, Worster BK, Chaudhry HJ, Rhyne JA, Arora VM, American College of Physicians Ethics, ProfessionalismHuman Rights Committee, American College of Physicians Council of Associates, Federation of State Medical Boards Special Committee on EthicsProfessionalism*. Online medical professionalism: patient and public relationships: policy statement from the American College of Physicians and the Federation of State Medical Boards. Ann Intern Med 2013 Apr 16;158(8):620-627 [FREE Full text] [doi: 10.7326/0003-4819-158-8-201304160-00100] [Medline: 23579867]

40. Kearney E, Wojcik A, Babu D. Artificial intelligence in genetic services delivery: utopia or apocalypse? J Genet Couns 2020 Feb;29(1):8-17 [FREE Full text] [doi: 10.1002/jgc4.1192] [Medline: 31749317]
41. 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]
42. Nazareth S, Hayward L, Simmons E. Hereditary cancer risk using a genetic chatbot before routine care visits. Obstet Gynecol 2021;138:860-870 [FREE Full text] [doi: 10.1097/aog.0000000000004596]
43. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T Peer-Rev J Formul Manag 2014;39:491-520.
44. Pew Research Center. Social Media Fact Sheet. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2022-12-05]

Abbreviations

ADOA: autosomal dominant optic atrophy
GC: genetic counseling
GT: genetic testing
IRB: Institutional Review Board

Edited by R Kukafka; submitted 22.08.22; peer-reviewed by K Backor, F Houyez; comments to author 27.09.22; revised version received 01.11.22; accepted 13.11.22; published 21.12.22.

Please cite as:
Yabumoto M, Miller E, Rao A, Tabor HK, Ormond KE, Halley MC
Perspectives of Rare Disease Social Media Group Participants on Engaging With Genetic Counselors: Mixed Methods Study
J Med Internet Res 2022;24(12):e42084
URL: https://www.jmir.org/2022/12/e42084
doi:10.2196/42084
PMID:36542454

Corrigenda and Addenda

Correction: Digital Device Exposure and Cognition Levels of Children in Low- and Middle-Income Countries: Cross-sectional Study in Cambodia

Hye Hyeon Kim[1*], PhD; JooHyun Lee[1*], BSN; Ho Heon Kim[1], RN, MS; Sangho Hwang[1], MS; Ilcheong Yi[2], PhD; Sambath Kao[3], MSc, MD; DooRa Kim[4], BS; Hyuk-Sang Sohn[5], PhD; Joohye Kim[6], PhD; Yejin Choi[4], BS; Sangchul Yoon[7], MD, PhD; Yu Rang Park[1], PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
[2]United Nations Research Institute for Social Development, Geneva, Switzerland
[3]Department of Malnutrition and Neurology, National Pediatric Hospital, Phnom Penh, Cambodia
[4]DoBrain Inc, Seoul, Republic of Korea
[5]Graduate School of Public Policy & Civic Engagement, Kyung Hee University, Seoul, Republic of Korea
[6]Department of Special Education, Baekseok University, Cheonan, Republic of Korea
[7]Department of Medical Humanities and Social Sciences, Yonsei University College of Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 10 5240 3434
Email: yurangpark@yuhs.ac

Related Article:

Correction of: https://www.jmir.org/2022/8/e31206/

(J Med Internet Res 2022;24(12):e44106) doi:10.2196/44106

In "Digital Device Exposure and Cognition Levels of Children in Low- and Middle-Income Countries: Cross-sectional Study in Cambodia (J Med Internet Res 2022;24(8):e31206)" the authors made one correction.

In the originally published article, Ilcheong Yi's country of organizational affiliation appeared as:

> *United Nations Research Institute for Social Development, Geneva, Swaziland*

It has now been corrected to:

> *United Nations Research Institute for Social Development, Geneva, Switzerland*

The correction will appear in the online version of the paper on the JMIR Publications website on December 6, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 06.11.22; this is a non–peer-reviewed article;accepted 21.11.22; published 06.12.22.

Please cite as:
Kim HH, Lee J, Kim HH, Hwang S, Yi I, Kao S, Kim D, Sohn HS, Kim J, Choi Y, Yoon S, Park YR
Correction: Digital Device Exposure and Cognition Levels of Children in Low- and Middle-Income Countries: Cross-sectional Study in Cambodia
J Med Internet Res 2022;24(12):e44106
URL: https://www.jmir.org/2022/12/e44106
doi:10.2196/44106
PMID:36473173

Corrigenda and Addenda

Correction: Ambient Assisted Living: Scoping Review of Artificial Intelligence Models, Domains, Technology, and Concerns

Mladjan Jovanovic[1], PhD; Goran Mitrov[2], MSc; Eftim Zdravevski[2], PhD; Petre Lameski[2], PhD; Sara Colantonio[3], PhD; Martin Kampel[4], PhD; Hilda Tellioglu[4], PhD; Francisco Florez-Revuelta[5], PhD

[1]Department of Computer Science, Singidunum University, Belgrade, Serbia

[2]Faculty of Computer Science and Engineering, University Saints Cyril and Methodius, Skopje, North Macedonia

[3]Signals & Images Lab, Institute of Information Science and Technologies, National Research Council of Italy, Pisa, Italy

[4]Faculty of Informatics, Vienna University of Technology, Vienna, Austria

[5]Department of Computing Technology, University of Alicante, Alicante, Spain

Corresponding Author:
Mladjan Jovanovic, PhD
Department of Computer Science, Singidunum University
Danijelova 32
Belgrade, 11000
Serbia
Phone: 381 603831844
Email: mjovanovic@singidunum.ac.rs

Related Article:

Correction of: https://www.jmir.org/2022/11/e36553/

(J Med Internet Res 2022;24(12):e45081) doi:10.2196/45081

In “Ambient Assisted Living: Scoping Review of Artificial Intelligence Models, Domains, Technology, and Concerns” (J Med Internet Res 2022;24(11):e36553) the authors noted one correction.

Under “Acknowledgments”, the sentence:

> *This work was part of and supported by GoodBrother, COST Action 19121—Network on Privacy-Aware Audio- and Video-Based Applications for Active and Assisted Living.*

has been replaced by:

> *This publication is based upon work from COST Action GoodBrother—Network on Privacy-Aware Audio- and Video-Based Applications for Active and Assisted Living (CA19121), supported by COST (European Cooperation in Science and Technology).*

The correction will appear in the online version of the paper on the JMIR Publications website on December 20, 2022 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 15.12.22; this is a non–peer-reviewed article;accepted 15.12.22; published 20.12.22.

Please cite as:
Jovanovic M, Mitrov G, Zdravevski E, Lameski P, Colantonio S, Kampel M, Tellioglu H, Florez-Revuelta F
Correction: Ambient Assisted Living: Scoping Review of Artificial Intelligence Models, Domains, Technology, and Concerns
J Med Internet Res 2022;24(12):e45081
URL: https://www.jmir.org/2022/12/e45081
doi:10.2196/45081
PMID:36538785

Original Paper

The Effect of Nonpharmaceutical Interventions Implemented in Response to the COVID-19 Pandemic on Seasonal Respiratory Syncytial Virus: Analysis of Google Trends Data

Hersh D Ravkin[1], BMedSc; Elad Yom-Tov[2,3], PhD; Lior Nesher[1,4], MD

[1]Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel

[2]Microsoft Research, Herzeliya, Israel

[3]Faculty of Industrial Engineering and Management, Technion, Haifa, Israel

[4]Infectious Disease Institute, Soroka University Medical Center, Beer Sheva, Israel

Corresponding Author:
Elad Yom-Tov, PhD
Microsoft Research
3 Alan Turing
Herzeliya, 4672415
Israel
Phone: 972 779391359
Email: eladyt@microsoft.com

Abstract

Background: Respiratory syncytial virus (RSV) is a major cause of respiratory infection in children. Despite usually following a consistent seasonal pattern, the 2020-2021 RSV season in many countries was delayed and changed in magnitude.

Objective: This study aimed to test if these changes can be attributed to nonpharmaceutical interventions (NPIs) instituted around the world to combat SARS-CoV-2.

Methods: We used the internet search volume for RSV, as obtained from Google Trends, as a proxy to investigate these abnormalities.

Results: Our analysis shows a breakdown of the usual correlation between peak latency and magnitude during the year of the pandemic. Analyzing latency and magnitude separately, we found that the changes therein are associated with implemented NPIs. Among several important interventions, NPIs affecting population mobility are shown to be particularly relevant to RSV incidence.

Conclusions: The 2020-2021 RSV season served as a natural experiment to test NPIs that are likely to restrict RSV spread, and our findings can be used to guide health authorities to possible interventions.

(J Med Internet Res 2022;24(12):e42781) doi:10.2196/42781

KEYWORDS

RSV; respiratory syncytial virus; search engine; Google Trends; Google; respiratory; children; pharmaceutical; intervention; COVID-19; pandemic; virus; infection; health

Introduction

Respiratory syncytial virus (RSV) is the leading cause of lower respiratory tract infection in children worldwide. Preterm gestation and several other underlying conditions particularly increase the risk of hospitalization and severe disease [1]. Although no specific treatment exists, prophylactic administration of monoclonal antibodies, when timed correctly, mitigates some of the risks. The American Academy of Pediatrics guidelines limit the duration of treatment with monoclonal antibodies to 5 months, with maximal benefit derived when treatment is initiated prior to the onset of the local RSV season [1]. Seasonality—including the start, peak, and end weeks—in RSV has been studied extensively and generally follows a set pattern within each country, with little variation from year to year. For the start week, even relatively major variations, when they rarely occur, do not exceed 1 month [2]. This regularity is key for proper timing of prophylaxis administration [1]. Beyond timing, a consistent spatiotemporal pattern of RSV epidemics has been established in previous years [3]. As there is no known animal reservoir of human RSV, transmission occurs solely through close contact with other

humans [4]. Changes in human behavior, therefore, are likely integral to the dynamics and seasonality of RSV epidemics.

Nonpharmaceutical interventions (NPIs) are policy-based strategies used to mitigate the effects of infectious diseases. When vaccines are unavailable, NPIs are the primary recourse for reducing transmission rate and decreasing the burden on health care systems. NPIs may be grossly categorized as personal, communal, or environmental [5]. While the latter 2 may be reasonably implemented across an entire population, it is difficult to enforce adherence to personal NPIs in very young children. This is a particularly important consideration in RSV, where young children are the primary at-risk group.

Surveillance of RSV outbreaks is not uniformly rigorous across the world [6]. Changes in health-seeking behaviors and viral surveillance during the COVID-19 pandemic further complicate the interpretation of epidemiological data [7]. However, previous research has shown that the volume of search engine queries can serve as a proxy for the incidence of respiratory diseases [8-11]. Initial attempts to harness internet search data to monitor viral incidence were shown to be naïve. For instance, Google Flu Trends, a system that predicted the influenza load from the Google search volume for specific terms, was shown to overestimate these loads [12]. However, work since then has improved the models that predict loads of influenza-like illness from these data [13,14]. In the case of RSV, the Google query volume for the term "RSV" has been demonstrated to be a good proxy for RSV incidence [15]. This correlation has been used to draw conclusions regarding the dynamics of RSV transmission when epidemiological data are insufficient [15,16].

In 2020, countries around the world instituted various NPIs to combat the COVID-19 pandemic [17]. Researchers have reported that the 2020-2021 RSV season was exceptional, in that its peak was both delayed and changed in magnitude [16,18-20]. Here, we re-establish the correlation between RSV incidence and internet search data and use the latter as a proxy to investigate the association of the various NPIs instituted in response to the COVID-19 pandemic with the abnormalities of the 2020-2021 RSV season.

Methods

Data Sources

Nonsentinel observational data on RSV incidence were obtained from the European Centre for Disease Prevention and Control's Surveillance Atlas of Infectious Diseases, an interactive tool that pools data collected from its member states through the European Surveillance System [21]. Data included in the study range from week 40 of 2014 to week 42 of 2021.

Search query volume data were gathered from Google Trends using the Google Trends Anchor Bank package [22]. Search query volume data for the Google Trends topic "Respiratory syncytial virus" were gathered for the period between week 9 of 2016 and week 43 of 2021.

Data on NPIs were taken from Worldwide Non-pharmaceutical Interventions Tracker for COVID-19 (WNTRAC). Briefly, WNTRAC is a comprehensive data set consisting of over 7000 NPIs implemented worldwide since the start of the COVID-19 pandemic. WNTRAC includes NPIs implemented in countries across the world, classifying them into a taxonomy of 16 NPI categories. NPI events are automatically extracted daily from Wikipedia articles using natural language processing techniques and are manually validated to ensure accuracy and veracity [17]. WNTRAC data up to December 17, 2021, are included in this study. Figures 1 and 2 provide a schematic overview of the data collection and processing.

Figure 1. Data sources and process used to validate correlation between RSV incidence and the internet search volume for RSV. ECDC: European Centre for Disease Prevention and Control; RSV: respiratory syncytial virus.

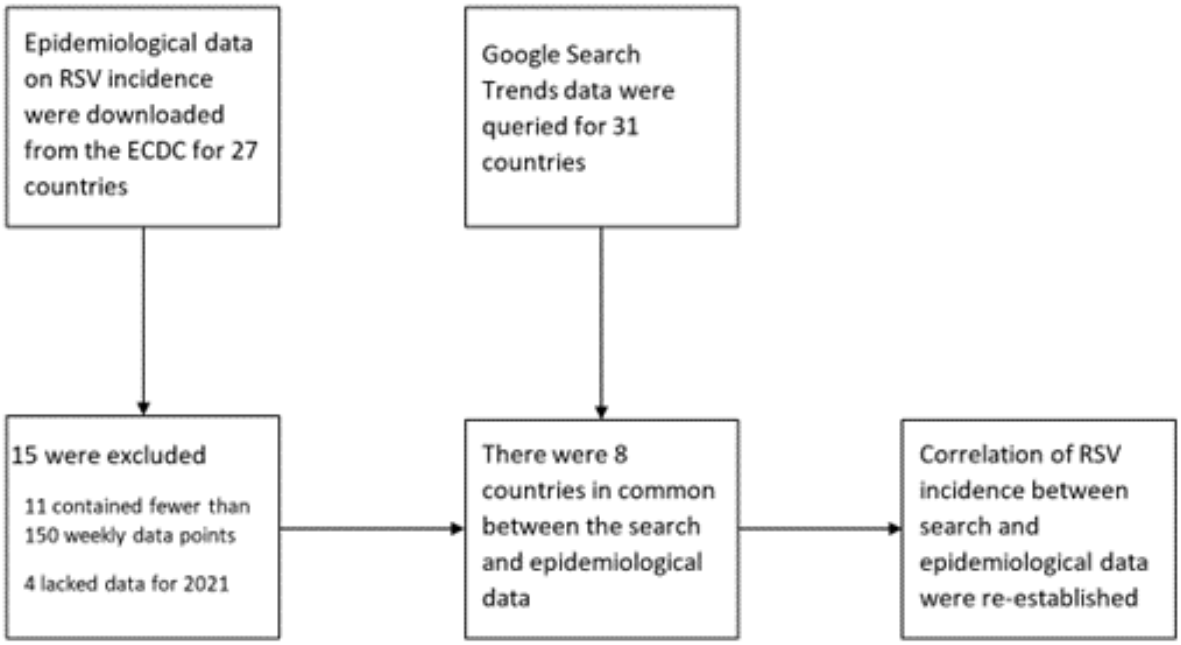
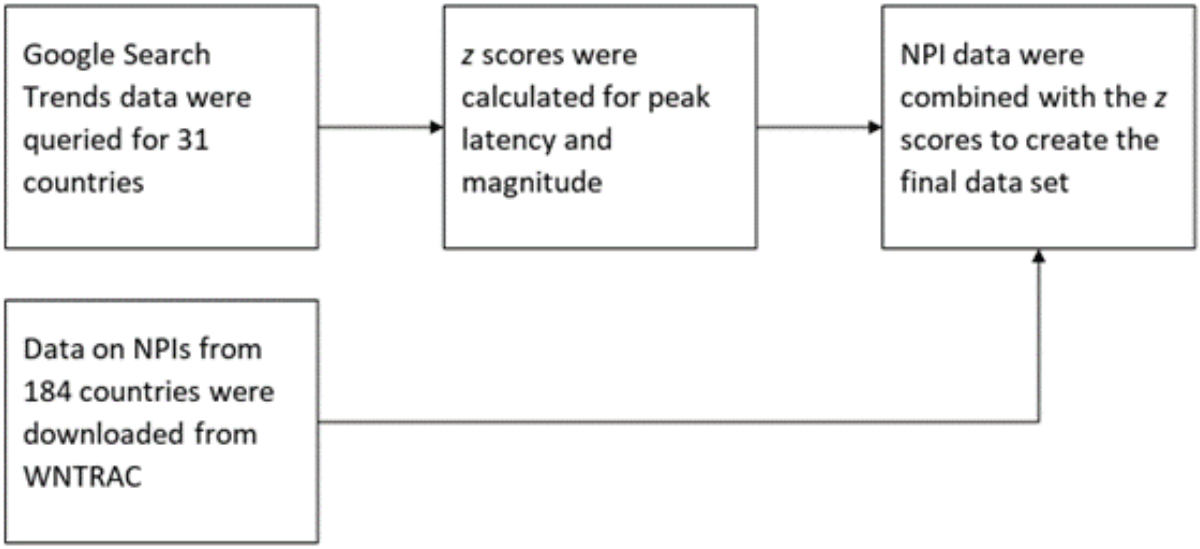


Figure 2. Data sources and processing used to generate the final data set. NPI: nonpharmaceutical intervention; WNTRAC: Worldwide Non-pharmaceutical Interventions Tracker for COVID-19.



Preprocessing of RSV Incidence Data

The original European Centre for Disease Prevention and Control data included 27 countries. Of these, 11 countries were excluded because they contained fewer than 150 weekly data points. An additional 4 countries were excluded because they contained no data for 2021, leaving a matrix of epidemiological data for 12 countries available for confirming the correlation between epidemiological and search trends data.

Weekly search trend data were collected for 31 countries using Google Trends Anchor Bank. Only countries with sufficient Google Trends data were included in the study. An effort was made to include countries from both hemispheres and various continents to improve generalizability. Data were smoothed to remove noise by calculating a rolling average with a 5-week moving window. Annual peaks were identified by calculating the local maxima of each country's weekly RSV incidence using a minimum horizontal distance of 35 samples between adjacent peaks (ie, a minimum of 35 weeks had to be present between adjacent peaks; this was achieved with the scipy.signal.find_peaks function, distance=35). Of the 31 countries for which search data were gathered, 8 were also present in the epidemiological data and could be used in the correlation analysis.

To allow comparison of the deviation from average of peak latency (ie, the extent to which the peak was delayed) and magnitude among different countries, standard scores were calculated for each in accordance with the function $Z = (x - \mu) / \sigma$, where Z is the standard score, x is the observed peak week or peak magnitude for 2021, μ is the mean peak per week divided by the magnitude as calculated on the basis of prior years included in the study, and σ is the SD value of the peak week divided by the magnitude as calculated on the basis of prior years included in the study. These standard scores represent the 2 target features (ie, outcome variables) used in the study.

Preprocessing of NPI Data

NPI data for the countries included in the study were obtained from WNTRAC. Several NPI types were recategorized to make them amenable to representation in a tabular format; namely, NPIs with unique values (eg, specific countries from which there were travel restrictions) and restrictions on mass gatherings. The NPIs with unique features were reorganized as either "some" or "all" based on whether they referred to restrictions pertaining to specific countries or blanket restrictions on all countries, respectively. Restrictions on mass gatherings, which originally displayed a specific numerical limit, were binned, grouping restrictions on gatherings of 10 or less, 10-100, 100-250, 250-500, and ≥500 persons. The NPI "changes in prison-related policies" was removed as it was instituted in very few countries and because symptomatic RSV predominantly affects young children. Additionally, the NPIs "declared state of emergency" and "contact tracing for COVID-19 patients" were deemed unrelated to the dynamics of the RSV outbreak and were therefore removed from the data set to decrease the effect of multicollinearity. NPI subtypes "other" and "na" were combined and recategorized as "unspecified" in the interest of interpretability, with no subsequent change in performance.

Rather than considering all NPIs instituted throughout the study period, NPIs included in the final matrix were limited only to those interventions instituted during the 3 months preceding the expected (average calculated over the past years included in the study) peak week.

Statistical Analysis

Spearman correlation analysis was used to calculate the correlation between epidemiological and search data. Correlation between peak magnitude and latency was determined by fitting a linear regression model.

Linear regression models were fitted to each of the two main target variables, peak latency and peak magnitude, after a subset of highly performing features was chosen through sequential backward selection. Backward selection was implemented using the scikit-learn SequentialFeatureSelector with default

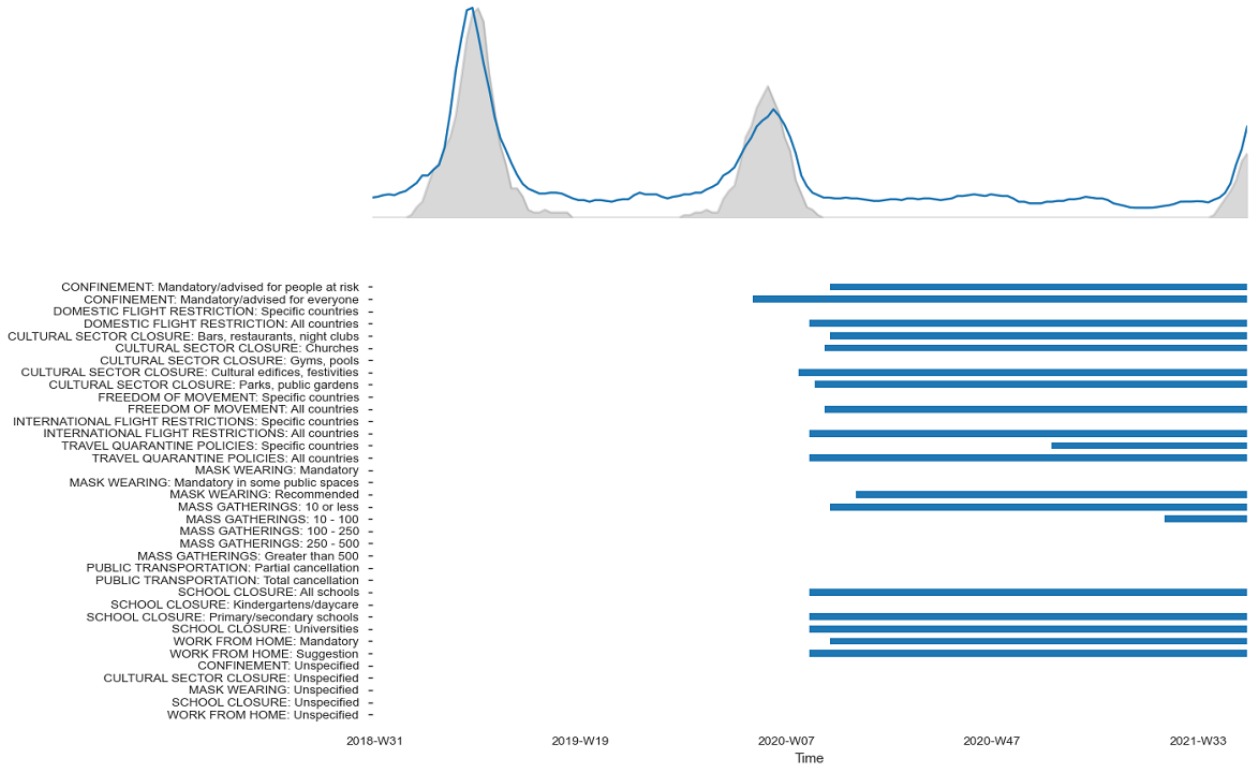
cross-validation and R^2 scoring. Relative contribution of the various NPIs to each of the outcomes was evaluated by calculating their feature importance using the SHAP (Shapley Additive Explanations) package (version 0.39.0), one of the most robust approaches currently available for explaining machine learning outputs [23]. The SHAP package's LinearExplainer was used to account for the correlation among various NPIs. For general pipeline development and validation, scikit-learn (version 0.22.1) was used. All analysis was conducted in Python (version 3.7.7).

Results

Overview

We first re-established the correlation between RSV case incidence and Google query volume for RSV, which was then used to infer RSV incidence for the 31 countries in our study. We then used a regression model to predict the normalized peak latency and the peak magnitude of RSV incidence in each country during the 2020-2021 season. Relative contribution of the various NPIs to each of the outcomes was then estimated using Shapley values. An example of the timing between the NPIs included in the study and the spread of RSV during the 2021-2022 season in one of the countries (Germany) is shown in Figure 3.

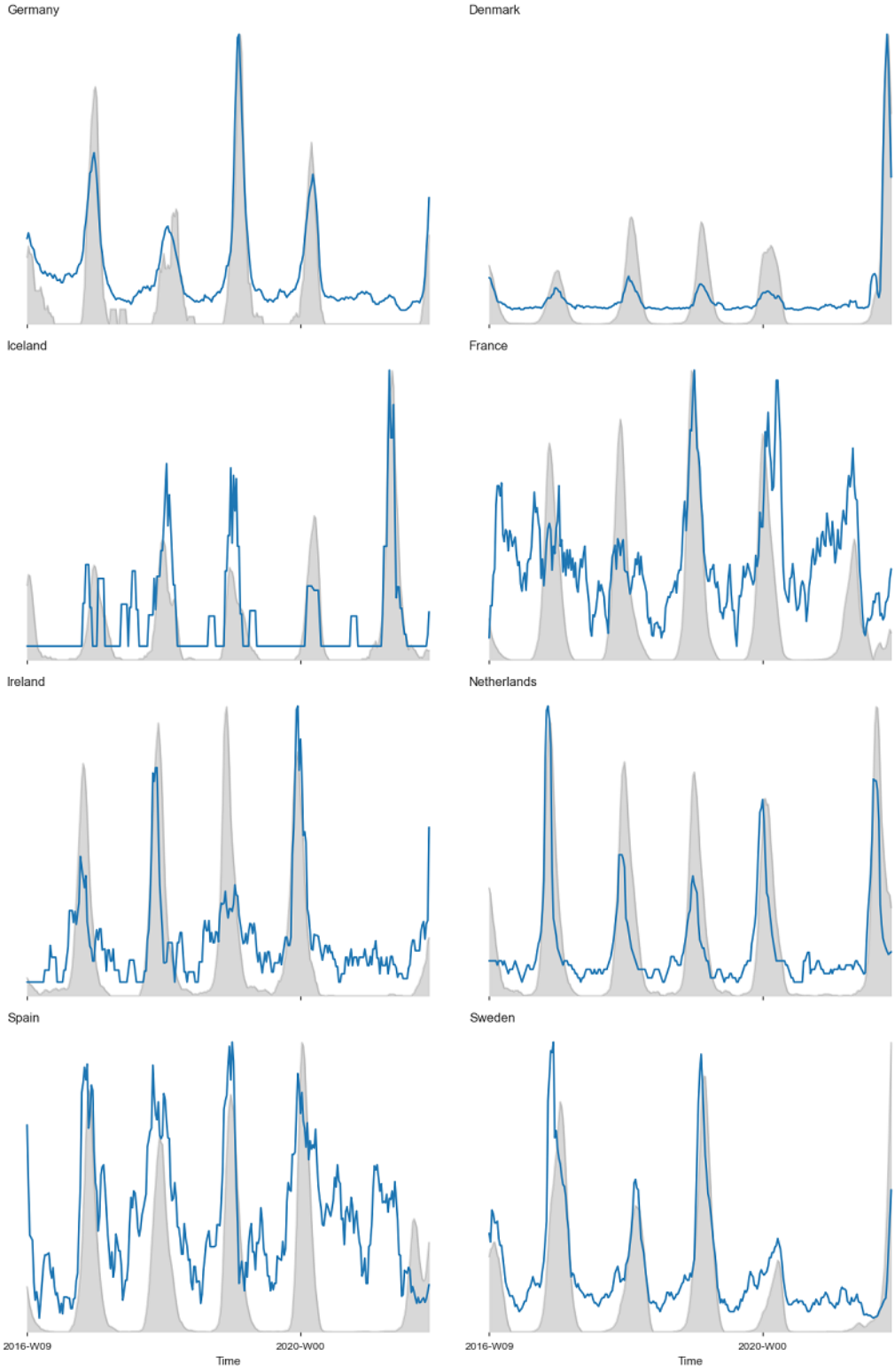
Figure 3. Changes in timing of respiratory syncytial virus incidence relative to the implementation of nonpharmaceutical interventions during the 2020-2021 season in Germany.



Validating the Correlation Between Search and Epidemiological Data

The correlation of epidemiological incidence data on RSV with internet search volume in countries for which both were available was, on average, 0.61 (n=8 countries, estimated over a period of 291 weeks; Figure 4), validating past research and suggesting that search data could indeed be used as a proxy for RSV incidence during the 2020-2021 season. This correlation is displayed graphically in Figure 3, which shows that changes in search trends closely follow the epidemiological data. The gray shaded area represents epidemiological data on RSV incidence; the blue line depicts the search data volume.

Figure 4. Correlation of epidemiological and search data in countries for which both were available.



Correlation Between Peak Latency and Magnitude

The rank regression model of peak latency as a function of country and peak magnitude for the years preceding 2021 was significant (R^2=0.13, P<.001), with country not shown as a significant explanatory variable. Applying this model to the year of the COVID-19 pandemic yielded an R^2 of 0.07 (P=.07). This shows that the correlation between RSV peak latency and magnitude broke down during the 2020-2021 RSV season.

Evaluation of Linear Regression Models

Two target variables, peak latency and peak magnitude, were examined in this study. For each, a subset of the most indicative attributes was chosen using backward stepwise selection. A linear regression model was fit to each of the target variables using each set of the chosen attributes. Both regression models were significant (adjusted R^2=0.815, $F_{19,11}$=7.967, P<.001) and

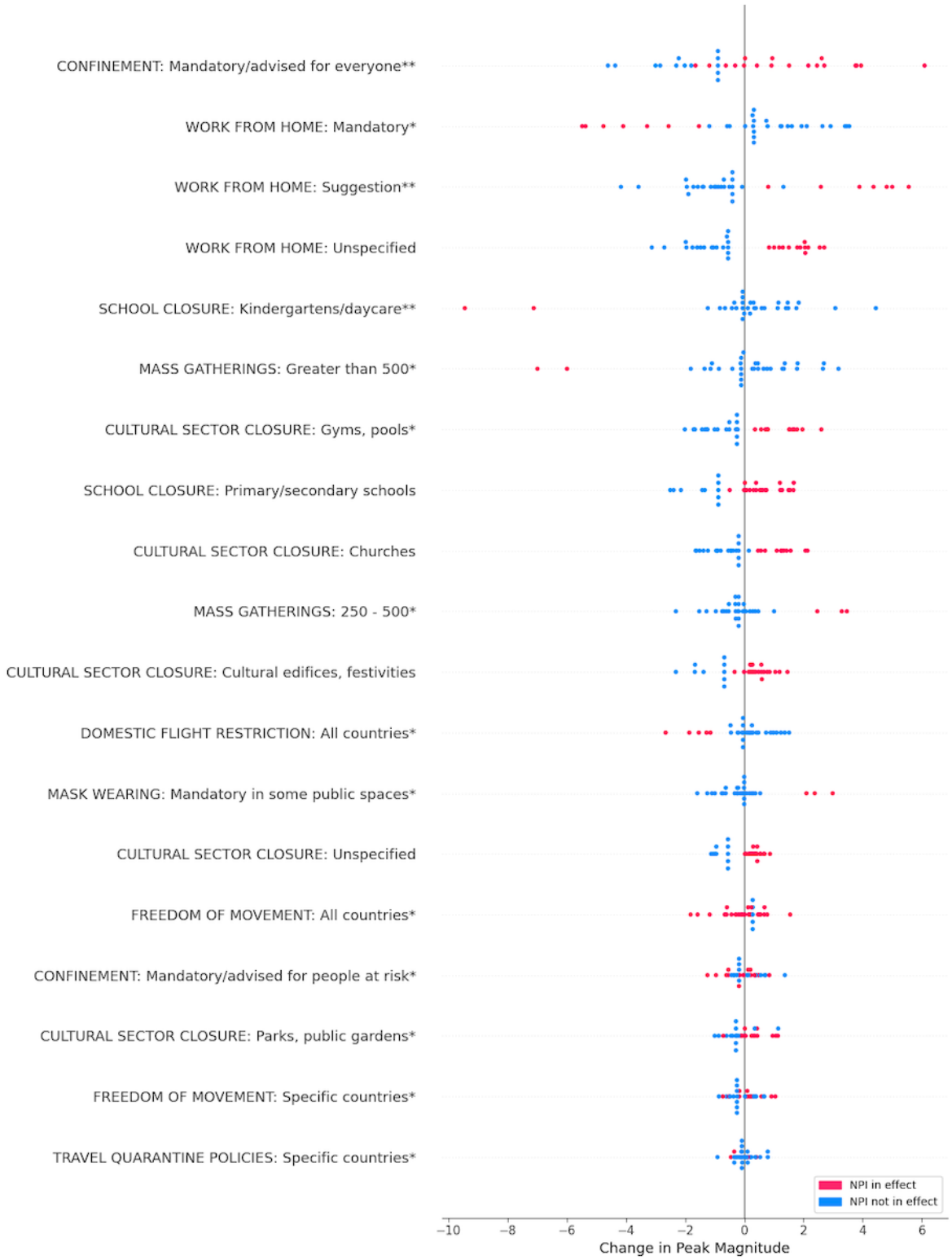
(adjusted R^2=0.799, $F_{19,11}$=7.261, P<.001) for peak latency and peak magnitude, respectively.

Insights From the Linear Regression Models

We analyzed the contribution of different NPIs to each of the two target variables. Multivariate feature importance (SHAP value) was calculated for all features in each of the two final models. Figures 5 and 6 highlight the effect of the various NPIs on peak latency and peak magnitude, respectively. Each point on the plot represents a country in the study; the color of the point indicates whether the NPI was in effect. For each feature, horizontal position relative to the midline (ie, expected value) indicates its contribution to model output. Features are arranged in descending order of mean absolute importance. It is important to note that all countries in the study experienced delays in the RSV peak. Leftward dispersion of data points in the peak latency plot, therefore, does not suggest that a feature caused an earlier peak than usual, but rather that it had a relatively lower association with the peak delay.

Figure 5. SHAP (Shapley Additive Explanations) summary plot for peak latency. Statistical significance: *P<.05, **P<.001. NPI: nonpharmaceutical intervention.

Figure 6. SHAP (Shapley Additive Explanations) summary plot for peak magnitude. Statistical significance: *P<.05, **P<.001. NPI: nonpharmaceutical intervention.



Discussion

Principal Findings

Our results support the hypothesis that NPIs instituted to combat the COVID-19 pandemic are strongly associated with changes in both the latency and magnitude of peak RSV incidence observed during the 2020-2021 RSV season. Owing to the lack of effective pharmacological treatment or vaccines for COVID-19, the global response to the pandemic relied mainly on the institution of NPIs. Various countries instituted disparate measures at different times [17]. The differential impact of these strategies has been posited to be partially responsible for the changes observed in the dynamics of respiratory viruses other than COVID-19 in 2021 [24]. Our study analyzed RSV incidence based on internet search volume and official epidemiological data reported by various agencies and used

data analysis techniques to model the effect of NPIs on RSV incidence.

Linear regression analysis of peak magnitude as a function of peak latency demonstrated a significant correlation between peak latency and magnitude for the years preceding the COVID-19 pandemic. This correlation, however, broke down during the pandemic. Furthermore, linear regression analyses modeling peak latency and peak magnitude as a function of instituted NPIs identified distinct influential features for each target feature.

Google Search Trends as a Proxy for RSV Incidence

Several studies have validated the efficacy of Google Search Trends query volume as an accurate proxy for RSV incidence [15,16]. We re-established the validity of these findings for the 2020-2021 RSV season. In keeping with previous studies, our analysis shows that changes in internet search data from most countries closely paralleled the observed epidemiological changes [15]. It is worth noting that while a recent study of global RSV seasonality included data on 27 countries, data for many of the countries had to be extracted from various national databases rather than being available from official RSV surveillance programs [2]. Although larger studies exist, these rely on data collected from published literature [25]. Surveillance data are particularly limited in middle- and low-income countries [2]. Although the need for increased epidemiological surveillance is undeniable, our study collected data on 31 countries in a manner that could potentially be automated and used for epidemiological decision-making in real time.

Correlation Between Peak Latency and Magnitude

We found that, prior to the pandemic, the yearly timing of the RSV peak was linearly correlated with its magnitude. This linear correlation was disrupted during the pandemic. Our results substantiate the hypothesis that the institution of various NPIs accounts for a high degree of variation in the timing and magnitude of the RSV outbreaks experienced by various countries.

Effects of Specific NPIs

A group of interventions related to reduced population mobility was associated with both a delayed RSV peak and a reduced peak magnitude. Among these were both domestic flight restrictions and restrictions on international arrivals from all countries, although the effects of domestic flight restrictions were more unequivocal. Restrictions on international arrivals from selected countries also showed a significant association with delayed and reduced peaks. These findings may lend credence to the hypothesis that the consistent spatiotemporal patterns of RSV spread are indeed linked to population mobility and human behavior. In the United States, for instance, yearly RSV activity begins in Florida during November and ends in February-March in the upper Midwestern United States [3]. A recent study identified the same spatiotemporal pattern, albeit shifted, in the out-of-season RSV epidemics during the 2020-2021 RSV season. The same study also suggested that increased volume of domestic air travel, which coincided with the out-of-season peaks in their study, may have been responsible [16]. Interestingly, mandatory quarantine for all arriving travelers was not significantly associated with changes in peak latency or magnitude, while mandatory quarantine on arrivals from select countries had a significant but unclear effect on peak magnitude.

Closure of school at the kindergarten or daycare level led to a significantly reduced peak magnitude. This corroborates what is known about RSV transmission in this high-risk age group. Other NPIs, particularly personal NPIs such as mask-wearing and social distancing, are extremely difficult to implement for this demographic. Reopening of schools (at all age groups) was significantly associated with an increased risk for RSV recurrence in another study that did not consider different age groups separately [26]. In our study, however, school closure at other age levels was not associated with a reduction in peak magnitude. This may be due to the greater ability to implement mask-wearing and other NPIs while maintaining school attendance at these ages. Additionally, it is possible that relaxation of school closure regulations had an unintended opposite effect of reducing social distancing in these groups.

Limiting mass gatherings to 10 people or fewer, effectively restricting interactions to the size of 1 or 2 nuclear families, was significantly associated with delayed peak. This intervention is widely considered one of the most drastic and effective measures for preventing the spread of respiratory viruses. A similar effect was demonstrated in western Washington in February 2019, where extremely high snowfall led to citywide social isolation at the level of individual household units. Researchers found that such a high-intensity intervention instituted close to the onset of an epidemic had the predominant effect of delaying the peak, while initiation of NPIs at the height of epidemic intensity predominantly decreased the peak's magnitude. In the case of RSV, which was at the height of its peak during their intervention, a 95% decrease in incidence was recorded [27]. Restricting mass gatherings to 10-100 people had the same effect; interestingly, restricting mass gatherings to 250-500 people was significantly associated with both reduced peak delays and a higher peak magnitude. This suggests that intervention at this threshold is less effective. Restriction of mass gatherings at the level of 500 people was associated with reduced peak magnitude, possibly echoing a similar conclusion to that of reduced population mobility. Restriction of mass gatherings at higher thresholds is also more difficult to interpret owing to greater variation in the types of gatherings. For instance, high-risk gatherings such as weddings, concerts, or clubs likely have different effects from lower-risk gatherings such as professional conferences or outdoor events.

Limitations

As with any study involving predictive modeling without incorporating dedicated experimental variation, associations identified here cannot be used to infer causal insights without further study. However, NPIs that were implemented to handle the COVID-19 pandemic influenced RSV incidence, as we have shown. Thus, institution of these NPIs should be considered a natural experiment from which a causal effect can be inferred. Furthermore, the high granularity of the NPIs used in this study, while vastly improving model robustness, makes it more difficult

to draw conclusive insights regarding the differential efficacy of NPIs. For instance, there are several instances in which more granular interventions, such as advised confinement for at-risk people or freedom of movement restrictions for specific countries, had a more significant effect on delaying peak incidence than their broader alternatives. While this most likely reflects the pattern adopted by many countries, of initially instituting more specific restrictions and gradually broadening them to include a greater segment of the population, it is difficult to substantiate this without conducting further research. Furthermore, some collinearity exists among the NPIs—for example, it stands to reason that countries with higher COVID-19 caseloads would implement a greater number of NPIs in parallel—thus increasing the multicollinearity of the data. We calculated the SHAP values presented using correlation-dependent feature perturbation to mitigate the effect of this collinearity to the greatest extent possible. Our study also did not consider the effects of climate, which is universally considered a significant factor contributing to RSV incidence.

Conclusions

Successfully anticipating the timing of RSV outbreaks is crucial for maximizing the prophylaxis of at-risk neonates. Current American Academy of Pediatrics guidelines recommend prophylactic treatment of at-risk neonates with monoclonal antibodies. The efficacy of this treatment is limited, however, and timely administration of the prophylactic drug before the yearly RSV outbreak is key to maximizing outcomes. Beyond the timing of prophylaxis, concerns over the timing of the 2020-2021 RSV epidemic have led governmental agencies to express concerns over concomitant viral outbreaks exceeding the capacity of health care systems [28]. This highlights the need for an efficient framework to predict changes in RSV seasonality in real time.

Identifying which interventions have the most pronounced effects on attenuating RSV outbreaks is important not only to further the understanding of RSV dynamics, but also as a tool for decision-making in future viral outbreaks. While further research is needed, we believe that our work may be a stepping stone on the path to accumulating sufficient literature and expertise to begin incorporating internet search trends as surrogate data for viral surveillance. By providing additional evidence to support the role of population mobility and human behavior on both spatial and temporal elements of RSV spread, we believe we have also shed light on the viral dynamics of RSV.

Conflicts of Interest

None declared.

References

1. American Academy of Pediatrics Committee on Infectious Diseases, American Academy of Pediatrics Bronchiolitis Guidelines Committee. Updated guidance for palivizumab prophylaxis among infants and young children at increased risk of hospitalization for respiratory syncytial virus infection. Pediatrics 2014 Aug;134(2):415-420. [doi: 10.1542/peds.2014-1665] [Medline: 25070315]
2. Obando-Pacheco P, Justicia-Grande A, Rivero-Calle I, Rodríguez-Tenreiro C, Sly P, Ramilo O, et al. Respiratory syncytial virus seasonality: a global overview. J Infect Dis 2018 Apr 11;217(9):1356-1364 [FREE Full text] [doi: 10.1093/infdis/jiy056] [Medline: 29390105]
3. Pitzer VE, Viboud C, Alonso WJ, Wilcox T, Metcalf CJ, Steiner CA, et al. Environmental drivers of the spatiotemporal dynamics of respiratory syncytial virus in the United States. PLoS Pathog 2015 Jan 8;11(1):e1004591 [FREE Full text] [doi: 10.1371/journal.ppat.1004591] [Medline: 25569275]
4. Hall CB, Douglas RG. Modes of transmission of respiratory syncytial virus. J Pediatrics 1981 Jul;99(1):100-103. [doi: 10.1016/s0022-3476(81)80969-9]
5. Qualls N, Levitt A, Kanade N, Wright-Jegede N, Dopson S, Biggerstaff M, CDC Community Mitigation Guidelines Work Group. Community mitigation guidelines to prevent pandemic influenza - United States, 2017. MMWR Recomm Rep 2017 Apr 21;66(1):1-34 [FREE Full text] [doi: 10.15585/mmwr.rr6601a1] [Medline: 28426646]
6. Teirlinck AC, Broberg EK, Stuwitz Berg A, Campbell H, Reeves RM, Carnahan A, et al. Recommendations for respiratory syncytial virus surveillance at the national level. Eur Respir J 2021 Sep 22;58(3):2003766 [FREE Full text] [doi: 10.1183/13993003.03766-2020] [Medline: 33888523]
7. Olsen SJ, Winn AK, Budd AP, Prill MM, Steel J, Midgley CM, et al. Changes in influenza and other respiratory virus activity during the COVID-19 pandemic - United States, 2020-2021. MMWR Morb Mortal Wkly Rep 2021 Jul 23;70(29):1013-1019 [FREE Full text] [doi: 10.15585/mmwr.mm7029a1] [Medline: 34292924]
8. Polgreen P, Chen Y, Pennock D, Nelson F. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 01;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
9. Sousa-Pinto B, Halonen JI, Antó A, Jormanainen V, Czarlewski W, Bedbrook A, et al. Prediction of asthma hospitalizations for the common cold using Google Trends: infodemiology study. J Med Internet Res 2021 Jul 06;23(7):e27044 [FREE Full text] [doi: 10.2196/27044] [Medline: 34255692]
10. Samaras L, García-Barriocanal E, Sicilia M. Syndromic surveillance models using web data: the case of influenza in Greece and Italy using Google Trends. JMIR Public Health Surveill 2017 Nov 20;3(4):e90 [FREE Full text] [doi: 10.2196/publichealth.8015] [Medline: 29158208]

11. Husnayain A, Shim E, Fuad A, Su E. 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 Dec 22;23(12):e34178 [FREE Full text] [doi: 10.2196/34178] [Medline: 34762064]
12. Lazer D, Kennedy R, King G, Vespignani A. Google flu trends still appears sick: an evaluation of the 2013-2014 flu season. SSRN Journal 2014. [doi: 10.2139/ssrn.2408560]
13. Lu FS, Hattab MW, Clemente CL, Biggerstaff M, Santillana M. Improved state-level influenza nowcasting in the United States leveraging internet-based data and network approaches. Nat Commun 2019 Jan 11;10(1):147 [FREE Full text] [doi: 10.1038/s41467-018-08082-0] [Medline: 30635558]
14. Osthus D, Daughton AR, Priedhorsky R. Even a good influenza forecasting model can benefit from internet-based nowcasts, but those benefits are limited. PLoS Comput Biol 2019 Feb 1;15(2):e1006599 [FREE Full text] [doi: 10.1371/journal.pcbi.1006599] [Medline: 30707689]
15. Oren E, Frere J, Yom-Tov E, Yom-Tov E. Respiratory syncytial virus tracking using internet search engine data. BMC Public Health 2018 Apr 03;18(1):445 [FREE Full text] [doi: 10.1186/s12889-018-5367-z] [Medline: 29615018]
16. Zheng Z, Warren JL, Artin I, Pitzer VE, Weinberger DM. Relative timing of respiratory syncytial virus epidemics in summer 2021 across the United States was similar to a typical winter season. Influenza Other Respir Viruses 2022 Jul 08;16(4):617-620 [FREE Full text] [doi: 10.1111/irv.12965] [Medline: 35137538]
17. Suryanarayanan P, Tsou C, Poddar A, Mahajan D, Dandala B, Madan P, et al. AI-assisted tracking of worldwide non-pharmaceutical interventions for COVID-19. Sci Data 2021 Mar 25;8(1):94 [FREE Full text] [doi: 10.1038/s41597-021-00878-y] [Medline: 33767205]
18. Agha R, Avner J. Delayed seasonal RSV surge observed during the COVID-19 pandemic. Pediatrics 2021 Sep;148(3). [doi: 10.1542/peds.2021-052089] [Medline: 34108234]
19. Weinberger Opek M, Yeshayahu Y, Glatman-Freedman A, Kaufman Z, Sorek N, Brosh-Nissimov T. Delayed respiratory syncytial virus epidemic in children after relaxation of COVID-19 physical distancing measures, Ashdod, Israel, 2021. Euro Surveill 2021 Jul;26(29) [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.29.2100706] [Medline: 34296678]
20. Casalegno J, Ploin D, Cantais A, Masson E, Bard E, Valette M, VRS study group in Lyon, et al. Characteristics of the delayed respiratory syncytial virus epidemic, 2020/2021, Rhône Loire, France. Euro Surveill 2021 Jul;26(29) [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.29.2100630] [Medline: 34296674]
21. Surveillance atlas of infectious diseases. URL: https://atlas.ecdc.europa.eu/public/index.aspx [accessed 2022-12-08]
22. West R. Calibration of Google Trends time series. 2020 Presented at: CIKM '20: The 29th ACM International Conference on Information and Knowledge Management; October 19-23, 2020; Virtual, Ireland. [doi: 10.1145/3340531.3412075]
23. Lundberg S, Lee S. A unified approach to interpreting model predictions. 2017 Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017; Long Beach, CA.
24. Baker RE, Park SW, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. The impact of COVID-19 nonpharmaceutical interventions on the future dynamics of endemic infections. Proc Natl Acad Sci U S A 2020 Dec 01;117(48):30547-30553 [FREE Full text] [doi: 10.1073/pnas.2013182117] [Medline: 33168723]
25. Bloom-Feshbach K, Alonso WJ, Charu V, Tamerius J, Simonsen L, Miller MA, et al. Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. PLoS One 2013 Feb 14;8(2):e54445 [FREE Full text] [doi: 10.1371/journal.pone.0054445] [Medline: 23457451]
26. Li Y, Wang X, Cong B, Deng S, Feikin D, Nair H. Understanding the potential drivers for respiratory syncytial virus rebound during the coronavirus disease 2019 pandemic. J Infect Dis 2022 Mar 15;225(6):957-964 [FREE Full text] [doi: 10.1093/infdis/jiab606] [Medline: 35030633]
27. Jackson ML, Hart GR, McCulloch DJ, Adler A, Brandstetter E, Fay K, Seattle Flu Study Investigators. Effects of weather-related social distancing on city-scale transmission of respiratory viruses: a retrospective cohort study. BMC Infect Dis 2021 Apr 09;21(1):335 [FREE Full text] [doi: 10.1186/s12879-021-06028-4] [Medline: 33836685]
28. Mahase E. Winter pressure: RSV, flu, and covid-19 could push NHS to breaking point, report warns. BMJ 2021 Jul 14;374:n1802. [doi: 10.1136/bmj.n1802] [Medline: 34266879]

Abbreviations

NPI: nonpharmaceutical intervention
RSV: respiratory syncytial virus
SHAP: Shapley Additive Explanations
WNTRAC: Worldwide Non-pharmaceutical Interventions Tracker for COVID-19

Edited by M Focsa; submitted 18.09.22; peer-reviewed by A Carnahan, J Paget; comments to author 10.11.22; revised version received 28.11.22; accepted 04.12.22; published 21.12.22.

Please cite as:
Ravkin HD, Yom-Tov E, Nesher L
The Effect of Nonpharmaceutical Interventions Implemented in Response to the COVID-19 Pandemic on Seasonal Respiratory Syncytial Virus: Analysis of Google Trends Data
J Med Internet Res 2022;24(12):e42781
URL: https://www.jmir.org/2022/12/e42781
doi:10.2196/42781
PMID:36476385

Original Paper

How COVID-19 Affected the Journal Impact Factor of High Impact Medical Journals: Bibliometric Analysis

Orestis Delardas[1*], BEng, MSc; Panagiotis Giannos[1,2*], BSc, MSc

[1]Promotion of Emerging and Evaluative Research Society, London, United Kingdom

[2]Department of Life Sciences, Faculty of Natural Sciences, Imperial College London, London, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Panagiotis Giannos, BSc, MSc
Department of Life Sciences
Faculty of Natural Sciences
Imperial College London
South Kensignton
London, SW7 2AZ
United Kingdom
Phone: 44 7765071907
Email: panagiotis.giannos19@imperial.ac.uk

Related Article:

Comment in: https://www.jmir.org/2023/1/e45322

Abstract

Background: Journal impact factor (IF) is the leading method of scholarly assessment in today's research world, influencing where scholars submit their research and where funders distribute their resources. COVID-19, one of the most serious health crises, resulted in an unprecedented surge of publications across all areas of knowledge. An important question is whether COVID-19 affected the *gold standard of scholarly assessment*.

Objective: In this paper, we aimed to comprehensively compare the productivity trends of COVID-19 and non–COVID-19 literature as well as track their evolution and scholarly impact across 3 consecutive calendar years.

Methods: We took as an example 6 high-impact medical journals (Annals of Internal Medicine [Annals], The British Medical Journal [The BMJ], Journal of the American Medical Association [JAMA], The Lancet, Nature Medicine [NatMed], and The New England Journal of Medicine [NEJM]) and searched the literature using the Web of Science database for manuscripts published between January 1, 2019, and December 31, 2021. To assess the effect of COVID-19 and non–COVID-19 literature in their scholarly impact, we calculated their annual IFs and percentage changes. Thereafter, we estimated the citation probability of COVID-19 and non–COVID-19 publications along with their rates of publication and citation by journal.

Results: A significant increase in IF change for manuscripts including COVID-19 published from 2019 to 2020 (P=.002; Annals: 283%; The BMJ: 199%; JAMA: 208%; The Lancet: 392%; NatMed: 111%; and NEJM: 196%) and to 2021 (P=.007; Annals: 41%; The BMJ: 90%; JAMA: 6%; The Lancet: 22%; NatMed: 53%; and NEJM: 72%) was seen, against non–COVID-19 ones. The likelihood of highly cited publications was significantly increased in COVID-19 manuscripts between 2019 and 2021 (Annals: z=3.4, P<.001; The BMJ: z=4.0, P<.001; JAMA: z=3.8, P<.001; The Lancet: z=3.5, P<.001; NatMed: z=5.2, P<.001; and NEJM: z=4.7, P<.001). The publication and citation rates of COVID-19 publications followed a positive trajectory, as opposed to non–COVID-19. The citation rate for COVID-19 publications peaked by the second quarter of 2020 while that of the publication rate approximately a year later.

Conclusions: The rapid surge of COVID-19 publications emphasized the capacity of scientific communities to respond against a global health emergency, yet inflated IFs create ambiguity as benchmark tools for assessing scholarly impact. The immediate implication is a loss in value of and trust in journal IFs as metrics of research and scientific rigor perceived by academia and society. Loss of confidence toward procedures employed by highly reputable publishers may incentivize authors to exploit the

publication process by monopolizing their research on COVID-19 and encourage them to publish in journals of *predatory* behavior.

(J Med Internet Res 2022;24(12):e43089) doi:10.2196/43089

KEYWORDS

COVID-19; journal impact factor; scientometrics; bibliometrics; infometrics; journal; assessment; research; resources; medical journal; literature; database; community; behavior

Introduction

The COVID-19 pandemic has presented unique challenges to modern societies. The spread of SARS-CoV-2 worldwide during the early 2020 posed unprecedented disruption in human lives and significant stresses to public health structures and socioeconomic systems [1]. *Feverish* academic activity on COVID-19 research has been recorded across all areas of knowledge with almost immediate effects since the disease was discovered in December 2019. Surges in COVID-19 infections in China and Italy initiated the first wave of COVID-19 publications within the first 3 months of the pandemic [2]. The die had been cast.

From an early stage, it was evident that the rate of publications was the most of any disease published thus far; however, very few constituted high-level original research including meta-analyses, systematic reviews, or control trials while preprints, opinion pieces, editorials or commentaries filled the void [2,3]. Adding to the exponential growth in COVID-19 publications was also the expedited editorial peer-review processes put in place for manuscripts that generated record speeds in processing times and article acceptance [4]. Indeed, Palayew et al [5] revealed that median time from submission to acceptance of COVID-19 research was reduced to 6 days from 84 days when compared to non–COVID-19 content during the first months of 2020. Horbach [6] further showed that the majority of this decrease was attributed to an acceleration of the review process.

These steps of disproportionately shorter-than-usual processes taken by academic journals reflect the particular urgency for information-sharing and altmetric dissemination across scientific fields [7]. However, fast-track review methods and processes developed during the COVID-19 pandemic are “here to stay” [8]. Although reasonable to expect such policies during times of extraordinary mobilization for novel treatments and best practices to combat a disease of this scale, not all outcomes are rosy. Much as the COVID-19 infodemic might have been paved with good intentions, efforts to loosen the demanding process of peer reviewing can seriously undermine research quality and the potential merit of journals in their attempt.

Critiques of such approaches have been described in the literature. Palayew et al [5] made the case for more training of peer reviewers before they are allowed to review in such short time frames to avoid weakening of scientific evidence. El-Menyar et al [4] argued about the necessity to uphold research ethics and best practices in fast-track processes and highlighted the scarcity of original research data, which often led to resources being reused, carrying forward flaws and inaccuracies that shaped public opinion and policies [4,9]. Glasziou [10] pointed out that the COVID-19 research corpus consists largely of preprints and duplicate studies, and Bero et al [11] drew attention to the decreased trustworthiness and validity of less rigorously reviewed COVID-19 research.

A discussion on the real repercussions of an overwhelming focus on COVID-19 research for scientometrics, such as that of the journal impact factor (IF), lacks in the existing literature. Given the role of these measures as benchmarks of research productivity and scholarly impact, editorial practices in favor of fast-track COVID-19 research output have fueled critiques, which view these as attempts to artificially inflate metrics at the potential expense of research quality.

Journal IF constitutes the principal approach to assess scholarly impact in modern research. This appraisal often guides scholars to select where to submit their research and funding bodies to decide where to allocate their resources. Considering the surge of COVID-19 research from the start of the pandemic, a crucial question arises on its influence upon the *gold standard of scholarly assessment* in journals of highest rank. We focused on 6 exemplar high-impact medical journals (Annals of Internal Medicine [Annals], The British Medical Journal [The BMJ], Journal of the American Medical Association [JAMA], The Lancet, Nature Medicine [NatMed], and The New England Journal of Medicine [NEJM]). The aim of our study was to comprehensively compare the productivity trends of COVID-19 and non–COVID-19 literature and track their evolution and scholarly impact across 3 consecutive calendar years.

Methods

Data Collection

To fulfil the purpose of our study, we selected 6 high-impact medical journals, namely Annals, The BMJ, JAMA, The Lancet, NatMed, and NEJM. We conducted a comprehensive search of the literature using the Web of Science database for manuscripts published between January 1, 2019, and December 31, 2021. To distinguish between COVID-19 and non–COVID-19 publications, we filtered manuscripts based on their title, abstract, or keywords using the following terms: “COVID-19” OR “SARS-COV2” OR “Coronavirus” OR “2019-nCoV.” Citation counts for each manuscript were retrieved using the Clarivate report function. The search of the literature was performed on a single day to reduce daily updates of the database. Manuscripts were restricted to peer-reviewed original research and review articles. No further exclusion criteria were applied to our search.

Data Processing

Calculation of a journal's IF in our study was based on the ratio between the number of citations and manuscripts published in a given journal over a single year for that journal.

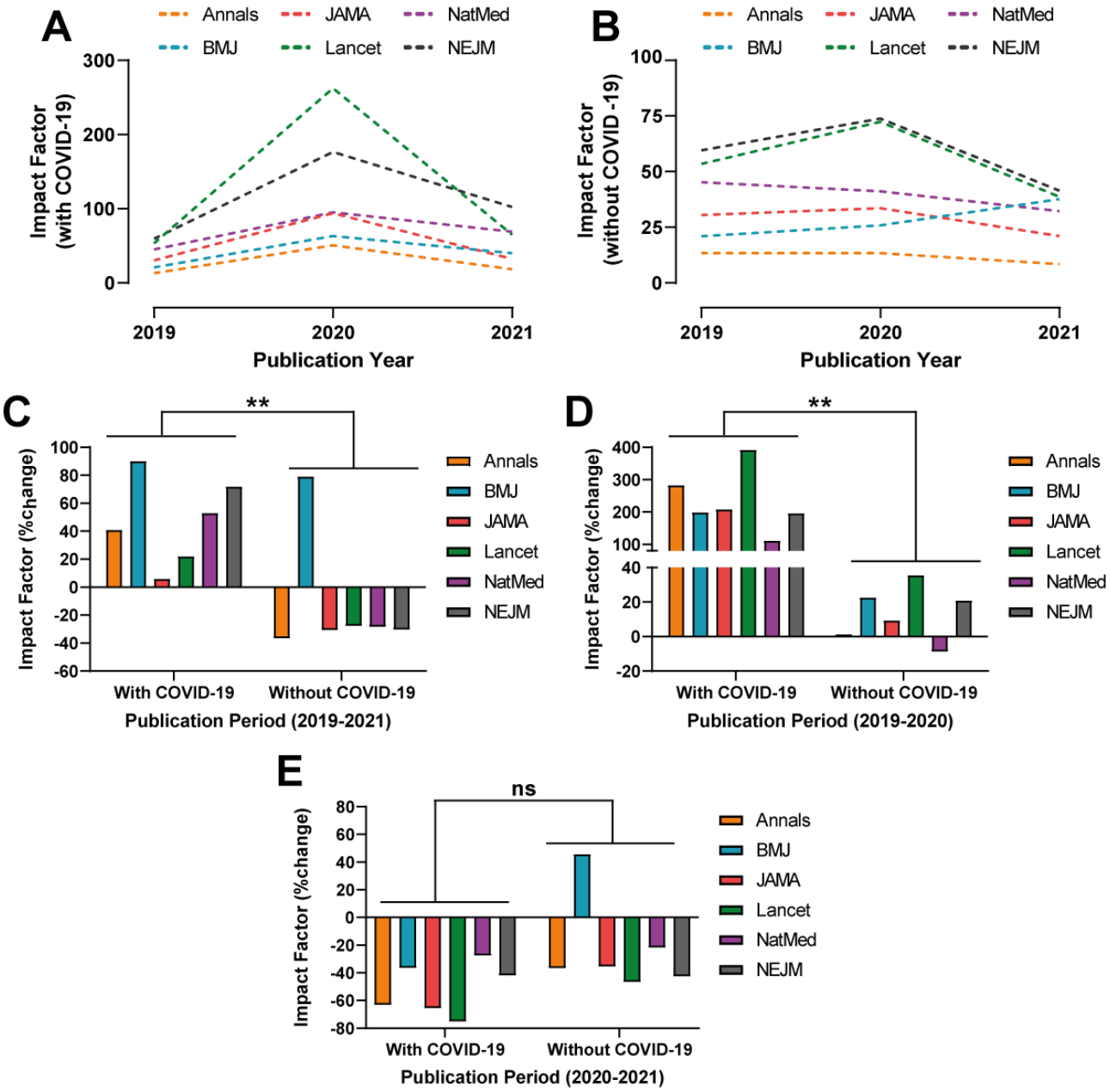
This approach was employed to enhance the temporal resolution of the analysis of scholarly influence from journals in publishing. To assess the effect of COVID-19 and non–COVID-19 literature on scholarly impact of these journals, we initially tracked the evolution of their IFs yearly from 2019 to 2021. We then calculated the percentage change in IFs year on year. Thereafter, we estimated the citation probability of any given COVID-19 and non–COVID-19 publication by journal amid the whole duration. These were expressed as normal distributions and calculated using the normal distribution function (NORMDIST) in Microsoft Excel 2016. On a more granular level, we estimated the publication and citation rate of COVID-19 and non–COVID-19 manuscripts on a monthly basis. Statistical significance was established as $P<.05$, differences in means were examined using a paired sample *t* test (two-tailed), and differences in distribution curves were assessed using an independent *z* test (two-tailed). Statistical analysis was performed using Microsoft Excel 2016 and SPSS statistics software, Version 28.0 (IBM Corp).

Results

Journal Impact Factors

The IFs of all 6 high-impact medical journals significantly increased for manuscripts including COVID-19 published from 2019 to 2020 ($P=.002$; Annals: 283%; The BMJ: 199%; JAMA: 208%; The Lancet: 392%; NatMed: 111%; and NEJM: 196%) and to 2021 ($P=.007$; Annals: 41%; The BMJ: 90%; JAMA: 6%, The Lancet: 22%; NatMed: 53%; and NEJM: 72%), when accounting for non–COVID-19 ones (Figure 1 and Table S1-2 in Multimedia Appendix 1). During the former period, The Lancet and Annals experienced the highest increase with a change in IF of 392% and 283% (as opposed to 36% and 1%), respectively. An exception to this trend was NatMed, which saw a decrease in IF of 9% (as opposed to 111%). During the latter period, a more moderate increase was observed across all journals and most prominently of 90% and 72% (as opposed to 79% and -31%) in The BMJ and NEJM, respectively. Notably, The BMJ was the only to experience sustained increase in IF from 2019 to 2020 and 2021. No significant changes were observed from 2020 to 2021 ($P=.06$).

Figure 1. Annual impact factor of 6 high-impact medical journals (Annals of Internal Medicine [Annals], The British Medical Journal [BMJ], Journal of the American Medical Association [JAMA], The Lancet, Nature Medicine [NatMed], and The New England Journal of Medicine [NEJM]) based on (A) manuscripts with and (B) without COVID-19 publications between 2019 and 2021. Changes in annual impact factor comparing manuscripts (C-E) with and without COVID-19 publications between 2019 and 2021. ns: not significant; **P<.01.



Probability of Citations

The probability of highly cited manuscripts published between 2019 and 2021 across all journals was significantly increased for COVID-19 manuscripts compared to non–COVID-19 ones (Annals: z=3.4, P<.001; The BMJ: z=4.0, P<.001; JAMA: z=3.8, P<.001; The Lancet: z=3.5, P<.001; NatMed: z=5.2, P<.001; and NEJM: z=4.7, P<.001; Figure 2 and Table S3 in Multimedia Appendix 1). The highest citation probability was seen in manuscripts published in NatMed (z=5.2) during this period. The likelihood of highly cited manuscripts was visually increased across all journals except in that of The BMJ, when considering manuscripts with COVID-19 against those without (Figure 2). Equally, the probability of highly cited COVID-19 manuscripts published during 2019-2021 appeared highest in The Lancet and NEJM compared to the majority of the remaining journals (Figure 3A and Table S4 in Multimedia Appendix 1).

Figure 2. Probability distributions of time-adjusted citation count for 6 high-impact medical journals—(A) Annals of Internal Medicine, (B) The British Medical Journal, (C) Journal of the American Medical Association, (D) The Lancet, (E) Nature Medicine, and (F) The New England Journal of Medicine—based on non–COVID-19, COVID-19–only, and combined publications between 2019 and 2021. ***P<.001.

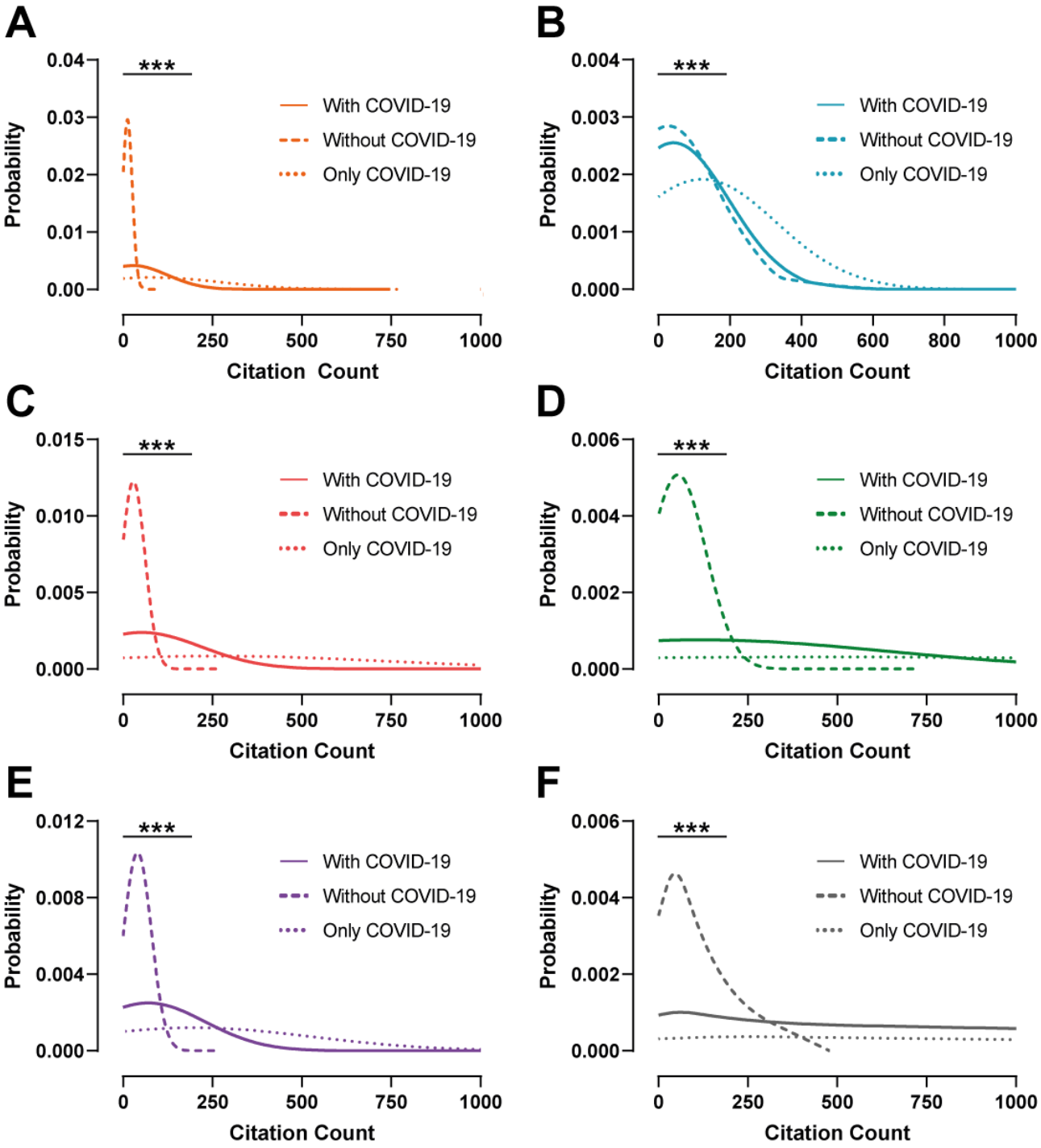
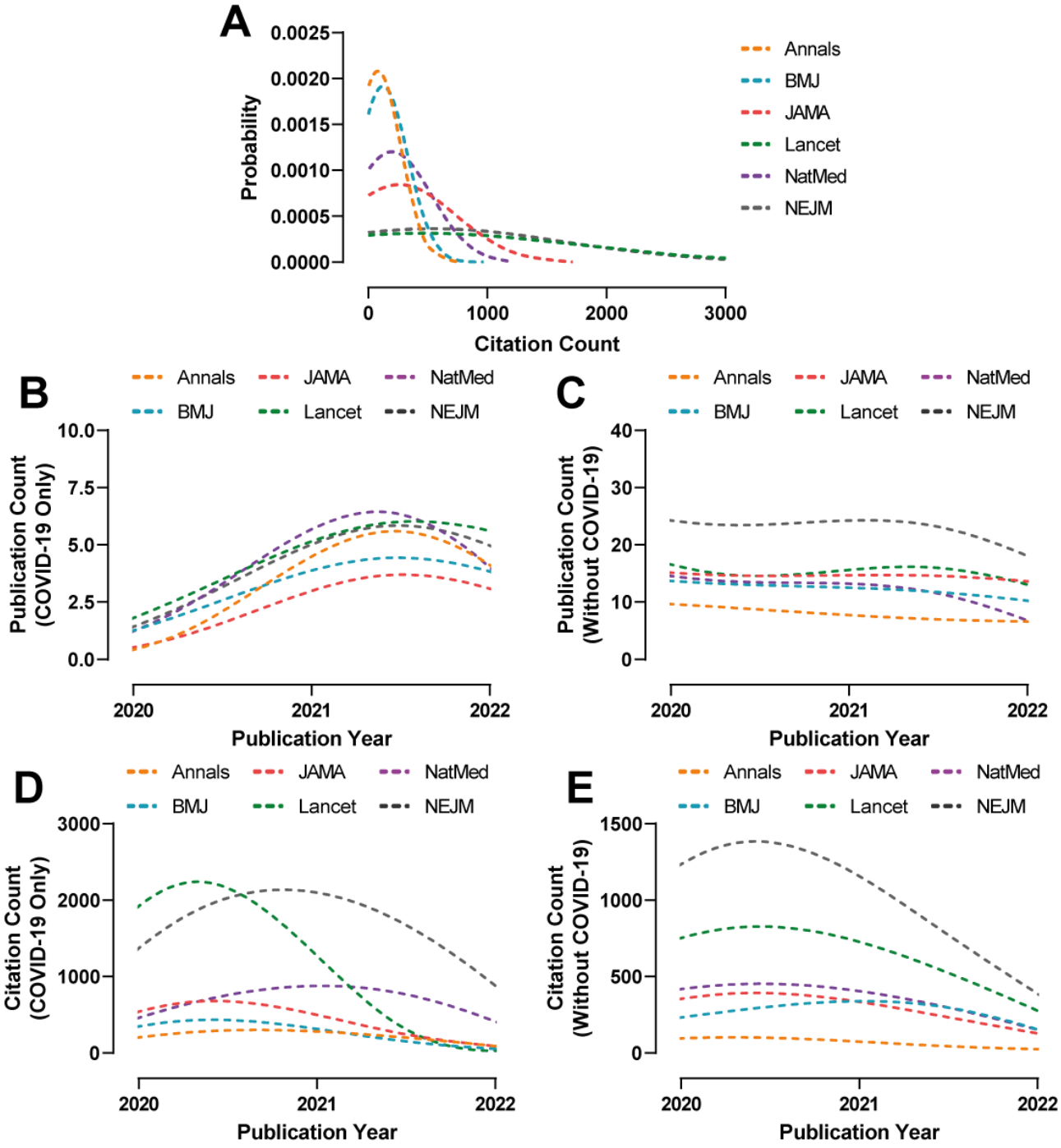


Figure 3. Probability distributions of (A) time-adjusted citation count across 6 high-impact medical journals (Annals of Internal Medicine [Annals], The British Medical Journal [BMJ], Journal of the American Medical Association [JAMA], The Lancet, Nature Medicine [NatMed], and The New England Journal of Medicine [NEJM]) based on COVID-19–only publications between 2019 and 2021. Publication and citation rates based on (B and D) COVID-19–only and (C and E) non–COVID-19 publications between 2020 and 2021.



Rate of Publications and Citations

The publication rate of COVID-19 manuscripts across all journals saw an increase between 2020 and 2021 with a peak by the second quarter of 2021 (Figure 3B). By contrast, the publication rate of non–COVID-19 manuscripts saw a moderate decrease throughout the elapsed duration (Figure 3C). Moreover, the citation rate of COVID-19 manuscripts peaked in the first 2 quarters of 2020 and strongly subsided afterward (Figure 3D). Conversely, non–COVID-19 manuscripts saw a continuous and extensive downward decrease in their citation rate from the start (Figure 3E).

Discussion

Principal Findings

Our study showed a significant increase in IF change across 6 high-impact medical journals (Annals, The BMJ, JAMA, The Lancet, NatMed, and NEJM) based on publications including COVID-19 manuscripts from 2019 to 2020 and to 2021, when compared to non–COVID-19 ones. The probability of highly

cited manuscripts was significantly increased in COVID-19 manuscripts across most journals and throughout the entire duration, when compared to non–COVID-19 ones. The citation rate for COVID-19 publications peaked by the second quarter of 2020 and that of the publication rate approximately a year later.

Interpretation of Findings

Our results reflect the capacity of the scientific community to respond against a global health emergency with high-impact publications on COVID-19 at an exponentially expanding rate. With high hopes for a breakthrough, scientists have indeed rushed to publish positive results on the disease [12]. However, this raises concerns whether scientific standards are being met both by researchers and the journals [13]. High-impact medical journals, including The Lancet, Nature, and JAMA embarked on a rapid peer-review initiative to accelerate the dissemination of COVID-19 manuscripts to the public and across the scientific community [14-16]. Nature explicitly invited researchers to shorten review times and decided to reduce the publication of non–COVID-19 content. JAMA expedited the publication of COVID-19 manuscripts within 10 to 12 days from submission [14,15]. A later analysis further confirmed that among other journals, The Lancet and Nature shortened their review processes for COVID-19 articles by almost two-thirds for the duration of the pandemic, when compared to non–COVID-19 submissions [17]. Another report showed that when the quality of peer-reviewed COVID-19 publications was assessed in the 3 most influential medical journals (ie, The Lancet, NEJM, and JAMA), high rates of retraction, withdrawal, or expression of concern were observed [18,19].

Fast-track publications practices have frequently been scrutinized for the rigidity of the research output. This underlies concerns about the quality control of the external peer review and internal editorial evaluation, thorough revision by authors, and journal editing of the manuscript. Most notably, The Lancet and NEJM came under intense fire during the second half of 2020 due to the publishing of false data in a highly influential study regarding the benefit of hydroxychloroquine or chloroquine for the treatment of COVID-19. This information found its way under the public spotlight causing a controversial deluge, leading to its retraction and a barrage of criticism at the integrity and quality of the research and its peer reviewing [20]. Notably, The Lancet has now reflected on the risks of rushed review processes employed as part of their early action against the pandemic and reiterated the need to “slow down” in their publication processes [16]. Nevertheless, a scarcity of explicit information regarding other high-impact medical journals, including The BMJ, NEJM, and Annals, remains. It would be no surprise that similar recorded patterns for COVID-19 publications in these journals could have been attributed to the expedited reviewing processes in an attempt to ease submission bottlenecks.

The growing concern that editorial practices can be as much responsible for the influx in publications as the heightened popularity of the topic among the academic community becomes evident. The attributed responsibility on editorial processes is mainly based on the asymmetrical treatment of COVID-19 research and the consequential encouragement of scientists to focus on COVID-19 by journals. These two acts invite certain types of research by making the route to publication more certain and less time intensive.

There are bearing implications to the potential inflation of journal metrics of research productivity and scholarly impact, such as IF. Journal IFs are commonly used by educational or research groups and various funding bodies to make decisions on the promotion of research proposals, grant applications, but also the awarding of positions to individuals and even salary considerations. In a sense, they provide a way to gauge a scientist or research group’s academic value, or an academic journal’s scientific rigor. Journal IFs, parallel to money, constitute a value system of scholarly influence. To maintain their value across time, they need to rely on stable and transparent processes that remain intact and are always faithfully followed. For academic journals, the main mechanism that controls publication rates and incentivizes research quality is a well-established and thorough peer-reviewing process. Similar to how currency manipulation works, when peer reviewing is altered, there is a risk of distorting the value of and trust in journal IFs as perceived by academia and society.

Apart from the obvious loss of confidence toward the procedures employed by highly reputable publishers, academic journals also face the risk of losing the interest of researchers in publishing to other competitors, and this might be damaging from a business perspective. COVID-19 articles can be a contributing factor to this phenomenon which is exacerbated when fast-track reviewing is made a priority. The lack of transparency and information on how and where exactly fast-track reviewing was implemented during the pandemic magnifies this issue as there is no real way for external parties to assess how much of this is artificially driven. This sows confusion among scholars on how to evaluate the quality of published research and may encourage authors to publish in journals of *predatory* behavior.

Another disservice that journal IF inflation and sudden changes in standards might cause is putting honest and hard-working researchers at a disadvantage. Shortened and sometimes less rigorous peer-review processes, combined with the observed surge in preprints, opinion pieces, and commentaries, while by no means unimportant, increase scientific noise and can waste resources that could be used in a lengthier but more impactful research. The rearrangement of peer reviewing might also benefit authors who are willing to exploit the system in order to inflate their productivity metrics and get an edge over colleagues who are less inclined to take advantage of the *hype*. This can reinforce a deluge of COVID-19 submissions of worrying quality as increasingly more researchers *get the trick* and do not want to miss out on the effortless opportunity to transform their career.

Limitations

Our study was prone to various inherent limitations. Assessment of IF by year can provide an enhanced temporal resolution of the scholarly influence presented by journals from their research output. However, overtime citations become inflated, and calculating year-specific IFs becomes challenging for a

retrospective analysis. To overcome this, we applied a time adjustment on the citations count based on the time elapsed from the start of the search up to date. However, we were not able to account for any traction cycles or short-term events that articles might have experienced over time. Although the IF of all included journals in our study was affected symmetrically by this inherent pitfall, it is likely that the derived yearly IFs were underestimated, especially in articles published at later years. Nevertheless, this phenomenon portrays the crudeness and imperfect abstraction of IF in gaining a more granular investigation. Similarly, COVID-19 manuscripts were restricted to article and review types without taking into consideration related editorials, opinions, or commentaries that constituted a significant portion of the surge in COVID-19 publications from the start. In the same manner, the protocol of data acquisition employed to collect manuscript count was limited to a single database (ie, Web of Science), which could have consequently magnified the quantity of eligible publications. Lastly, a manual screening of the derived publications was not possible, which led to an automatic filtering based on title, abstract, or keywords that best describe COVID-19 terminology. Hence, we could not establish in full whether the retrieved manuscripts indeed focused on COVID-19 and not on other research domains related to it. Taken together, our results and the conclusions derived may be considered more conservative and should be interpreted with caution.

Conclusions

The rise of COVID-19 has resulted in a surge of scientific production across all areas of knowledge globally. Our findings ultimately demonstrated that the IF, likelihood of being highly cited, and publication and citation rates of manuscripts published across 6 high impact medical journals (Annals, The BMJ, JAMA, The Lancet, NatMed, and NEJM), between 2019 and 2021, were positively skewed by COVID-19 manuscripts. The eruption of COVID-19 publications reinforced the capacity of the scientific community to step up to the challenge, but casted doubt on the reliability of highly susceptible IFs—as shown here—in evaluating scholarly impact. The loss of trust on journal IFs as measures of scientific rigor and confidence in the procedures employed by highly influential publishers may incentivize a culture of exploitation by researchers and journals against the scientific process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 28 KB - jmir_v24i12e43089_app1.docx]

References

1. Delardas O, Kechagias KS, Pontikos PN, Giannos P. Socio-Economic Impacts and Challenges of the Coronavirus Pandemic (COVID-19): An Updated Review. Sustainability 2022 Aug 06;14(15):9699 [FREE Full text] [doi: 10.3390/su14159699] [Medline: 26701262]
2. Kambhampati SBS, Vaishya R, Vaish A. Unprecedented surge in publications related to COVID-19 in the first three months of pandemic: A bibliometric analytic report. J Clin Orthop Trauma 2020 May;11(Suppl 3):S304-S306 [FREE Full text] [doi: 10.1016/j.jcot.2020.04.030] [Medline: 32405191]
3. Gianola S, Jesus TS, Bargeri S, Castellini G. Characteristics of academic publications, preprints, and registered clinical trials on the COVID-19 pandemic. PLoS One 2020;15(10):e0240123 [FREE Full text] [doi: 10.1371/journal.pone.0240123] [Medline: 33022014]
4. El-Menyar A, Mekkodathil A, Asim M, Consunji R, Rizoli S, Abdel-Aziz Bahey A, et al. Publications and retracted articles of COVID-19 pharmacotherapy-related research: A systematic review. Sci Prog 2021;104(2):368504211016936 [FREE Full text] [doi: 10.1177/00368504211016936] [Medline: 33989091]
5. Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus JV. Pandemic publishing poses a new COVID-19 challenge. Nat Hum Behav 2020 Jul;4(7):666-669. [doi: 10.1038/s41562-020-0911-0] [Medline: 32576981]
6. Horbach SPJM. Pandemic publishing: Medical journals strongly speed up their publication process for COVID-19. Quantitative Science Studies 2020 Aug;1(3):1056-1067. [doi: 10.1162/qss_a_00076]
7. Giannos P, Kechagias KS, Katsikas Triantafyllidis K, Falagas ME. Spotlight on early COVID-19 research productivity: A 1-year bibliometric analysis. Front Public Health 2022;10:811885 [FREE Full text] [doi: 10.3389/fpubh.2022.811885] [Medline: 35712290]
8. Coombes R. Fast track research approval "here to stay," says regulator. BMJ 2021 Mar 12;372:n705. [doi: 10.1136/bmj.n705] [Medline: 33712495]
9. Chirico F, Teixeira da Silva JA, Magnavita N. "Questionable" peer review in the publishing pandemic during the time of COVID-19: implications for policy makers and stakeholders. Croat Med J 2020 Jul 05;61(3):300-301 [FREE Full text] [doi: 10.3325/cmj.2020.61.300] [Medline: 32643350]

10. Bero L, Lawrence R, Leslie L, Chiu K, McDonald S, Page MJ, et al. Cross-sectional study of preprints and final journal publications from COVID-19 studies: discrepancies in results reporting and spin in interpretation. BMJ Open 2021 Jul 16;11(7):e051821 [FREE Full text] [doi: 10.1136/bmjopen-2021-051821] [Medline: 34272226]
11. Glasziou P, Sanders S, Hoffmann T. Waste in covid-19 research. BMJ 2020 May 12;369:m1847. [doi: 10.1136/bmj.m1847] [Medline: 32398241]
12. Hossain MM. Current status of global research on novel coronavirus disease (COVID-19): a bibliometric analysis and knowledge mapping. F1000Res 2020 May 18;9:374. [doi: 10.12688/f1000research.23690.1]
13. Kupferschmidt K, Cohen J. Will novel virus go pandemic or be contained? Science 2020 Feb 07;367(6478):610-611. [doi: 10.1126/science.367.6478.610] [Medline: 32029604]
14. Bauchner H, Fontanarosa PB, Golub RM. Editorial Evaluation and Peer Review During a Pandemic: How Journals Maintain Standards. JAMA 2020 Aug 04;324(5):453-454. [doi: 10.1001/jama.2020.11764] [Medline: 32589195]
15. Coronavirus pandemic: Nature's pledge to you. Nature. 2020. URL: https://www.nature.com/articles/d41586-020-00882-z [accessed 2022-12-02]
16. The Lancet Global Health. Publishing in the time of COVID-19. The Lancet Global Health 2020 Jul;8(7):e860. [doi: 10.1016/s2214-109x(20)30260-6]
17. Kousha K, Thelwall M. Covid-19 refereeing duration and impact in major medical journals. Quantitative Science Studies 2022 Apr 12;3(1):1-17. [doi: 10.1162/qss_a_00176]
18. Bramstedt KA. The carnage of substandard research during the COVID-19 pandemic: a call for quality. J Med Ethics 2020 Dec;46(12):803-807. [doi: 10.1136/medethics-2020-106494] [Medline: 33004545]
19. Zdravkovic M, Berger-Estilita J, Zdravkovic B, Berger D. Scientific quality of COVID-19 and SARS CoV-2 publications in the highest impact medical journals during the early phase of the pandemic: A case control study. PLoS One 2020 Nov 5;15(11):e0241826 [FREE Full text] [doi: 10.1371/journal.pone.0241826] [Medline: 33152034]
20. Furlong A. Lancet to change peer review process following COVID-19 retraction. Politico. 2020. URL: https://www.politico.eu/article/lancet-review-process-following-covid-19-saga-coronavirus/ [accessed 2022-12-02]

Abbreviations

Annals: Annals of Internal Medicine
IF: impact factor
JAMA: Journal of the American Medical Association
NatMed: Nature Medicine
NEJM: The New England Journal of Medicine
The BMJ: The British Medical Journal

Edited by G Eysenbach; submitted 29.09.22; peer-reviewed by AY Gasparyan, F Elleuch; comments to author 21.11.22; revised version received 21.11.22; accepted 30.11.22; published 21.12.22.

Please cite as:
Delardas O, Giannos P
How COVID-19 Affected the Journal Impact Factor of High Impact Medical Journals: Bibliometric Analysis
J Med Internet Res 2022;24(12):e43089
URL: https://www.jmir.org/2022/12/e43089
doi:10.2196/43089
PMID:36454727

Original Paper

Evolution of Public Opinion on COVID-19 Vaccination in Japan: Large-Scale Twitter Data Analysis

Ryota Kobayashi[1,2,3*], PhD; Yuka Takedomi[4*], MA; Yuri Nakayama[1*], BSc; Towa Suda[4*], MA; Takeaki Uno[4], PhD; Takako Hashimoto[5], PhD; Masashi Toyoda[6], PhD; Naoki Yoshinaga[6], PhD; Masaru Kitsuregawa[6,7], PhD; Luis E C Rocha[8,9], PhD

[1]Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan
[2]Mathematics and Informatics Center, The University of Tokyo, Tokyo, Japan
[3]PRESTO, Japan Science and Technology Agency, Kawaguchi, Japan
[4]Principles of Informatics Research Division, National Institute of Informatics, Tokyo, Japan
[5]Faculty of Commerce and Economics, Chiba University of Commerce, Ichikawa, Japan
[6]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
[7]National Institute of Informatics, Tokyo, Japan
[8]Department of Economics, Ghent University, Ghent, Belgium
[9]Department of Physics and Astronomy, Ghent University, Ghent, Belgium
[*]these authors contributed equally

Corresponding Author:
Ryota Kobayashi, PhD
Graduate School of Frontier Sciences
The University of Tokyo
5-1-5 Kashiwanoha
Kashiwa, 277-8561
Japan
Phone: 81 4 7136 4134
Email: r-koba@k.u-tokyo.ac.jp

Abstract

Background: Vaccines are promising tools to control the spread of COVID-19. An effective vaccination campaign requires government policies and community engagement, sharing experiences for social support, and voicing concerns about vaccine safety and efficiency. The increasing use of online social platforms allows us to trace large-scale communication and infer public opinion in real time.

Objective: This study aimed to identify the main themes in COVID-19 vaccine-related discussions on Twitter in Japan and track how the popularity of the tweeted themes evolved during the vaccination campaign. Furthermore, we aimed to understand the impact of critical social events on the popularity of the themes.

Methods: We collected more than 100 million vaccine-related tweets written in Japanese and posted by 8 million users (approximately 6.4% of the Japanese population) from January 1 to October 31, 2021. We used Latent Dirichlet Allocation to perform automated topic modeling of tweet text during the vaccination campaign. In addition, we performed an interrupted time series regression analysis to evaluate the impact of 4 critical social events on public opinion.

Results: We identified 15 topics grouped into the following 4 themes: (1) personal issue, (2) breaking news, (3) politics, and (4) conspiracy and humor. The evolution of the popularity of themes revealed a shift in public opinion, with initial sharing of attention over personal issues (individual aspect), collecting information from news (knowledge acquisition), and government criticism to focusing on personal issues. Our analysis showed that the Tokyo Olympic Games affected public opinion more than other critical events but not the course of vaccination. Public opinion about politics was significantly affected by various social events, positively shifting attention in the early stages of the vaccination campaign and negatively shifting attention later.

Conclusions: This study showed a striking shift in public interest in Japan, with users splitting their attention over various themes early in the vaccination campaign and then focusing only on personal issues, as trust in vaccines and policies increased. An interrupted time series regression analysis showed that the vaccination rollout to the general population (under 65 years) increased the popularity of tweets about practical advice and personal vaccination experience, and the Tokyo Olympic Games

disrupted public opinion but not the course of the vaccination campaign. The methodology developed here allowed us to monitor the evolution of public opinion and evaluate the impact of social events on public opinion, using large-scale Twitter data.

(J Med Internet Res 2022;24(12):e41928) doi:10.2196/41928

KEYWORDS

COVID-19; vaccine; vaccination; Twitter; public opinion; topic modeling; longitudinal study; topic dynamics; social events; interrupted time series regression

Introduction

Vaccination is an effective mechanism to reduce the numbers of hospitalizations and deaths associated with the emergent coronavirus disease (COVID-19). With the advent of efficient vaccines after the first wave of the COVID-19 pandemic, public health efforts moved to strategies to cost-effectively immunize the population to increase survival and resume economic activity. Dose availability and uptake rates are fundamental to reaching sufficient vaccination coverage, but those numbers vary across countries in the current pandemic. One particular concern was the hesitancy regarding the safety and effectiveness of COVID-19 vaccines [1], which affected individuals' willingness to get vaccinated in not only low- and middle-income countries [2,3], but also high-income countries [4-7]. Japan stood out among developed economies as having one of the lowest vaccine confidence levels in the population [8]. This resulted from safety concerns about the human papillomavirus (HPV) vaccine that emerged in the early 2010s as a result of misinformation spread on the adverse effects of the HPV vaccine [9,10], prompting the Japanese Ministry of Health, Labour and Welfare to suspend the proactive recommendation of the HPV vaccine from June 2013 until November 2021. Such low public confidence delayed the start of mass vaccination against COVID-19, and Japan was 2 months behind the United States, China, and European countries, leading to safety concerns and inquiries regarding the Tokyo Olympic Games that had already been postponed to August 2021. Although mass vaccination started late in Japan, the country achieved high vaccination coverage in a short time and had one of the highest vaccination rates in the world (ranking 14th among 229 countries). Japan achieved a full vaccination rate of 72.4% on October 31, 2021 [11], ranking ahead of early adopters, such as the United Kingdom (67.0%), Germany (66.2%), and the United States (58.6%).

It is unclear how public opinion affected government policies that were also influenced by the domestic economic slowdown and the concerns about the Tokyo Olympic Games. Public opinion typically reacts to policies and might serve as a barometer of government strategies. Monitoring public opinion, however, is challenging. The largest study of vaccination intention in Japan surveyed 30,000 participants [7] and found that a large proportion of the population was unsure (33%) or unwilling (11%) to receive the COVID-19 vaccine, with side effects and safety being the main reasons. Classic survey studies like this are costly, relatively slow, and, with few exceptions [8], cannot trace changes in public opinion in real time [2,6,7,12]. Large-scale studies aiming to increase accuracy and the spatiotemporal resolution of responses require advanced survey techniques. In recent years, human activity has been increasingly mediated by digital devices, leaving footprints that can be exploited to assess the population's health and opinions [13-17]. In the context of COVID-19, social media data have been used to predict the number of new cases (incidence) [18,19] and to interpret the public perception of the pandemic [20]. Twitter has been particularly useful to monitor public opinion because users engage and react timely to environmental changes, for example, reacting to epidemic outbreaks [21,22], expressing concerns about the disease [23], accepting the pandemic situation [24], and reacting to vaccination issues [25-27]. Twitter is widely used in Japan, where more than 60% of people below 40 years old are actively engaged [28]. The pervasiveness of Twitter provides a unique source of data to monitor the evolution of public opinion during the various stages of the Japanese vaccination campaign.

In line with previous studies [25,26], we assumed that Twitter activity is a barometer of the public perception of COVID-19 vaccination. We thus focused on quantifying the public perception during the mass vaccination campaign in Japan by analyzing more than 100 million vaccine-related tweets posted by over 8 million users (approximately 6.4% of the Japanese population). The main goal was to understand the dynamics of public opinion during the vaccination campaign in Japan, which initially delayed the rollout of vaccines compared with other high-income countries. We hypothesized that such major social disruptions would lead the population to focus the debate on a few topics directly related to their daily experiences, in particular, their personal experiences with the vaccines. This debate could potentially generate social support and confidence to engage more people in the vaccination campaign. To examine the hypothesis, we identified the main topics on Twitter using the Latent Dirichlet Allocation (LDA) model [29]. We also hypothesized that public opinion would timely and semantically react to critical social events. The reactions would be for not only the stages of the vaccination campaign, but also major sports events like the Tokyo Olympic Games taking place during the vaccine rollout and the 5th COVID-19 wave. To examine the hypothesis, we quantified the effect of these critical events on the content of the debates on Twitter, using interrupted time series analysis [30].

Methods

Data Collection

We downloaded all Japanese tweets with the word "waku-chin" (vaccine in Japanese) posted between January 1, 2021, and October 31, 2021. The data set was provided by the NTT DATA Corporation [31]. We used data made available by NTT DATA Corporation to analyze all the vaccination-related tweets. The

Twitter application programming interface (API) has a limit for the number of tweets that can be downloaded in a month. The study period was chosen to include a short period before the launch of the vaccination campaign in Japan (February 17, 2021) and a short period after the end of the Tokyo Olympic Games when the full vaccination rate reached 70% of the Japanese population (October 25, 2021). The data set contained 114,357,691 tweets. We further collected data on the tweet text, the time stamp (posting time), and whether the tweet was an original tweet or a retweet. Using data from Our World in Data [32], we obtained the daily incidence (number of new cases) of COVID-19 and the full vaccination rate (the percentage of the population who received the second dose of the COVID-19 vaccine) in Japan [11]. The COVID-19 vaccines available in Japan were Pfizer, Moderna, AstraZeneca, and Takeda (Novavax), all of which require 2 doses.

Data Processing

Data processing and analysis were performed using Python software, version 3.9.7 (Python Software Foundation). We first extracted the plain text from the remaining tweets and removed emojis. Afterward, we segmented each text into Japanese words using the morphological analyzer MeCab [33] and removed stop words that have little analytic value (eg, “kore,” “sore,” and “suru” meaning “this,” “it,” and “do,” respectively, in Japanese). Finally, we changed words to their root forms (eg “boku” to “watashi” [“I” in Japanese] or “Utta” to “Utsu” [“inject” in Japanese]). This normalization corresponds to, for example, “viruses” to “virus” or “went” to “go” in English.

Topic Modeling

The LDA model [29] implemented in the *Gensim* Python package [34] was used to identify topics in the Twitter data. Before the topic modeling analysis, we removed rare words, that is, words appearing in fewer than 1000 tweets that corresponded to 0.0004% of the tweets, and the most frequent words “waku-chin” (vaccine) and “sessyu” (vaccination). In addition, we identified “bot” tweets by reading typical tweets in each topic obtained by the LDA model and removed the bot tweets until the LDA model did not identify artificial topics due to bots. To determine the number of topics, we calculated the topic coherence score C_V [35], which quantifies the quality of the topics obtained by the LDA model based on the probability distribution of the words. The coherence score C_V is defined as a complex function of the joint probability distribution of the words [35], and a high coherence score indicates that the topics are highly interpretable for humans, that is, the subject of the tweets within a topic is likely the same. We adopted the number of topics with the highest coherence score: K=15 (Multimedia Appendix 1). Finally, we assigned each tweet to a topic with the highest posterior probability that was calculated based on all the words in the tweet.

Interrupted Time Series Regression

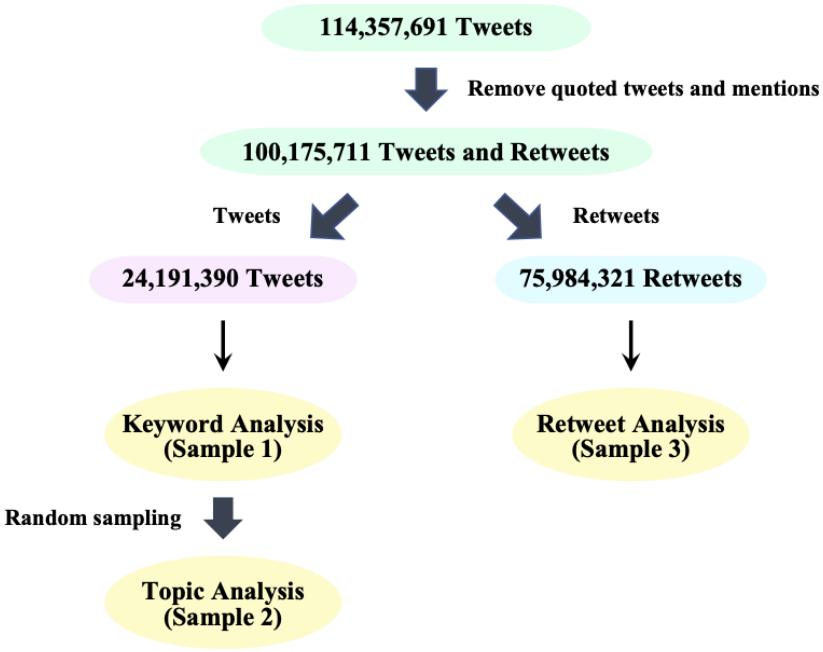
We used interrupted time series regression [30] to quantify the impact of major events (eg, the start of the Olympic games) on the popularity of a theme in tweets. The *statsmodels* Python package [36] was used for this analysis. The theme was defined based on a subset of keywords (Multimedia Appendix 2), and the analysis was based on all the tweet data (sample 1: 24 million tweets in Figure 1). We assumed the following regression model:

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t + u_t \text{ , } \textbf{(1)}$$

where Y_t is the percentage of a theme in tweets at time t (day), T is the number of days from the start of the observation period (T=0, which represents 30 days before each event), X_t is a dummy variable that equals to 0 and 1 before and after the event, respectively, and u_t is the error term. Here, β_0 represents the baseline popularity (in percentage) at T=0, β_1 represents the slope before the event, and β_2 and β_3 represent the level and slope change after the event, respectively.

A time series often exhibits autocorrelation, that is, the error terms are correlated over time, whereas the regression analysis assumes that the error terms u_t are uncorrelated. To evaluate the confidence intervals of the estimated parameters, we calculated the Newey-West standard error [37,38], also known as the heteroskedasticity- and autocorrelation-consistent standard error, which is robust to the autocorrelation. We also calculated the Newey-West standard error to evaluate the confidence intervals of the linear regression analysis.

Figure 1. Data processing workflow.



Results

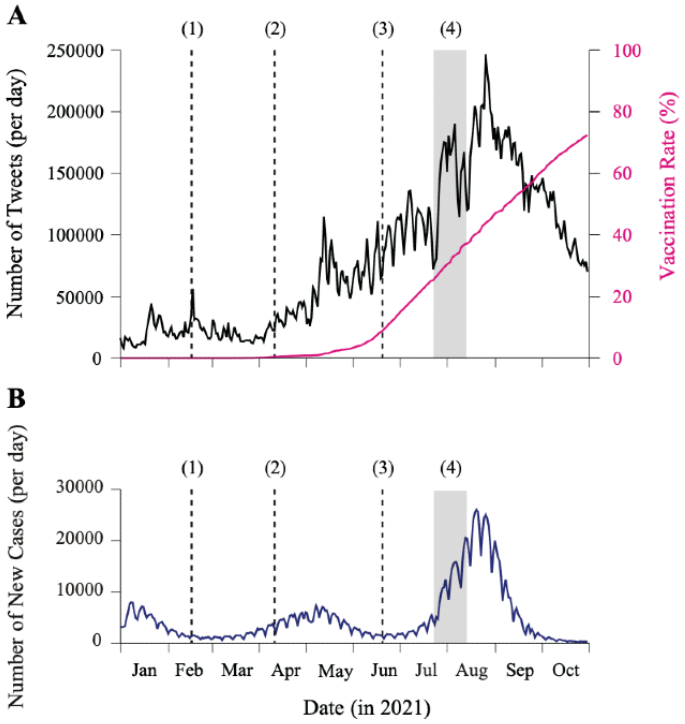
Dataset

The original data set contained 114,357,691 vaccine-related tweets written in Japanese from January 1 to October 31, 2021. Our analysis is based on 3 samples containing either the original tweets (24,191,390/114,357,691, 21.2%) or retweets (75,984,321/114,357,691, 66.4%) that do not contain any comments (Figure 1). Quoted tweets, that is, retweets with comments (5,765,735/114,357,691, 5.0%) and mentioned tweets (8,416,245/114,357,691, 7.4%) were excluded from our analysis because they were much fewer than the tweets and retweets. The first sample (Sample 1) contained 24,191,390 tweets posted by 6,034,435 users and was used to study the evolution of public opinion, including disruptions due to critical events. A random sample of the original data (Sample 2, N=1,000,000) was then used to identify the main topics and themes, and a sample of all retweets (Sample 3) was used to study the spread of opinions.

Vaccine-Related Tweets

Figure 2A shows the number of vaccine-related tweets per day during the study period and highlights the following 4 critical events [39]: (1) the launch of the COVID-19 vaccination campaign by the Japanese government on February 17, 2021, focusing initially on essential workers (eg, health care workers); (2) the start of vaccination of the elderly population (above 65 years old) on April 12, 2021; (3) the start of general public vaccination on June 21, 2021; and (4) the Tokyo Olympic Games taking place from July 23 to August 8, 2021. The first peak occurred on January 21, 2021, when Prime Minister Yoshihide Suga made a statement that “high coverage of vaccination is not a precondition for holding the Olympics in Tokyo” and the Ministry of Health, Labour and Welfare signed a contract with Pfizer Inc to supply a total of 72 million doses of its COVID-19 vaccine. Although a spike was observed at the very start of the vaccination campaign (event 1), vaccine-related tweets started to increase after event 2, when the vaccination of nonessential workers began. This coincided with the outbreak of the 4th wave in Japan (early April 2021; Figure 2B) and was followed by increased interest during the peak of infections in the 4th wave (May 13, 2021), when the online booking of vaccine appointments was launched but became overwhelmed, leaving many people without a vaccination slot. There was a sharp relative decrease in tweets at the start and end of the Olympic Games, followed by the largest peak on August 26, 2021, when a contamination scandal (approximately 1.6 million doses of the Moderna vaccine were discarded) was publicized. This last peak also coincided with the peak of the 5th wave and was followed by a substantial decrease in vaccine-related tweets, likely because of the high vaccination rate in the population (Figure 2). We found a low to moderate correlation between the number of tweets and the number of new COVID-19 cases (Spearman rank correlation: 0.322), suggesting that the effect of the pandemic situation on public opinion was not strong.

Figure 2. Vaccine-related tweets, vaccination rates, and incidence of COVID-19. (A) The number of vaccine-related tweets per day written in Japanese (black, left y-axis) and the fraction of the fully vaccinated population in Japan (magenta, right y-axis) between January 1 and October 31, 2021. (B) Daily incidence of COVID-19 in Japan. The vertical lines indicate 4 main events during the study period: (1) the launch of the COVID-19 vaccination campaign for essential workers; (2) the launch of vaccination for the elderly population (above 65 years old); (3) the launch of vaccination for the general population (under 65 years old); and (4) the period of the Tokyo Olympic Games.



Clustering Vaccine-Related Tweets

The ranking of the most used words on vaccine-related tweets (Sample 2 in Figure 1) revealed that 242,627 (24.3%) of them explicitly contained the word "COVID-19" (see Multimedia Appendix 3 for the frequent words). While the prevalence of specific words in tweets can reveal patterns of popular words, this measure is unable to unveil hidden semantic relations among tweets. We thus applied a machine learning methodology, the LDA model, to a sample of 1,000,000 tweets (100,000 per month, Sample 2) to automatically identify and classify (ie, cluster) tweets into meaningful topics. This monthly sampling was used here to remove the nonstationarity of the tweet activity given the imbalance in the number of vaccine-related tweets during the study period. Using LDA, we automatically identified 15 topics from the tweets (solely based on the textual content) and manually grouped them into the following 4 general themes: (1) personal issue, (2) breaking news, (3) politics, and (4) conspiracy and humor. Table 1 shows examples of representative tweets and the most popular words in each topic. Contributing terms were manually extracted from the top 30 weighted terms in the LDA model (Multimedia Appendix 4).

The most popular theme that emerged from the topic analysis was *personal issue* (Theme 1; 493,296/989,339 tweets, 49.9%), and it was formed by 2 topics about personal issues before being vaccinated, that is, personal view on vaccination and personal schedule of vaccination, and 4 topics about personal experiences after being vaccinated, that is, 1 topic about live reporting on the vaccination experience (eg, waiting room or to/from the vaccination center) and 3 topics about individual vaccination experiences including (1) complaints about discomfort, and side effects and personal life after vaccination; (2) reporting body temperature after taking the vaccine; and (3) advice to overcome side effects (Table 1).

The second most popular theme was *breaking news* (Theme 2; 210,550/989,339 tweets, 21.3%), and it included 2 topics about news on COVID-19 vaccines, such as vaccine development and approval, and vaccine effectiveness. The first topic included tweets about the development of Moderna, AstraZeneca, and Pfizer vaccines (clinical trials and government approvals) in Japan and other countries. The second topic was about the effectiveness of vaccines and contained information about mRNA vaccines, the effectiveness of vaccines against new variants, and serious side effects (eg, thrombus) of the AstraZeneca vaccine. The last topic was about booking an appointment for vaccination, in particular, about availability and whether users could successfully book a timeslot (Table 1).

Politics was the third most popular theme (Theme 3; 169,663/989,339 tweets, 17.1%), with 3 topics. The first topic was related to opinions on the government. For instance, users complained that the vaccination schedule in Japan was behind other countries and disagreed on holding the Tokyo Olympic Games given the low vaccination coverage. Opinions on mass media, such as complaints about unreliable information from the media and the attitude of the press inciting unrest, formed the second topic. Finally, the vaccination policy, including casual chats, for example, tweets mentioning the assignment of Mr Taro Kono (a politician famous among the young population) as vaccine minister, formed the third topic (Table 1).

Table 1. Topics identified from vaccine-related tweets before and during the COVID-19 vaccination campaign in Japan.

Themes and topics	Tweets (N=989,339), n (%)	Top terms contributing to the topic model	Representative tweet[a]
Theme 1: Personal issue	493,296 (49.9)		
Personal view on vaccination	170,095 (17.2)	I, think, myself, scary, absolutely, feeling, alright	"I'll be vaccinated because I want to. If you don't want to, you don't have to. I don't think I should tell others to be vaccinated!"
Personal schedule of vaccination	57,763 (5.8)	tomorrow, today, finish, clinic, appointment, next week, this week	"I'm finally getting the Pfizer COVID-19 vaccine tomorrow. I'm so excited."
Live reports of before/after vaccination	31,952 (3.2)	pain, go back, venue, swell, 30 minutes, sleepy, wait	"I've arrived the vaccination venue too early. I'm waiting and killing time."
Journal about vaccination experience	132,843 (13.4)	second time, adverse reaction, first time, yesterday, side effects, work, fine, temperature	"I've got the second shot. I was fine after the first shot, but I don't think I'm fine this time. Side effect will come sooner or later."
Perception after vaccination	65,490 (6.6)	pain, arm, injection, left arm, feel, discomfort	"The injection was given very quickly and was not very painful. It has been five hours since the injection, and I feel a little bit of discomfort in my left arm..."
Preparation for vaccination	35,153 (3.6)	fever, condition, second day, prepare, lighten, better, helpful	"I've got the second shot of vaccine! I need to buy sports drink when I go home... and most important of all, food for the cat!"
Theme 2: Breaking news	210,550 (21.3)		
Clinical trial and use authorization	79,247 (8.0)	Pfizer, Moderna, Ministry of Health, Labour and Welfare, development, start, clinical trial, approved	"Approval by the Ministry of Health, Labour and Welfare (MHLW) of the only vaccine for new coronavirus from US pharmaceutical giant Pfizer."
Effectiveness of vaccination	74,120 (7.5)	death, effectiveness, mRNA, report, research, Israel, variant	"It is reported that mRNA vaccines are effective against corona #corona #mRNA vaccine #effective."
Booking vaccination appointment	57,183 (5.8)	booking, preparation, group, available, campaign, system, local government	"The unprecedented scale of Mass vaccination: What is the preparation status of local municipalities?"
Theme 3: Politics	169,663 (17.1)		
Opinion about politics	95,219 (9.6)	Japan, measures, country, impossible, government, declaration of a state of emergency, Tokyo	"To the idiots in the government: if you can inoculate corona vaccine to all Japanese citizens, you can hold Olympic and Paralympic, but if you can't, cancel them."
Opinion about mass media	41,094 (4.2)	anxiety, information, news coverage, rumor, media, explanation, fact	"The mass media raised fears with coronas, and now they are raising fears with vaccines. Media should report the facts unbiasedly instead of raising fears."
Vaccination policy	33,350 (3.4)	third time, recently, tweet, video, shit, laugh	"It may be better to take a wait-and-see approach to the vaccination. It could be bad."
Theme 4: Conspiracy and humor	115,830 (11.7)		
Population control	41,428 (4.2)	population, human being, world, cause, conspiracy theory, reduction	"They developed the corona vaccine for the purpose of the Deep State agenda: global human enslavement, depopulation and money making!"
Effect on the body	30,221 (3.1)	children, 5G, freedom, destruction, discrimination	"It's very exciting to be able to connect to 5G when you are vaccinated!"
Internet meme	44,181 (4.5)	cine-cine	"Vac-vac-cine-cine! Vac-vac-cine-cine! Vac-vac! Cine-cine! Cine-cine vac-vac!"

[a]Original tweets are in Japanese.

The least popular theme contained topics related to conspiracy and humor (Theme 4; 115,830/989,339 tweets, 11.7%). The first topic was about control of the population, for example, the conspiracy theory that "the purpose of COVID-19 vaccination was to reduce the global population," and the second topic was about the effects on the body, for example, the theory that

"COVID-19 vaccines are a ploy to connect people to the 5G network." Internet memes formed the third topic, for example, the popular "Vac-vac-cine-cine" (from "vaccine vaccine" because a person needs 2 vaccine shots to be fully vaccinated and because the combination of these words sounds like "exciting" and "male genitalia" in Japanese) (Table 1).

Evolution of the Popularity of Themes

Previous research has shown that the number of tweets about a particular topic reflects the users' attention to that topic [40,41]. We thus estimated the popularity of tweets for each topic (grouped in 4 major themes; see the previous section) to monitor temporal changes in the interest of users (Figure 3). *Personal issue* (Theme 1) continuously increased, starting at nearly 30% and increasing to over 70% by the end of the study period. *Breaking news* (Theme 2) and *politics* (Theme 3), on the other hand, declined steadily from nearly 30% and 25%, respectively, to around 10%, dropping more significantly after June, when vaccination became available for people under 65 years old (the majority of Twitter users). *Conspiracy and humor* (Theme 4) also reduced slightly during the period and overall remained relatively low. We further validated this result by creating a subset of keywords for each theme (Multimedia Appendix 2) and then extracting all tweets of each theme from the original data set (24 million tweets) (Multimedia Appendix 5). The linear regression analysis (Table 2) showed a statistically significant increase in the tweets about *personal issue* (Theme 1) and a decrease in the other themes, with *breaking news* (Theme 2) and *politics* (Theme 3) decreasing 5 times in comparison to *conspiracy and humor* (Theme 4). These trends revealed a shift in the concerns of Twitter users, who initially shared their attention over personal issues (individual aspect), collecting information from the news (knowledge acquisition), and government decisions (the course of the vaccination campaign) and then focused mostly on personal issues once the vaccination campaign was effectively implemented in the general population.

The evolution of specific topics reflected finer aspects of the opinion dynamics. The combined topics about personal issues before being vaccinated (ie, personal view and personal schedule) increased after May followed by a slight decrease after August (Figure 4A). This pattern reflected increasing concerns with vaccination and the Tokyo Olympic Games that ended in early August. The combined topics about a user's experience after being vaccinated (ie, live reports, journal, perception, and preparation) showed a sharp increase after June, when the vaccination of the general population began (Figure 4A). Moreover, 17.9% (17,760/99,461) of the tweets belonged to the topic about personal issues after being vaccinated, even in January before the vaccination campaign in Japan. This is because the LDA model assigned a topic based on the words in a tweet (see Multimedia Appendix 4 for the top 30 terms).

In contrast, the popularity of conspiracy theories (population control and effect on the body) decreased steadily, indicating that education built up confidence in the vaccines (Figure 4B). Opinions on the booking of vaccination appointments peaked in May, when the booking system was launched. Opinions on politics peaked in April and then decreased substantially, reflecting an initial criticism toward the government for the late implementation of mass vaccination, followed by approval once the campaign rolled out. Again, we validated these findings by extracting the corresponding tweets using a subset of keywords for each topic or aggregated topic (Multimedia Appendix 2) and confirmed the trends (Multimedia Appendix 6), with a low prevalence of words related to conspiracy theories (1,452,528/24,032,297 tweets, 6.0%). This result also confirmed that the initial concerns about the government and the reliability of the vaccines became secondary once the vaccination reached most of the population and personal experiences became dominant.

Figure 3. Popularity of the study themes. Each line represents the percentage of tweets in each theme (Table 1) over time. The percentage is calculated monthly from a sample of vaccine-related tweets (1 million tweets: Sample 2).

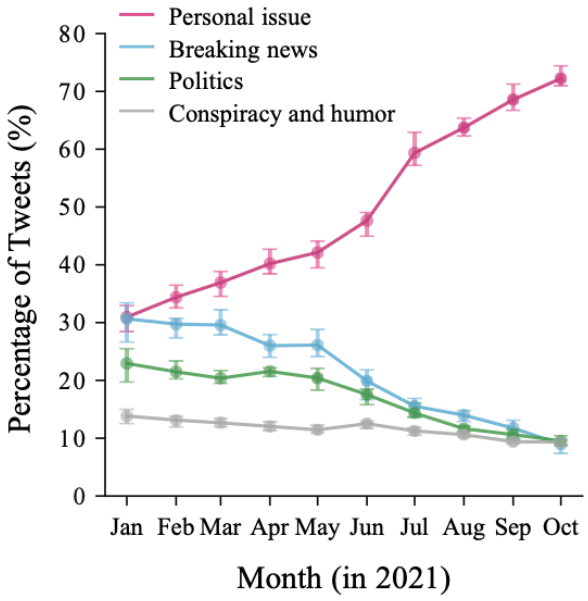


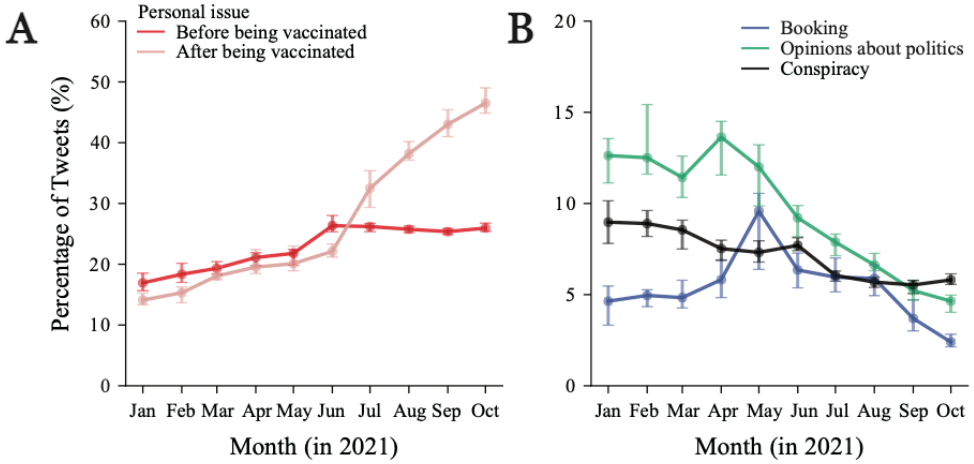
Table 2. Linear regression analysis of the popularity time series of the themes extracted by keywords.

Variable	Theme[a]			
	Theme 1	Theme 2	Theme 3	Theme 4
Intercept				
Coefficient	25.8[b]	44.0[b]	24.0[b]	8.75[b]
95% CI	23.9 to 27.6	41.2 to 46.8	22.2 to 25.8	8.22 to 9.27
Slope				
Coefficient	0.142[b]	−0.082[b]	−0.070[b]	−0.014[b]
95% CI	0.133 to 0.152	−0.097 to −0.067	−0.078 to −0.061	−0.017 to −0.011

[a]Theme 1: personal issue; Theme 2: breaking news; Theme 3: politics; and Theme 4: conspiracy and humor.

[b]Statistically significant change (P<.05).

Figure 4. Popularity of the topics. Each line represents the percentage of tweets over time. (A) Aggregated topics about personal issue before/after being vaccinated (Theme 1). (B) Topics about booking vaccination appointment (Booking in Theme 2), opinions about politics (Opinions about politics in Theme 3), and aggregated topics about conspiracy theories, that is, population control and effect on the body (Conspiracy in Theme 4). The percentage is calculated monthly from a sample of vaccine-related tweets (1 million tweets: Sample 2).



Shift in Interest After Critical Events

Specific events may have social and individual consequences and may affect public opinion and discussion of different themes. Four critical events marked the vaccination campaign in Japan during 2021 (the various stages of the vaccination campaign and the Tokyo Olympic Games; Figure 2). To test our hypothesis of critical events on opinion dynamics, we performed interrupted time series regression [30] to estimate the changes in the popularity of themes (see the Methods section). We first calculated the popularity (ie, the percentage of tweets) of 4 themes defined by subsets of keywords (Multimedia Appendix 2). The themes were as follows: Theme 1, personal issue; Theme 2, breaking news; Theme 3, politics; and Theme 4, conspiracy and humor. In this analysis, the level parameter (β_2 in Equation 1) indicates a shift in the relative attention, whereas the slope parameter (β_3 in Equation 1) indicates a shift in the rate of popularity increase of a given theme. Table 3 shows that *politics* (Theme 3) was the theme most affected by these events. The impact of the general population vaccination rollout and the Tokyo Olympic Games on public opinion was larger than that of the other critical events, and they affected all aspects of public opinion. The vaccination of health workers positively shifted the popularity of the *politics* theme, likely because of increasing expectations of rolling out mass vaccination. The vaccination of the elderly population only positively shifted the trend. On the other hand, both the vaccination of the general population and the Tokyo Olympic Games negatively shifted the interest in politics, suggesting relatively fewer concerns with government policies. Furthermore, the vaccination rollout of the general population increased the rate of tweets about practical advice and personal experience. Finally, the start of the Tokyo Olympic Games caused an increase in interest in personal issues that remained nearly constant afterward (Figure 5), likely because of the large vaccination coverage achieved during this period.

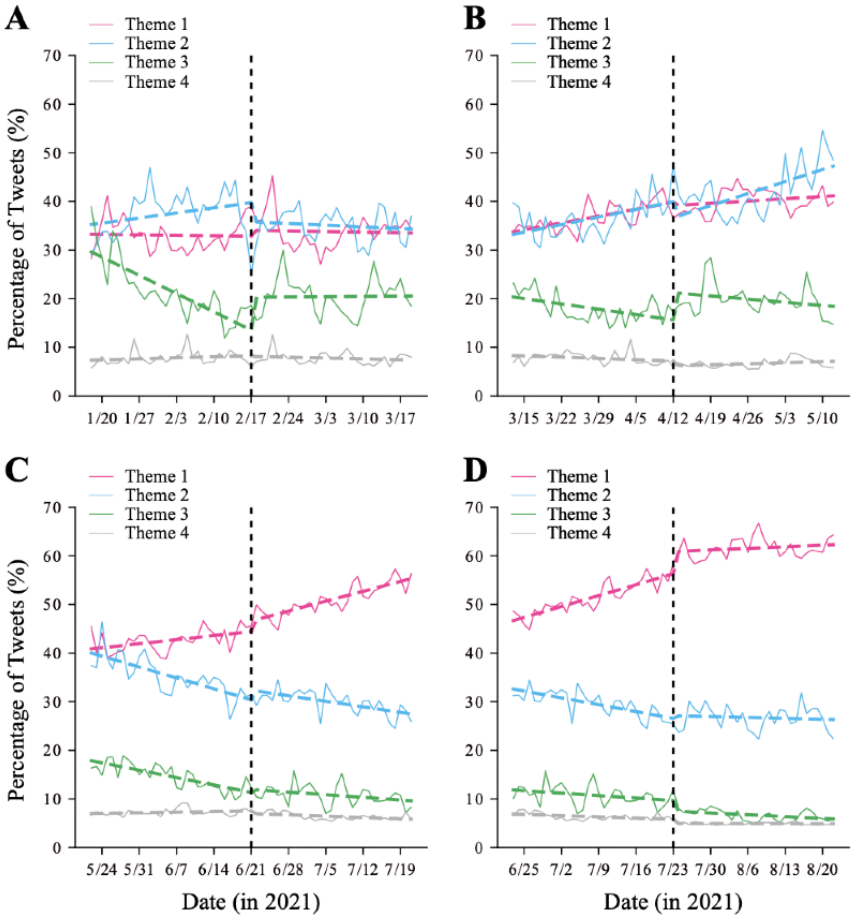
Table 3. Changes in the popularity of the 4 themes at critical events.

Variable	Theme[a]			
	Theme 1	Theme 2	Theme 3	Theme 4
Health workers				
Level (β_2)				
Coefficient	1.21	−3.93	6.74[b]	−0.14
95% CI	−4.51 to 6.93	−8.88 to 1.03	1.27 to 12.2	−1.45 to 1.18
Slope (β_3)				
Coefficient	−0.01	−0.20	0.54[b]	−0.05
95% CI	−0.32 to 0.31	−0.55 to 0.15	0.30 to 0.80	−0.14 to 0.03
Elderly population				
Level (β_2)				
Coefficient	−0.71	−3.25	5.59[b]	−0.98
95% CI	−3.54 to 2.13	−7.79 to 1.28	1.60 to 9.58	−2.11 to 0.15
Slope (β_3)				
Coefficient	−0.14	0.13	0.06	0.07
95% CI	−0.28 to 0.01	−0.17 to 0.42	−0.15 to 0.27	−0.01 to 0.14
General population				
Level (β_2)				
Coefficient	2.17 [b]	1.95	0.60	−0.51
95% CI	0.49 to 3.85	−0.15 to 4.05	−1.21 to 2.42	−1.33 to 0.30
Slope (β_3)				
Coefficient	0.17[b]	0.16[b]	0.14[b]	−0.06[b]
95% CI	0.05 to 0.29	0.06 to 0.26	0.06 to 0.22	−0.09 to −0.02
Olympic Games				
Level (β_2)				
Coefficient	4.57[b]	0.54	−2.17[b]	−0.89[b]
95% CI	2.80 to 6.35	−1.42 to 2.51	−3.19 to −1.15	−1.32 to −0.46
Slope (β_3)				
Coefficient	−0.27[b]	0.18[b]	0.02	0.03[b]
95% CI	−0.37 to −0.17	0.04 to 0.32	−0.05 to 0.08	0.00 to 0.07

[a]Theme 1: personal issue; Theme 2: breaking news; Theme 3: politics; and Theme 4: conspiracy and humor.

[b]Statistically significant change (P<.05).

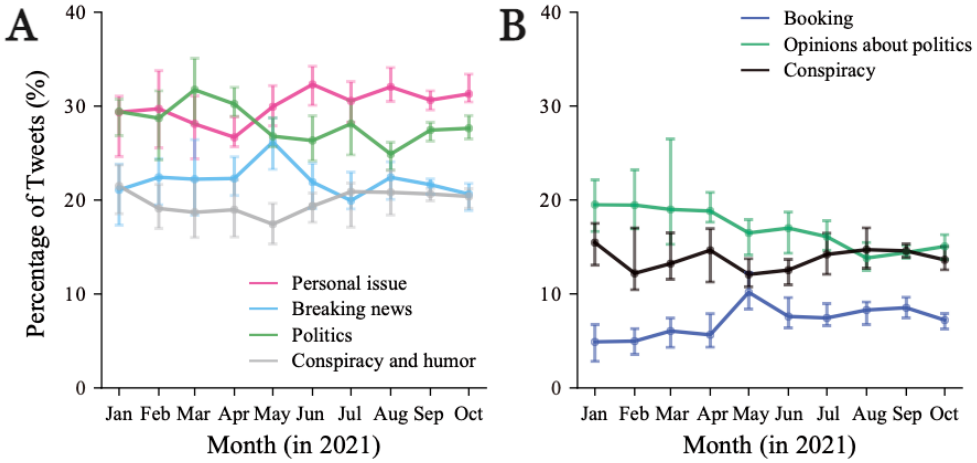
Figure 5. Impact of social events on the popularity of the themes. We applied interrupted time series regression to the popularity time series of each theme (Theme 1: personal issue; Theme 2: breaking news; Theme 3: politics; and Theme 4: conspiracy and humor). We examined the following 4 major events during the vaccination period: (A) vaccination start for health workers, (B) vaccination start for older people, (C) vaccination start for the general population (under 65 years), and (D) start of the Olympic Games in Tokyo.



Spread of Opinions

A tweet is a unidirectional process of sharing information with the community. Retweeting, on the other hand, is a social process where users engage and share tweets to spread opinions on their own social network [42]. The analysis of 75,984,321 retweets by 3,917,181 users (Sample 3) showed a higher prevalence of retweets about *personal issue* (Theme 1) and *politics* (Theme 3) in comparison to *breaking news* (Theme 2) and *conspiracy and humor* (Theme 4) (Figure 6A). Those observations aligned with the theory of complex contagion, since users mostly engaged with tweets (by retweeting) related to personal experiences and political opinion rather than tweets sharing hard-to-verify information, such as vaccine reliability and conspiracy theories, that might have negative consequences and might affect the credibility of the user retweeting [43]. Similar to the popularity of certain topics, the social process is also intensified during certain periods (Figure 6B). For instance, the topic of booking an appointment exhibited a peak in May, coinciding with the popularity of this topic, whereas the topic of politics declined after April, when vaccination of the elderly population started.

Figure 6. Popularity of retweeted themes or topics. Each line represents the percentage of frequently retweeted tweets (retweeted more than 10 times in a day) in each theme (A) or topic (B) over time. The topics of Conspiracy represent combined topics of population control and effect on the body.



Discussion

Principal Findings

This study aimed to understand the dynamics of public opinion during the vaccination campaign in Japan, which initially delayed the rollout of vaccines compared with other high-income countries. We leveraged the textual information in tweets and performed a topic analysis of vaccine-related tweets to identify 15 topics further grouped into the following 4 major themes: (1) personal issue, (2) breaking news, (3) politics, and (4) conspiracy and humor, during the vaccination campaign in Japan (from January 1 to October 31, 2021). We found a striking shift in public interest, with users splitting their attention over various themes early in the campaign and then focusing on personal issues, as trust in vaccines and policies built up with an effective vaccination campaign. Next, we examined the effect of critical social events on the popularity of the tweet themes. We found that the vaccination rollout to the general population (under 65 years old) increased the popularity of tweets about practical advice and personal vaccination experience. This result implies that the start of vaccination of the general population was a critical event for Twitter users (mostly 20-30 years old in Japan). We also found that the popularity of the themes remained at the same level during the Olympic Games.

Comparison With Prior Work

Previous studies using social media (Twitter and Reddit) [25-27] to study public opinions of COVID-19 vaccination in different countries were limited in sample size and did not cover the whole vaccination campaign. Therefore, only topics related to breaking news [25-27] and politics [25] were identified. We showed, however, that personal issue is a common topic emerging during critical periods and is fundamental for a successful mass vaccination campaign, since it bonds people via social support. Furthermore, our findings are more robust than the findings of existing studies [25-27] because the main results (Figure 3 and 4) were confirmed by a robustness analysis using the whole data set (Multimedia Appendix 5 and Multimedia Appendix 6). While we could not collect all tweets via the Twitter API, we could still analyze all vaccination-related tweets by using comprehensive data from NTT DATA Corporation.

Furthermore, the interrupted time series regression analysis showed that the vaccination rollout of the general population and the Tokyo Olympic Games affected public opinion more than other critical events. Public opinion on politics was the most significantly affected debate, positively shifting attention early in the vaccination campaign and negatively later. In addition, social dialogue was maintained with tweets about personal issues mostly retweeted when vaccination reached the adult population, which is the most active user group on Twitter.

Limitations

There are limitations in our study. First, it was impossible to avoid sampling bias in the online data set even though we analyzed all Japanese tweets including the word “waku-chin” (vaccine in Japanese). We analyzed the tweets posted by 8 million users (approximately 6.4% of the Japanese population), which is comprehensive and represents the opinion of active users but might not fully represent the general public. Nevertheless, Twitter data are representative of the opinion of the younger generation (20-30 years old) in Japan, which is supported by a survey [28] reporting that more than 60% of the population below 40 years old is actively engaged on Twitter. To minimize potential sampling biases, we resampled the original data of 6 million users to remove temporal effects. Unlike standard survey studies, we were unable to collect sociodemographic information and thus could not stratify the analysis to age group, location, education, and gender [3,7]. Stratification would help us to assess the extent to which certain social groups (eg, adults vs elderly) and locations (eg, Tokyo during the Olympic Games) were affected. Second, the study population was limited to those using Twitter in Japan. While this limitation enables us to understand the public opinion of Japanese Twitter users, the results may not be generalizable to other countries, such as the United States, China, and European and African countries. Future work is necessary to compare the public opinions of users in Japan to those in other countries. Finally, the inclusion criterion of the keyword “vaccine” may have captured tweets not relevant to COVID-19, such as those related to the HPV vaccine or to pet vaccination. To assess this aspect, we manually reviewed tweets and found that most of

them were not contaminated by discussions of other types of vaccines.

We used the LDA model to identify topics from tweets, which assigns a tweet to a topic based on the words present. However, a topic might contain several issues. For example, more than 10% of the tweets in January (before the vaccination campaign in Japan) were classified under the topic "after being vaccinated." This is because the terms that contributed to the topic (eg, "side reaction" and "mask") were used in January. Moreover, we applied interrupted time series analysis to examine the impact of critical social events on popular topics on Twitter. While the standard interrupted time series analysis [30] neglected the effect of autocorrelation in the time series, we incorporated it by calculating the Newey-West standard error to evaluate the confidence intervals. Notably, the low to moderate correlation between the number of tweets and the number of new cases in a day suggests that the pandemic status might impact the popular topics on Twitter. It would be interesting to further investigate the effect of the pandemic's status on the popular topics on Twitter and incorporate it into the analysis. Finally, we manually identified bot retweets that impacted the topics obtained by the LDA model. Bot detection is a challenging research issue, and there have been only few attempts to identify bots (eg, Botometer [44]). Although we applied Botometer to our data set, it was unable to identify the bots we excluded in this study. Further studies are required to establish guidelines to identify tweets posted by bot accounts.

Conclusions

We studied the evolution of public opinion regarding COVID-19 vaccination in Japan by analyzing more than 100 million vaccine-related tweets. We identified the following 4 themes in the tweets: (1) personal issue, (2) breaking news, (3) politics, and (4) conspiracy and humor. We found a striking shift in public interest. Users split their attention over various themes early in the campaign and then focused on personal issues, as trust in vaccines built up with an effective vaccination campaign. An interrupted time series regression analysis showed that the vaccination rollout to the general population (under 65 years old) increased the popularity of tweets about practical advice and personal vaccination experience, and the Tokyo Olympic Games disrupted public opinion but not the course of the vaccination campaign. The methodology developed here allowed us to monitor the evolution of public opinion and evaluate the impact of social events on public opinion, using large-scale Twitter data.

Acknowledgments

This research was conducted as part of the "COVID-19 AI & Simulation Project" run by Mitsubishi Research Institute commissioned by the Cabinet Secretariat. The methodology used in this study was developed through support from the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant numbers JP18K11560, JP19H0113, JP21H03559, JP21H04571, JP22H03695, JP22K12285, and JPJSBP120202201), Japan Science and Technology Agency CREST (grant number JP-MJCR1401), Japan Science and Technology Agency PRESTO (grant number JPMJPR1925), and Japan Agency for Medical Research and Development (grant number JP21wm0525004).

Authors' Contributions

RK, YT, YN, and TS are co-first authors. RK conceived the study, and RK and LECR designed the study. TH, MT, NY, and MK collected and preprocessed the data. YN, RK, YT, TS, and TU analyzed the data. RK, YN, and YT created the figures. YT, TS, YN, and RK created the tables. RK and LECR wrote the manuscript. All authors discussed the results and contributed to finalizing the manuscript. RK supervised the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Dependency of the number of topics on the coherence score.
[PNG File , 69 KB - jmir_v24i12e41928_app1.png]

Multimedia Appendix 2
List of keywords for the 4 themes and their subthemes.
[DOCX File , 459 KB - jmir_v24i12e41928_app2.docx]

Multimedia Appendix 3
Top 50 used words in vaccine-related tweets: Sample 2 (1 million tweets).
[DOCX File , 227 KB - jmir_v24i12e41928_app3.docx]

Multimedia Appendix 4
Top 30 contributing terms of each topic identified from vaccine-related tweets.

[DOCX File , 838 KB - jmir_v24i12e41928_app4.docx]

Multimedia Appendix 5
Popularity of the themes defined based on the subset of keywords (Multimedia Appendix 2). Each line represents the percentage of tweets of each theme over time. The percentage was calculated for each month (A) and day (B). Dashed lines in panel (B) represent the fitted lines obtained by the linear regression.
[PNG File , 386 KB - jmir_v24i12e41928_app5.png]

Multimedia Appendix 6
Popularity of the subthemes defined based on the subset of keywords (Multimedia Appendix 2). (A) Subthemes related to personal issue corresponding to Theme 1 (Figure 4A). (B) Subthemes related to reservation, politics, and conspiracy theories corresponding to Themes 2, 3, and 4, respectively (Figure 4B).
[PNG File , 240 KB - jmir_v24i12e41928_app6.png]

References

1. Wouters OJ, Shadlen KC, Salcher-Konrad M, Pollard AJ, Larson HJ, Teerawattananon Y, et al. Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment. The Lancet 2021 Mar;397(10278):1023-1034. [doi: 10.1016/s0140-6736(21)00306-8]
2. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb 20;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
3. Solís Arce JS, Warren SS, Meriggi NF, Scacco A, McMurry N, Voors M, et al. COVID-19 vaccine acceptance and hesitancy in low- and middle-income countries. Nat Med 2021 Aug 16;27(8):1385-1394 [FREE Full text] [doi: 10.1038/s41591-021-01454-y] [Medline: 34272499]
4. Dror AA, Eisenbach N, Taiber S, Morozov NG, Mizrachi M, Zigron A, et al. Vaccine hesitancy: the next challenge in the fight against COVID-19. Eur J Epidemiol 2020 Aug;35(8):775-779 [FREE Full text] [doi: 10.1007/s10654-020-00671-y] [Medline: 32785815]
5. Fisher KA, Bloomstone SJ, Walder J, Crawford S, Fouayzi H, Mazor KM. Attitudes toward a potential SARS-CoV-2 vaccine: A survey of U.S. adults. Ann Intern Med 2020 Dec 15;173(12):964-973 [FREE Full text] [doi: 10.7326/M20-3569] [Medline: 32886525]
6. Neumann-Böhme S, Varghese NE, Sabat I, Barros PP, Brouwer W, van Exel J, et al. Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. Eur J Health Econ 2020 Sep;21(7):977-982 [FREE Full text] [doi: 10.1007/s10198-020-01208-6] [Medline: 32591957]
7. Nomura S, Eguchi A, Yoneoka D, Kawashima T, Tanoue Y, Murakami M, et al. Reasons for being unsure or unwilling regarding intention to take COVID-19 vaccine among Japanese people: A large cross-sectional national survey. Lancet Reg Health West Pac 2021 Sep;14:100223 [FREE Full text] [doi: 10.1016/j.lanwpc.2021.100223] [Medline: 34368797]
8. de Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study. Lancet 2020 Sep 26;396(10255):898-908 [FREE Full text] [doi: 10.1016/S0140-6736(20)31558-0] [Medline: 32919524]
9. Hanley SJB, Yoshioka E, Ito Y, Kishi R. HPV vaccination crisis in Japan. Lancet 2015 Jun 27;385(9987):2571. [doi: 10.1016/S0140-6736(15)61152-7] [Medline: 26122153]
10. Simms KT, Hanley SJB, Smith MA, Keane A, Canfell K. Impact of HPV vaccine hesitancy on cervical cancer in Japan: a modelling study. The Lancet Public Health 2020 Apr;5(4):e223-e234 [FREE Full text] [doi: 10.1016/S2468-2667(20)30010-4] [Medline: 32057317]
11. 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. [doi: 10.1038/s41562-021-01122-8] [Medline: 33972767]
12. Murphy J, Vallières F, Bentall RP, Shevlin M, McBride O, Hartman TK, et al. Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. Nat Commun 2021 Jan 04;12(1):29 [FREE Full text] [doi: 10.1038/s41467-020-20226-9] [Medline: 33397962]
13. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
14. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using Twitter. In: EMNLP '11: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 27-31, 2011; Edinburgh, United Kingdom p. 1568-1576. [doi: 10.5555/2145432.2145600]
15. Proskurnia J, Grabowicz P, Kobayashi R, Castillo C, Cudré-Mauroux P, Aberer K. Predicting the Success of Online Petitions Leveraging Multidimensional Time-Series. In: WWW '17: Proceedings of the 26th International Conference on World

Wide Web. 2017 Presented at: 26th International Conference on World Wide Web; April 3-7, 2017; Perth, Australia p. 755-764. [doi: 10.1145/3038912.3052705]
16. 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]
17. Du Z, Wang L, Yang B, Ali ST, Tsang TK, Shan S, et al. Risk for international importations of variant SARS-CoV-2 originating in the United Kingdom. Emerg Infect Dis 2021 May;27(5):1527-1529 [FREE Full text] [doi: 10.3201/eid2705.210050] [Medline: 33760727]
18. O'Leary DE, Storey VC. A Google–Wikipedia–Twitter model as a leading indicator of the numbers of coronavirus deaths. Intell Sys Acc Fin Mgmt 2020 Sep 28;27(3):151-158. [doi: 10.1002/isaf.1482]
19. Qin L, Sun Q, Wang Y, Wu K, Chen M, Shia B, et al. Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. Int J Environ Res Public Health 2020 Mar 31;17(7):2365 [FREE Full text] [doi: 10.3390/ijerph17072365] [Medline: 32244425]
20. 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):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
21. Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An "infodemic": Leveraging high-volume Twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. Open Forum Infect Dis 2020 Jul;7(7):ofaa258 [FREE Full text] [doi: 10.1093/ofid/ofaa258] [Medline: 33117854]
22. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
23. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
24. Aiello LM, Quercia D, Zhou K, Constantinides M, Šćepanović S, Joglekar S. How epidemic psychology works on Twitter: evolution of responses to the COVID-19 pandemic in the U.S. Humanit Soc Sci Commun 2021 Jul 23;8(1):179. [doi: 10.1057/s41599-021-00861-3]
25. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on Twitter: Topic modeling and sentiment analysis. J Med Internet Res 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: 10.2196/24435] [Medline: 34115608]
26. Kwok SWH, Vadde SK, Wang G. Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: Machine learning analysis. J Med Internet Res 2021 May 19;23(5):e26953 [FREE Full text] [doi: 10.2196/26953] [Medline: 33886492]
27. Wu W, Lyu H, Luo J. Characterizing discourse about COVID-19 vaccines: A Reddit version of the pandemic story. Health Data Science 2021 Aug 27;2021:1-11. [doi: 10.34133/2021/9837856]
28. Survey on Information and Communication Media Usage Time and Behavior in Japan. Japanese Ministry of Internal Affairs and Communications. 2021. URL: https://www.soumu.go.jp/main_content/000765258.pdf [accessed 2022-08-15]
29. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. The Journal of Machine Learning Research 2003;3:993-1022. [doi: 10.5555/944919.944937]
30. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2017 Feb 01;46(1):348-355 [FREE Full text] [doi: 10.1093/ije/dyw098] [Medline: 27283160]
31. NTT DATA Corporation. URL: https://www.nttdata.com [accessed 2022-08-15]
32. Our World in Data. URL: https://ourworldindata.org [accessed 2022-08-15]
33. Mecab: Yet another part-of-speech and morphological analyzer. GitHub. URL: https://github.com/taku910/mecab [accessed 2022-08-15]
34. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. In: Proceedings of the LREC Workshop on New Challenges for NLP Frameworks. 2010 Presented at: LREC Workshop on New Challenges for NLP Frameworks; 2010; Valletta, Malta p. 46-50. [doi: 10.13140/2.1.2393.1847]
35. Röder M, Both A, Hinneburg A. Exploring the Space of Topic Coherence Measures. In: WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 Presented at: Eighth ACM International Conference on Web Search and Data Mining; February 2-6, 2015; Shanghai, China p. 399-408. [doi: 10.1145/2684822.2685324]
36. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. In: Proceedings of the 9th Python in Science Conference. 2010 Presented at: 9th Python in Science Conference; June 28-July 3, 2010; Austin, Texas p. 92-96. [doi: 10.25080/majora-92bf1922-011]
37. Newey WK, West KD. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 1987 May;55(3):703. [doi: 10.2307/1913610]
38. Andrews DWK. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 1991 May;59(3):817. [doi: 10.2307/2938229]
39. Yamakuma E. COVID-19 Vaccinations in Japan. The Japan Times. 2021. URL: https://www.japantimes.co.jp/how-to-coronavirus-vaccinations-japan [accessed 2022-08-15]

40. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? In: WWW '10: Proceedings of the 19th International Conference on World Wide Web. 2010 Presented at: 19th International Conference on World Wide Web; April 26-30, 2010; Raleigh, North Carolina p. 591-600. [doi: 10.1145/1772690.1772751]
41. Lehmann J, Gonçalves B, Ramasco J, Cattuto C. Dynamical classes of collective attention in Twitter. In: WWW '12: Proceedings of the 21st International Conference on World Wide Web. 2012 Presented at: 21st International Conference on World Wide Web; April 16-20, 2012; Lyon, France p. 251-260. [doi: 10.1145/2187836.2187871]
42. Boyd D, Golder S, Lotan G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. 2010 Presented at: 43rd Hawaii International Conference on System Sciences; January 05-08, 2010; Honolulu, Hawaii p. 1-10. [doi: 10.1109/hicss.2010.412]
43. State B, Adamic L. The Diffusion of Support in an Online Social Movement: Evidence from the Adoption of Equal-Sign Profile Pictures. In: CSCW '15: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2015 Presented at: 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; March 14-18, 2015; Vancouver, BC, Canada p. 1741-1750. [doi: 10.1145/2675133.2675290]
44. Sayyadiharikandeh M, Varol O, Yang KC, Flammini A, Menczer F. Detection of Novel Social Bots by Ensembles of Specialized Classifiers. In: CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020 Presented at: 29th ACM International Conference on Information & Knowledge Management; October 19-23, 2020; Virtual Event, Ireland p. 2725-2732. [doi: 10.1145/3340531.3412698]

Abbreviations

API: application programming interface
HPV: human papillomavirus
LDA: Latent Dirichlet Allocation

Edited by C Basch, G Eysenbach; submitted 16.08.22; peer-reviewed by Z Butt, T Tsang, H Mondal, SF Tsao; comments to author 07.09.22; revised version received 09.10.22; accepted 24.10.22; published 22.12.22.

Please cite as:
Kobayashi R, Takedomi Y, Nakayama Y, Suda T, Uno T, Hashimoto T, Toyoda M, Yoshinaga N, Kitsuregawa M, Rocha LEC
Evolution of Public Opinion on COVID-19 Vaccination in Japan: Large-Scale Twitter Data Analysis
J Med Internet Res 2022;24(12):e41928
URL: https://www.jmir.org/2022/12/e41928
doi:10.2196/41928
PMID:36343186

Original Paper

Sentiment Analysis of Insomnia-Related Tweets via a Combination of Transformers Using Dempster-Shafer Theory: Pre– and Peri–COVID-19 Pandemic Retrospective Study

Arash Maghsoudi[1], PhD; Sara Nowakowski[1], PhD; Ritwick Agrawal[1], MD; Amir Sharafkhaneh[1], MD; Mark E Kunik[1], MD; Aanand D Naik[2], MD; Hua Xu[3], PhD; Javad Razjouyan[1], PhD

[1]Department of Medicine, Baylor College of Medicine, Houston, TX, United States

[2]Department of Management, Policy, and Community Health, University of Texas School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[3]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States

Corresponding Author:
Javad Razjouyan, PhD
Department of Medicine
Baylor College of Medicine
One Baylor Plaza
Houston, TX, 77030
United States
Phone: 1 713 798 4951
Email: javad.razjouyan@bcm.edu

Abstract

Background: The COVID-19 pandemic has imposed additional stress on population health that may result in a change of sleeping behavior.

Objective: In this study, we hypothesized that using natural language processing to explore social media would help with assessing the mental health conditions of people experiencing insomnia after the outbreak of COVID-19.

Methods: We designed a retrospective study that used public social media content from Twitter. We categorized insomnia-related tweets based on time, using the following two intervals: the prepandemic (January 1, 2019, to January 1, 2020) and peripandemic (January 1, 2020, to January 1, 2021) intervals. We performed a sentiment analysis by using pretrained transformers in conjunction with Dempster-Shafer theory (DST) to classify the polarity of emotions as *positive*, *negative*, and *neutral*. We validated the proposed pipeline on 300 annotated tweets. Additionally, we performed a temporal analysis to examine the effect of time on Twitter users' insomnia experiences, using logistic regression.

Results: We extracted 305,321 tweets containing the word *insomnia* (prepandemic tweets: n=139,561; peripandemic tweets: n=165,760). The best combination of pretrained transformers (combined via DST) yielded 84% accuracy. By using this pipeline, we found that the odds of posting negative tweets (odds ratio [OR] 1.39, 95% CI 1.37-1.41; *P*<.001) were higher in the peripandemic interval compared to those in the prepandemic interval. The likelihood of posting negative tweets after midnight was 21% higher than that before midnight (OR 1.21, 95% CI 1.19-1.23; *P*<.001). In the prepandemic interval, while the odds of posting negative tweets were 2% higher after midnight compared to those before midnight (OR 1.02, 95% CI 1.00-1.07; *P*=.008), they were 43% higher (OR 1.43, 95% CI 1.40-1.46; *P*<.001) in the peripandemic interval.

Conclusions: The proposed novel sentiment analysis pipeline, which combines pretrained transformers via DST, is capable of classifying the emotions and sentiments of insomnia-related tweets. Twitter users shared more negative tweets about insomnia in the peripandemic interval than in the prepandemic interval. Future studies using a natural language processing framework could assess tweets about other types of psychological distress, habit changes, weight gain resulting from inactivity, and the effect of viral infection on sleep.

(J Med Internet Res 2022;24(12):e41517) doi:10.2196/41517

KEYWORDS

COVID-19; coronavirus; sleep; Twitter; natural language processing; sentiment analysis; transformers; Dempster-Shafer theory; sleeping; social media; pandemic; effect; viral infection

Introduction

The COVID-19 pandemic has imposed excessive stress on the world population [1,2] through financial instability, unemployment, social isolation, and a lack of social activities [3]. Prior studies established the association between this stress and sleep disturbances [4-6]. Additionally, due to the pandemic, restrictions such as social distancing have resulted in the increase of certain digital behaviors, including distance learning, web-based meetings, web-based shopping, and social media usage [7-9]. The rise in the usage of social media platforms, like Twitter, provides researchers with a new source of data for screening public behavior.

Several studies have reported the impact of the COVID-19 pandemic on sleep quality and mental health [10-17]. However, these studies were limited to small databases, data gathered through questionaries, or both, and they lacked a comparison group. For instance, one study used Twitter to report the effect of the COVID-19 pandemic on the sleep quality of pregnant women based on 192 tweets [18]. The sentiment analysis of social media content is a challenging task, since such texts are unstructured, brief, informal, and casual; are prone to mistakes in dictation and grammar; and are noisy (emojis, hashtags, URLs, etc); and they entail ambiguities, such as polysemy [19]. Therefore, using artificial intelligence and machine learning tools and techniques may prove to be beneficial for tackling these challenges. Among these tools are advanced, analytical natural language processing (NLP) algorithms called *transformers* [19-26]. They are newly proposed tools and extensions to previous versions of a deep artificial neural network—recurrent neural networks—for language modeling and language encoding.

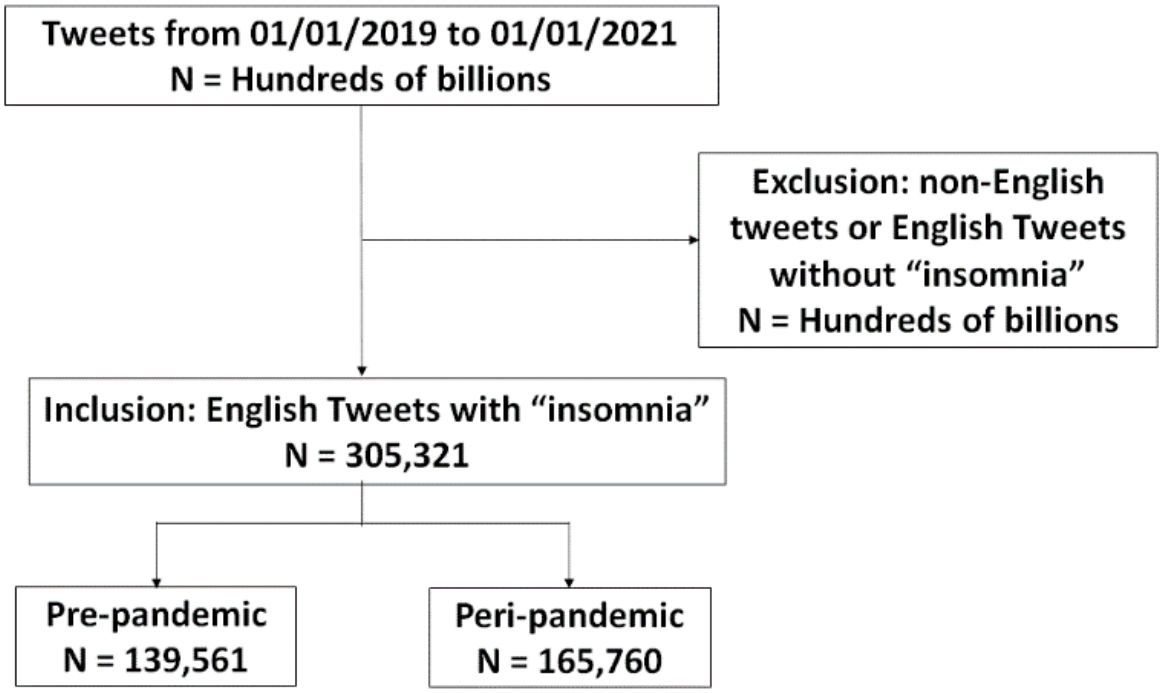
We hypothesized that using NLP to explore social media could help with assessing the mental health conditions of people experiencing insomnia after the outbreak of the COVID-19 pandemic. Mental health was defined by measuring negative sentiment, using NLP algorithms on publicly available data from Twitter. We designed a sentiment analysis pipeline based on pretrained transformers' architectures. The output of transformers was combined via Dempster-Shafer theory (DST; theory of belief) to achieve higher accuracy in the recognition of sentiments. The performance of this model was verified for accuracy by using a manually annotated data set. Subsequently, using this pipeline, we analyzed and compared the sentiments inherent in insomnia-related tweets that were posted within 1 year before the COVID-19 pandemic outbreak (prepandemic) and within 1 year during the pandemic (peripandemic). We also compared the results of the sentiment analysis of the tweets in terms of tweets' posting times (ie, temporal analysis; before midnight vs after midnight).

Methods

Study Design and Data Collection

This retrospective pilot study examined tweets that were posted in the 2019 calendar year (prepandemic interval) and the 2020 calendar year (peripandemic interval). We collected publicly available English tweets by using the Twitter application programming interface, which allowed us to collect tweets by matching keywords (ie, *insomnia*). The tweets were classified into two groups—prepandemic (January 1, 2019, to January 1, 2020) and peripandemic (January 1, 2020, to January 1, 2021) tweets—based on the posting dates and times. The inclusion criteria for tweets were that they must contain the word *insomnia* and be in English. Therefore, all non-English tweets and English tweets without the keyword *insomnia* were excluded (Figure 1). The data extracted from included tweets were used for sentiment analysis and for sentiment annotation.

Figure 1. STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) diagram.



Sampling Strategy and Annotation

To determine the minimum required sample size for the NLP algorithm performance measurement, we used the exact power calculation method [27]. We assumed that for an effect size of 0.3, an α of .05, a power of 80, and 5 *df*, 143 notes would be required. However, our team of annotators reviewed 300 randomly selected notes.

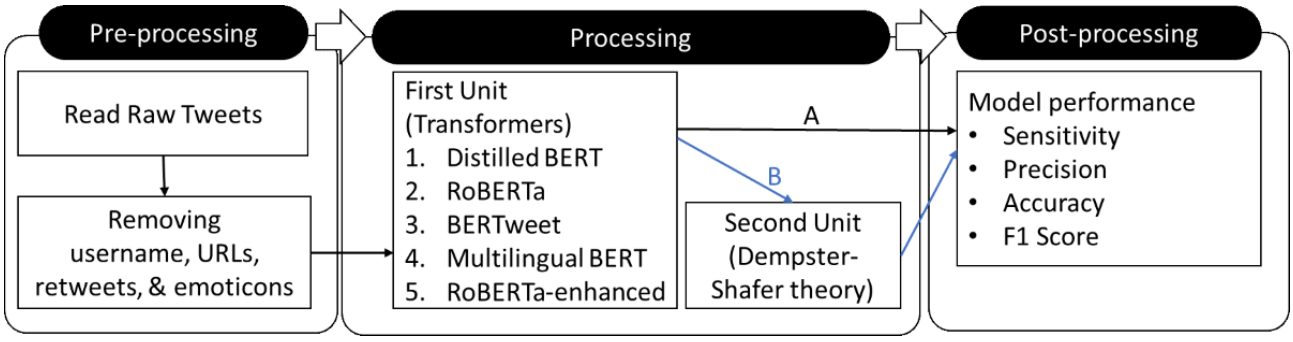
To verify the performance of the models in predicting the tweets' sentiments, we randomly chose 300 tweets from the data extracted (according to the *Study Design and Data Collection* section) and manually annotated them into the positive, negative, and neutral categories. Two nonnative English speakers with International English Language Testing System scores of ≥7 annotated the tweets. A third senior nonnative English speaker served as a final judge to adjudicate disagreements. We used the Cohen κ [28] parameter to measure the interrater reliability between annotators.

Developing a Sentiment Analysis Pipeline for Tweets

Sentiment Analysis Pipeline Overview

We devised an algorithm that had the following three steps: preprocess, process, and postprocess. In the preprocess step, we prepared the tweets for the process step by removing special characters, URLs, and hashtags. The process step consisted of 2 units. The first unit performed sentiment classification (ie, positive, negative, and neutral), using multiple models. The second unit used DST to combine the output from several models (ie, those from the previous step) to provide a more accurate prediction. Finally, in the postprocess step, we quantified the sentiment analysis performance of different models. These steps are discussed in more detail in the following sections and in Figure 2.

Figure 2. The machine learning natural language processing algorithm pipeline. (A) We calculated the performance of each transformer separately. (B) The output of transformers was combined, using the Dempster-Shafer theory to make the final decision. BERT: Bidirectional Encoder Representations From Transformers; RoBERTa: Robustly Optimized Bidirectional Encoder Representations From Transformers Pretraining Approach.



Preprocessing

Raw data scraped from Twitter contain irrelevant attributes (eg, usernames, URLs, retweets, emoticons, etc). The purpose of preprocessing was to filter undesired text content and obtain relevant parts of the tweets.

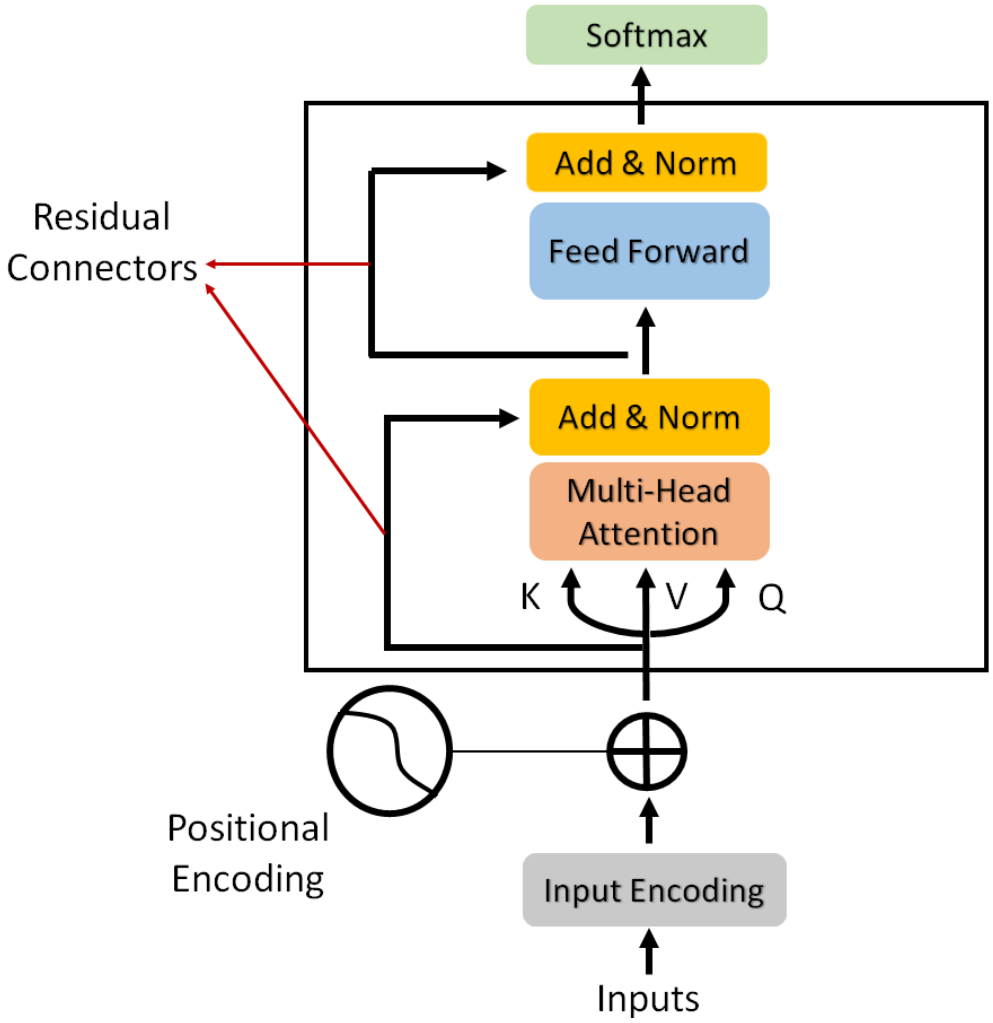
Process

The process step consisted of the following two units: NLP-based sentiment analysis classifiers and DST, which was used to combine the classifiers' outputs.

First Unit: Transformers

To perform the sentiment analysis on tweets, we took advantage of transformers, which are the new generation of deep artificial neural networks (also known as *recurrent neural networks*) that were introduced for machine translation [29] and were constructed by stacking transformer units on top of each other. They comprise two main blocks—an encoder and a decoder. The encoder is used for classification and inference, and the decoder is mainly used for language modeling; the complete architecture is used for machine translation [30]. A typical encoder of a transformer is shown in Figure 3 (Multimedia Appendix 1 provides a brief theory of transformers).

Figure 3. Classification procedure with a transformer.



A total of 5 different pretrained transformer-based models for the sentiment analysis of tweets were used. The five pretrained models provided by the Hugging Face AI community are as follows:

1. Distilled Bidirectional Encoder Representations From Transformers (BERT) [31], which was fine-tuned on the Stanford Sentiment Treebank v2 database [32]. Knowledge distillation [33,34] was used to reduce the size of a BERT model by 40% while preserving 97% of its language understanding capabilities and making it 60% faster.
2. Robustly Optimized BERT Pretraining Approach (RoBERTa) [35] for sentiment analysis, which was trained on around 58 million tweets. The RoBERTa model was based on the BERT structure; however, it was pretrained on not only the data that BERT was trained on (BookCorpus [34,36] and English Wikipedia; around 3.3 billion words) but also a news data and stories database [37]. RoBERTa was fine-tuned on 58 million tweets for sentiment analysis.
3. BERTweet [38], which was trained based on the RoBERTa pretraining procedure and pretrained on 850 million English tweets.
4. The multilingual BERT-based model, which was fine-tuned for sentiment analysis on product reviews in the following six languages: English, Dutch, German, French, Spanish, and Italian. It predicts the sentiment of a review by using stars (between 1 and 5 stars); 3 stars are considered neutral, <3 are considered negative, and ≥4 are considered positive.
5. The RoBERTa [35] model that was fine-tuned on 15 data sets from diverse text sources to enhance generalization across different types of texts (reviews, tweets, etc).

Second Unit: DST

To increase the performance of the transformer models discussed in the *First Unit: Transformers* section, we used DST [39,40], which has the ability to combine evidence from different experts. We let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_l\}$ be a finite set of possible hypotheses. This set is referred to as the *frame of discernment*, and its powerset is 2^{Θ}. We defined a function, *m(.)*, called a *basic belief assignment*, which maps every subset η of Θ to a value ranging from 0 to 1 and satisfies the following conditions:

$$m(\varphi)=0 \quad \textbf{(1)}$$

and

A subset ζ for which $m(\eta)$ is >0 is called a *focal element*. We defined another function called *the belief function*, *bel(.)*, which assigns a value ranging from 0 to 1 to every nonempty subset ζ of Θ and is defined as follows:

Given the above functions, we defined the combination rule. We assumed 2 basic belief assignments, $m_1(.)$ and $m_2(.)$, for belief functions $bel_1(.)$ and $bel_2(.)$ and let η_j and ζ_k be focal elements of bel_1 and bel_2, respectively. $m_1(.)$ and $m_2(.)$ were then combined to obtain the belief mass committed to $\vartheta \subseteq \Theta$, according to the following combination (ie, orthogonal sum formula):

where the denominator is essential for normalization.

Postprocess: Model Evaluation

To evaluate the performance of the models discussed in the *First Unit: Transformers* section, evaluation metrics—sensitivity, precision, accuracy, and F_1 score—extracted from the confusion matrix were used in this study and were calculated by using the following equations [41]:

Statistical Analysis

After performing the sentiment analysis and dividing the data into the negative, positive, and neutral categories, the categorical characteristics (number of negative, positive, and neutral tweets) of these tweets were analyzed by using the chi-square test and odds ratios (ORs). *P* values with a significance level of <.05, 95% CIs, and *z*-statistics were reported. Data management was performed with Python 3.8 [42], and the analysis was performed with SPSS version 27 (IBM Corporation).

Temporal Analysis

We also investigated the chronology of insomnia-related tweets by examining the overall hourly number of tweets. We extracted the posting times of tweets with a negative sentiment. The daily hours were then categorized into the following two time spans: before midnight (1 PM to midnight) and after midnight (1 AM to noon). We calculated the percentage of negative tweets in each interval and used a logistic regression analysis to compare the odds of posting negative tweets before and after midnight.

Results

Characteristics of Tweets

We retrieved 305,321 tweets that contained the word *insomnia* and were posted in the prepandemic and peripandemic periods. Of these, 139,561 were posted in the prepandemic period, and 165,760 (an 18.7% increase) were posted in the peripandemic interval. The tweets' length (number of words) was approximately the same between these two time periods (prepandemic: mean 26.3, SD 13.7 words; peripandemic: mean 29.3, SD 13.7 words). The number of tweet interactions, defined as the summation of the number of likes, retweets, and replies, did not differ significantly (P<001) (prepandemic: mean 6.2, SD 171.8 interactions; peripandemic: mean 5.4, SD 100.6 interactions).

Annotation

Of the 300 tweets that were annotated by the two reviewers, 167 (55.7%) were classified as negative, 102 (34%) were classified as neutral, and 31 (10.3%) were classified as positive. The interrater reliability reached 0.55 (95% CI 0.44-0.69).

Sentiment Analysis Pipeline Performance

In Table 1, we report the accuracy of the five models that were pretrained on 300 annotated tweets. Model 1—Distilled BERT—had the best performance (80.3%). After combining the models by using the DST approach, we observed that combining models 1, 2, 3, and 5 resulted in the highest performance (84%; Table 1).

Since Distilled BERT (model 1) showed the best performance for single-model classification, and to better understand how DST improves the performance of the pipeline, we analyzed the evaluation metrics of this model alongside those of the best combination of models (ie, the one reported in Table 1), which showed overall better performance for all 3 categories of sentiments (Table 2).

Table 1. Comparison of the performance of the models used to analyze the 300 annotated tweets.

Models	Accuracy (%)
Individual models	
Model 1 (Distilled BERT[a]) [31]	80.3
Model 2 (RoBERTa[b]) [35]	52.7
Model 3 (BERTweet[c]) [38]	53
Model 4 (BERT-multilingual) [35]	49.3
Model 5 (fine-tuned RoBERTa) [35]	45.3
Combined models based on Dempster-Shafer theory [39,40]	
Model 1+model 2+model 3	81
Model 1+model 2+model 3+model 5	84
Model 1+model 5	77.2
Model 1+model 2+model 3+model 4+model 5	81.7

[a]BERT: Bidirectional Encoder Representations From Transformers [31].

[b]RoBERTa: Robustly Optimized Bidirectional Encoder Representations From Transformers Pretraining Approach [35].

[c]BERTweet is a Robustly Optimized Bidirectional Encoder Representations From Transformers Pretraining Approach model that was trained on 850 million English tweets [38].

Table 2. Comparison of the performance of the individual model—Distilled Bidirectional Encoder Representations From Transformers—and the combined model based on Dempster-Shafer theory in identifying each sentiment class (positive, neutral, and negative).

Sentiment	Sensitivity (%)		Precision (%)		F_1 score		Accuracy (%)	
	Individual model[a]	Combined model[b]	Individual model	Combined model	Individual model	Combined model	Individual model	Combined model
Negative	92.8	93.4	77.9	81.7	84.7	87.1	81.3	84.6
Neutral	72.5	77.5	98.7	98.8	83.6	86.8	90.3	91.3
Positive	38.7	54.8	46.2	58.6	42.1	56.6	89	92

[a]The individual model is Distilled Bidirectional Encoder Representations From Transformers [31].

[b]The combined model is the combination of Distilled Bidirectional Encoder Representations From Transformers (BERT) [31], Robustly Optimized BERT Pretraining Approach (RoBERTa) [35], BERTweet [38], and fine-tuned RoBERTa [35].

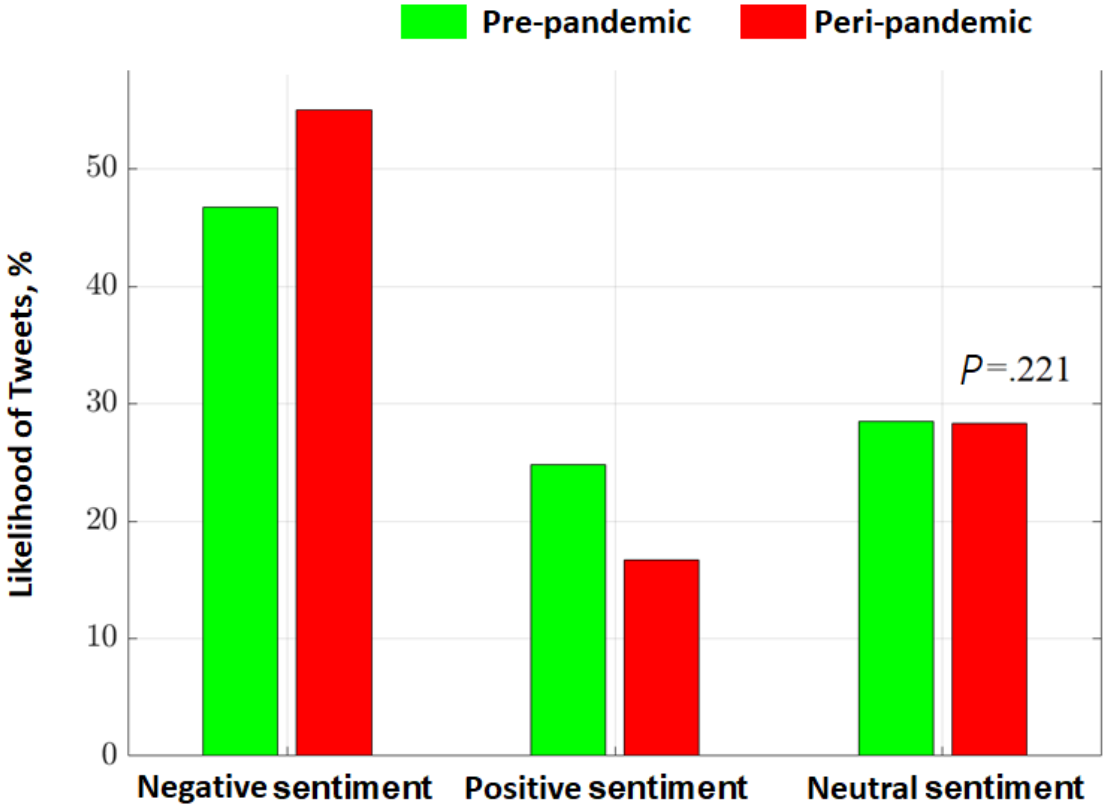
Sentiment Analysis

The results of the best combined model for sentiment analysis that was applied to all of the tweets are shown in Table 3. We observed a higher likelihood of posting negative tweets during the peripandemic period (91,242/165,760, 55%) compared to that during the prepandemic period (65,164/139,561, 46.7%). Accordingly, we observed a lower likelihood of posting positive tweets during the peripandemic period (27,621/165,760, 16.7%) compared to that during the prepandemic period (34,633/139,561, 24.8%). We also observed the same likelihood of posting neutral tweets during the peripandemic and postpandemic periods (Figure 4). We reported 39% higher odds of posting negative tweets during the peripandemic period compared to those during the prepandemic interval (OR, 1.39; 95% CI, 1.37-1.41, P<.001; Table 3).

Table 3. Characteristics of negative and positive prepandemic (calendar year 2019) tweets and peripandemic (calendar year 2020) tweets.

Tweet sentiment	Total tweets (N=305,321), n (%)	Prepandemic tweets (n=139,561), n (%)	Peripandemic tweets (n=165,760), n (%)	Prepandemic vs peripandemic		
				P value	*z*-statistic	Odds ratio (95% CI)
Negative tweets	156,406 (51.3)	65,164 (46.7)	91,242 (55)	<.001	45.94	1.39 (1.37-1.41)
Positive tweets	62,254 (20.4)	34,633 (24.8)	27,621 (16.7)	<.001	55.402	0.60 (0.59-0.61)
Neutral tweets	86,661 (28.3)	39,764 (28.5)	46,897 (28.3)	.22	1.22	0.99 (0.97-1.00)

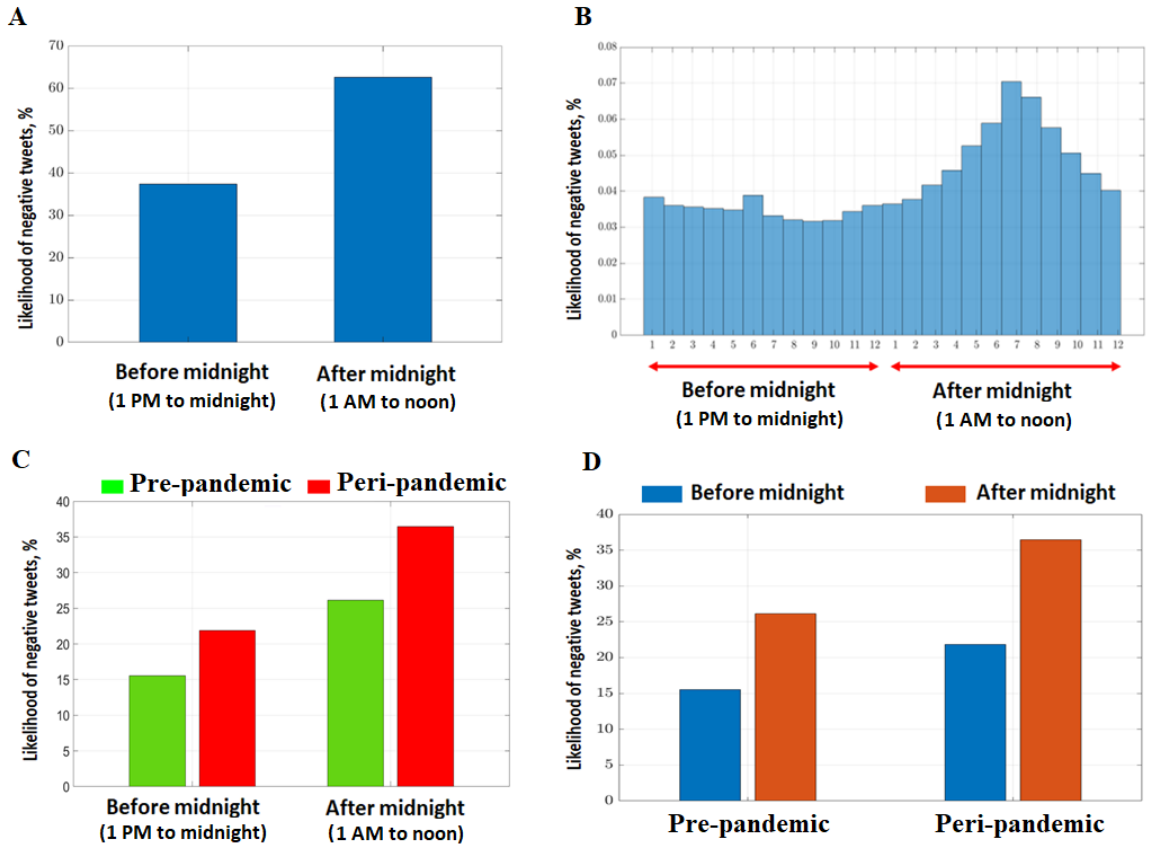
Figure 4. Likelihood of posting negative, positive, and neutral tweets in the prepandemic and peripandemic periods. *P<.001.



Temporal Analysis

The likelihood of posting negative tweets after midnight was higher than that before midnight (OR 1.21, 95% CI 1.19-1.23; P<.001; Figure 5A). An increasing trend was observed during after-midnight intervals when compared to before-midnight intervals, according to the hourly distribution of negative tweets (Figure 5B). The odds of posting negative tweets before midnight during the peripandemic period were 15% higher than those during the prepandemic period (OR 1.15, 95% CI 1.12-1.18; Figure 5C), while the odds posting negative tweets after midnight was 60% higher during the peripandemic period (OR 1.60; 95% CI 1.57-1.63; P<.001; Figure 5C). In the prepandemic period, the odds of posting negative tweets were 2% higher after midnight compared to those before midnight (OR 1.02, 95% CI 1.00-1.07; P=.008; Figure 5D); however, they were 43% higher in the peripandemic period (OR 1.43, 95% CI 1.40-1.46; P<.001; Figure 5D). The results of a quarterly (3-month) analysis of tweet sentiments for the prepandemic and peripandemic intervals are presented in Table S1 and Figure S2 in Multimedia Appendix 2.

Figure 5. Temporal analysis of tweets. (A) Percentage of negative tweets posted before midnight (1 PM to midnight) and after midnight (1 AM to noon). (B) Hourly distribution of negative tweets. (C) Comparison of the likelihood of posting negative tweets before midnight (1 PM to midnight) and after midnight (1 AM to noon) for the prepandemic and peripandemic periods. (D) Comparison of the likelihood of posting negative tweets before midnight (1 PM to midnight) and after midnight (1 AM to noon) for the prepandemic and peripandemic periods.



Discussion

Principal Findings

In this retrospective cohort study, we showed that NLP tools can monitor population health by using the sentiments expressed on a publicly available platform, such as Twitter, as a surrogate measure of public awareness and perception. We observed that the COVID-19 pandemic was negatively associated with a change in insomnia-related self-report tweets. We designed a novel NLP pipeline for sentiment analysis that was based on a combination of pretrained transformers (combined via DST; ie, theory of belief). By using this basis, which was validated on manually annotated tweets, we detected more negative tweets during the peripandemic interval than those detected during the prepandemic interval among people reporting insomnia on Twitter.

First, we developed a novel machine learning–based pipeline to analyze emotions. To verify the performance of models, we manually annotated 300 tweets. The κ analysis showed an agreement of 55% among different raters. This is not a very strong agreement, and this could have resulted from the inherent subjectivity of sentiment analysis tasks, in which everyone assigns a sentiment to a text according to their perspectives [43]. Next, using this annotated database, we verified the performance of each model individually and analyzed the performance of all of the models; Distilled BERT (model 1) performed the best, reaching an accuracy of 80.3%. In addition, the combined model yielded the best results (84% accuracy). It is worthy to note that the addition of RoBERTa (model 2) and BERTweet (model 3) did not improve the accuracy by much, but the addition of fine-tuned RoBERTa (model 5) resulted in a 4% increase in accuracy. Although the overall performance of fine-tuned RoBERTa (model 5) was lower than that of Distilled BERT (model 1), it had higher accuracy (71%) in detecting positive tweets than Distilled BERT (model 1; accuracy: 38.7%; confusion matrices are found in Figure S1 in Multimedia Appendix 2). Therefore, the combined model had superior accuracy in detecting positive tweets (54.8%) compared to Distilled BERT (model 1). Furthermore, based on Table 1, it can be deduced that keeping RoBERTa (model 2) and BERTweet (model 3) in the combination is necessary because the combination of Distilled BERT (model 1) and fine-tuned RoBERTa (model 5) yielded worse results (77.2%). This could be explained by the fact that while fine-tuned RoBERTa (model 5) had better performance in recognizing positive tweets, its performance in recognizing neutral and negative tweets was not very promising; thus, it reduced the overall accuracy. This shows the efficiency of DST in combining the models and exploiting the strength of each model to improve the overall classification of sentiments.

Having developed a reliable pipeline for sentiment analysis, we analyzed the emotions of tweets. During the peripandemic interval, we observed a significantly higher number of tweets with the keyword *insomnia* (P<001). A possible explanation is that social interactions shifted from in-person environments to

web-based environments, such as Twitter. The number of Twitter's annual users increased by 33.8%, from 138 million users in 2019 to 186 million users in 2020 [44,45]. We also observed a rise in the total number of insomnia-related tweets after the pandemic began. Considering this, in conjunction with the results of the sentiment analysis, we believe this spike could be related to the rise in negative tweets (Figure 4). According to Table 3, while there was an 8.1% decrease in the number of positive tweets related to insomnia, this number was overshadowed by an 8.3% spike in the number of negative tweets; the number of neutral tweets did not change meaningfully (0.2% decrease). Our findings on the significant increase in the number of negative tweets (P<001) during the pandemic is consistent with previously published literature [46]. Politis et al [47] showed an increase in negative sentiment on certain dates by analyzing tweets that were posted before and after the outbreak of the COVID-19 pandemic.

A previous study by Nota and Coles [48] showed that individuals experiencing sleep disruption exhibited diminished top-down inhibitory processes for controlling negative emotions and often engaged in repetitive negative thinking (rumination). We observed the same trend in our study; individuals with insomnia were more prone to rumination when they were awake and free from distractions at night (Figure 5B), suggesting a state of frustration after a poor night of sleep. This corresponds with the observation from Figure 5A, which shows that 62.4% (190,521/305,321) of the negative tweets were generated after midnight.

Our study showed that NLP tools can be used to monitor people's attitudes toward public stress, such as stress resulting from a pandemic. Policy makers and public health authorities may benefit from using such surveillance tools to better advocate for constituents [49]. Our study is classified as an infodemiology study, which offers an opportunity to analyze public sentiment in real time [50]. NLP tools are strong tools for analyzing and mining Twitter, which is a source of soft intelligence.

Limitations

In this study, we used Twitter as the source of data collection. As such, we might have excluded a large population that uses other social media platforms (eg, Facebook) or discussion forums (eg, Reddit) to express their perceptions about insomnia. Future studies should investigate publicly available data on other social media platforms in addition to those on Twitter. Further, as this study was based on tweets, it lacks validity measures, as no questionnaires or self-reported measures were used. A future study could use Twitter data and self-reported measures for individuals, health professionals, researchers, and nonprofit organizations in conjunction to assess the needs of pregnant women and the perceived available support and resources during the COVID-19 pandemic.

Of note, in this work, only the keyword *insomnia* was used to scrape the tweets. Although synonyms such as *sleeplessness* could have been used, we were interested only in the clinical term *insomnia*. A study that captures data on the broader area of sleep (ie, beyond insomnia) would be useful for further understanding the full effect of the pandemic. Additionally, several possible confounding factors, such as user location, were not available for all users; such factors may hinder the effect of geolocation on perceptions of insomnia.

Conclusion

In this study, we proposed a novel NLP pipeline that was based on a combination of transformers using DST to predict the sentiments inherent in text data. We manually annotated 300 tweets and combined various transformer architectures via DST. This combination resulted in higher accuracy for sentiment analysis. By using this pipeline on insomnia-related tweets, our study showed the negative effect of the COVID-19 pandemic on individuals' experiences of reporting insomnia on Twitter. To investigate the changes in Twitter users' reported sleep behaviors in the context of the COVID-19 pandemic, we analyzed tweets about insomnia that were posted before and during the pandemic (2019 and 2020). A strength of this study was using NLP and DST to identify tweets about insomnia and analyze their sentiments. In the future, we will assess the effects of changes in other aspects of mental health states (eg, boredom, fear, disgust, surprise, etc) and lifestyle changes (eg, changes in sleep duration, sleep schedules, substance use, physical activity, and sleep medication use) on insomnia symptoms during and after the pandemic based on Twitter and other social media platforms.

Acknowledgments

This work was supported by grants from National Institute of Health (NIH), National Heart, Lung, and Blood Institute K25 funding (grant 1K25HL152006-01; principal investigator: JR) and from NIH National Institute of Nursing Research funding (grant R01NR018342; principal investigator: SN). This study used the facilities and resources of the Center of Innovations in Quality, Effectiveness and Safety (grant number CIN 13-413). The opinions expressed are those of the authors and not necessarily those of the Department of Veterans Affairs, the US government, or Baylor College of Medicine.

Conflicts of Interest

HX and the University of Texas Health Science Center at Houston have research-related financial interest at Melax Technologies Inc.

Multimedia Appendix 1
A brief theory of transformers.
[DOCX File , 14 KB - jmir_v24i12e41517_app1.docx]

Multimedia Appendix 2
Supplementary tables and figures.
[DOCX File , 180 KB - jmir_v24i12e41517_app2.docx]

References

1. de Figueiredo CS, Sandre PC, Portugal LCL, Mázala-de-Oliveira T, da Silva Chagas L, Raony Í, et al. COVID-19 pandemic impact on children and adolescents' mental health: Biological, environmental, and social factors. Prog Neuropsychopharmacol Biol Psychiatry 2021 Mar 02;106:110171 [FREE Full text] [doi: 10.1016/j.pnpbp.2020.110171] [Medline: 33186638]
2. Lebrasseur A, Fortin-Bédard N, Lettre J, Raymond E, Bussières EL, Lapierre N, et al. Impact of the COVID-19 pandemic on older adults: Rapid review. JMIR Aging 2021 Apr 12;4(2):e26474 [FREE Full text] [doi: 10.2196/26474] [Medline: 33720839]
3. Elbogen EB, Lanier M, Blakey SM, Wagner HR, Tsai J. Suicidal ideation and thoughts of self-harm during the COVID-19 pandemic: The role of COVID-19-related stress, social isolation, and financial strain. Depress Anxiety 2021 May 05;38(7):739-748 [FREE Full text] [doi: 10.1002/da.23162] [Medline: 33949747]
4. Cellini N, Canale N, Mioni G, Costa S. Changes in sleep pattern, sense of time and digital media use during COVID-19 lockdown in Italy. J Sleep Res 2020 Aug;29(4):e13074 [FREE Full text] [doi: 10.1111/jsr.13074] [Medline: 32410272]
5. Partinen M, Bjorvatn B, Holzinger B, Chung F, Penzel T, Espie CA, ICOSS-collaboration group. Sleep and circadian problems during the coronavirus disease 2019 (COVID-19) pandemic: the International COVID-19 Sleep Study (ICOSS). J Sleep Res 2021 Feb;30(1):e13206 [FREE Full text] [doi: 10.1111/jsr.13206] [Medline: 33179820]
6. Partinen M, Kronholm E. Epidemiology: Principles and application in sleep medicine. In: Chokroverty S, editor. Sleep Disorders Medicine: Basic Science, Technical Considerations and Clinical Aspects. New York, NY: Springer; 2017:485-521.
7. Goel A, Gupta L. Social media in the times of COVID-19. J Clin Rheumatol 2020 Sep;26(6):220-223 [FREE Full text] [doi: 10.1097/RHU.0000000000001508] [Medline: 32852927]
8. Feldkamp J. The rise of TikTok: The evolution of a social media platform during COVID-19. In: Hovestadt C, Recker J, Richter J, Werder K, editors. Digital Responses to Covid-19: Digital Innovation, Transformation, and Entrepreneurship During Pandemic Outbreaks. Cham, Switzerland: Springer; 2021:73-85.
9. Colley RC, Bushnik T, Langlois K. Exercise and screen time during the COVID-19 pandemic. Health Rep 2020 Jul 15;31(6):3-11 [FREE Full text] [doi: 10.25318/82-003-x202000600001-eng] [Medline: 32672923]
10. Beck F, Léger D, Fressard L, Peretti-Watel P, Verger P, Coconel Group. Covid-19 health crisis and lockdown associated with high level of sleep complaints and hypnotic uptake at the population level. J Sleep Res 2021 Feb;30(1):e13119 [FREE Full text] [doi: 10.1111/jsr.13119] [Medline: 32596936]
11. de Azevedo Barros MB, Lima MG, Malta DC, Szwarcwald CL, de Azevedo RCS, Romero D, et al. Report on sadness/depression, nervousness/anxiety and sleep problems in the Brazilian adult population during the COVID-19 pandemic. Epidemiol Serv Saude 2020;29(4):e2020427 [FREE Full text] [doi: 10.1590/s1679-49742020000400018] [Medline: 32844918]
12. Zhou SJ, Wang LL, Yang R, Yang XJ, Zhang LG, Guo ZC, et al. Sleep problems among Chinese adolescents and young adults during the coronavirus-2019 pandemic. Sleep Med 2020 Oct;74:39-47 [FREE Full text] [doi: 10.1016/j.sleep.2020.06.001] [Medline: 32836185]
13. Idrissi AJ, Lamkaddem A, Benouajjit A, El Bouaazzaoui MB, El Houari F, Alami M, et al. Sleep quality and mental health in the context of COVID-19 pandemic and lockdown in Morocco. Sleep Med 2020 Oct;74:248-253 [FREE Full text] [doi: 10.1016/j.sleep.2020.07.045] [Medline: 32862008]
14. Passos L, Prazeres F, Teixeira A, Martins C. Impact on mental health due to COVID-19 pandemic: Cross-sectional study in Portugal and Brazil. Int J Environ Res Public Health 2020 Sep 17;17(18):6794 [FREE Full text] [doi: 10.3390/ijerph17186794] [Medline: 32957702]
15. Mautong H, Gallardo-Rumbea JA, Alvarado-Villa GE, Fernández-Cadena JC, Andrade-Molina D, Orellana-Román CE, et al. Assessment of depression, anxiety and stress levels in the Ecuadorian general population during social isolation due to the COVID-19 outbreak: a cross-sectional study. BMC Psychiatry 2021 Apr 28;21(1):212 [FREE Full text] [doi: 10.1186/s12888-021-03214-1] [Medline: 33910550]
16. Gualano MR, Lo Moro G, Voglino G, Bert F, Siliquini R. Effects of Covid-19 lockdown on mental health and sleep disturbances in Italy. Int J Environ Res Public Health 2020 Jul 02;17(13):4779 [FREE Full text] [doi: 10.3390/ijerph17134779] [Medline: 32630821]
17. Alkhamees AA, Alrashed SA, Alzunaydi AA, Almohimeed AS, Aljohani MS. The psychological impact of COVID-19 pandemic on the general population of Saudi Arabia. Compr Psychiatry 2020 Oct;102:152192 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152192] [Medline: 32688022]
18. Talbot J, Charron V, Konkle AT. Feeling the void: Lack of support for isolation and sleep difficulties in pregnant women during the COVID-19 pandemic revealed by Twitter data analysis. Int J Environ Res Public Health 2021 Jan 06;18(2):393 [FREE Full text] [doi: 10.3390/ijerph18020393] [Medline: 33419145]

19. Chaudhry HN, Javed Y, Kulsoom F, Mehmood Z, Khan ZI, Shoaib U, et al. Sentiment analysis of before and after elections: Twitter data of U.S. election 2020. Electronics (Basel) 2021 Aug 27;10(17):2082 [FREE Full text] [doi: 10.3390/electronics10172082]
20. He J, Mai S, Hu H. A unimodal reinforced transformer with time squeeze fusion for multimodal sentiment analysis. IEEE Signal Process Lett 2021 May 07;28:992-996. [doi: 10.1109/lsp.2021.3078074]
21. Jalil Z, Abbasi A, Javed AR, Khan MB, Hasanat MHA, Malik KM, et al. COVID-19 related sentiment analysis using state-of-the-art machine learning and deep learning techniques. Front Public Health 2022 Jan 14;9:812735 [FREE Full text] [doi: 10.3389/fpubh.2021.812735] [Medline: 35096755]
22. Jiang M, Wu J, Shi X, Zhang M. Transformer based memory network for sentiment analysis of web comments. IEEE Access 2019 Dec 02;7:179942-179953 [FREE Full text] [doi: 10.1109/access.2019.2957192]
23. Naseem U, Razzak I, Khushi M, Eklund PW, Kim J. COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. IEEE Trans Comput Soc Syst 2021 Jan 29;8(4):1003-1015 [FREE Full text] [doi: 10.1109/TCSS.2021.3051189] [Medline: 35783149]
24. Naseem U, Razzak I, Musial K, Imran M. Transformer based deep intelligent contextual embedding for Twitter sentiment analysis. Future Gener Comput Syst 2020 Dec;113:58-69 [FREE Full text] [doi: 10.1016/j.future.2020.06.050]
25. Sousa MG, Sakiyama K, de Souza Rodrigues L, Moraes PH, Fernandes ER, Matsubara ET. BERT for stock market sentiment analysis. 2020 Presented at: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI); November 4-6, 2019; Portland, OR. [doi: 10.1109/ictai.2019.00231]
26. Wang T, Lu K, Chow KP, Zhu Q. COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model. IEEE Access 2020 Jul 28;8:138162-138169 [FREE Full text] [doi: 10.1109/ACCESS.2020.3012595] [Medline: 34812342]
27. Kang H. Sample size determination and power analysis using the G*Power software. J Educ Eval Health Prof 2021;18:17 [FREE Full text] [doi: 10.3352/jeehp.2021.18.17] [Medline: 34325496]
28. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37-46. [doi: 10.1177/001316446002000104]
29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. 2017 Presented at: 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA p. 6000-6010 URL: https://dl.acm.org/doi/pdf/10.5555/3295222.3295349
30. Tay Y, Dehghani M, Bahri D, Metzler D. Efficient transformers: A survey. arXiv Preprint posted online on September 14, 2020. [FREE Full text]
31. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online on October 11, 2018. [FREE Full text]
32. Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng A, et al. Recursive deep models for semantic compositionality over a sentiment treebank. 2013 Presented at: 2013 Conference on Empirical Methods in Natural Language Processing; October 18-21, 2013; Seattle, WA p. 1631-1642 URL: https://aclanthology.org/D13-1170.pdf
33. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv Preprint posted online on March 9, 2015. [FREE Full text]
34. Buciluǎ C, Caruana R, Niculescu-Mizil A. Model compression. 2006 Aug Presented at: KDD06: The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 20-23, 2006; Philadelphia, PA p. 535-541. [doi: 10.1145/1150402.1150464]
35. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A robustly optimized BERT pretraining approach. arXiv Preprint posted online on July 26, 2019. [FREE Full text]
36. Zhu Y, Kiros R, Zemel R, Salakhutdinov R, Urtasun R, Torralba A, et al. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. 2016 Presented at: 2015 IEEE International Conference on Computer Vision (ICCV); December 7-13, 2015; Santiago, Chile. [doi: 10.1109/iccv.2015.11]
37. Trinh TH, Le QV. A simple method for commonsense reasoning. arXiv Preprint posted online on June 7, 2018. [FREE Full text]
38. Nguyen DQ, Vu T, Nguyen AT. BERTweet: A pre-trained language model for English tweets. arXiv Preprint posted online on October 5, 2020. [FREE Full text]
39. Dempster AP. Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics 1967 Apr;38(2):325-339. [doi: 10.1214/aoms/1177698950]
40. Shafer G. A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press; 1976.
41. Razjouyan J, Freytag J, Dindo L, Kiefer L, Odom E, Halaszynski J, et al. Measuring adoption of patient priorities-aligned care using natural language processing of electronic health records: Development and validation of the model. JMIR Med Inform 2021 Feb 19;9(2):e18756 [FREE Full text] [doi: 10.2196/18756] [Medline: 33605893]
42. Van Rossum G, Drake FL. Python/C Api Manual - Python 3: (Python Documentation Manual Part 4). Scotts Valley, CA: CreateSpace; 2009.

43. Wearne T, Osborne-Crowley K, Rosenberg H, Dethier M, McDonald S. Emotion recognition depends on subjective emotional experience and not on facial expressivity: evidence from traumatic brain injury. Brain Inj 2018 Oct 08;33(1):12-22. [doi: 10.1080/02699052.2018.1531300] [Medline: 30296178]
44. Iqbal M. Twitter revenue and usage statistics (2022). Business of Apps. URL: https://www.businessofapps.com/data/twitter-statistics/ [accessed 2022-11-24]
45. Haman M. The use of Twitter by state leaders and its impact on the public during the COVID-19 pandemic. Heliyon 2020 Nov 19;6(11):e05540 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e05540] [Medline: 33294685]
46. 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]
47. Politis I, Georgiadis G, Kopsacheilis A, Nikolaidou A, Papaioannou P. Capturing Twitter negativity pre- vs. mid-COVID-19 pandemic: An LDA application on London public transport system. Sustainability 2021 Dec 02;13(23):13356 [FREE Full text] [doi: 10.3390/su132313356]
48. Nota JA, Coles ME. Shorter sleep duration and longer sleep onset latency are related to difficulty disengaging attention from negative emotional images in individuals with elevated transdiagnostic repetitive negative thinking. J Behav Ther Exp Psychiatry 2018 Mar;58:114-122. [doi: 10.1016/j.jbtep.2017.10.003] [Medline: 29111422]
49. Tsai MH, Wang Y. Analyzing Twitter data to evaluate people's attitudes towards public health policies and events in the era of COVID-19. Int J Environ Res Public Health 2021 Jun 10;18(12):6272 [FREE Full text] [doi: 10.3390/ijerph18126272] [Medline: 34200576]
50. Marshall C, Lanyi K, Green R, Wilkins GC, Pearson F, Craig D. Using natural language processing to explore mental health insights from UK tweets during the COVID-19 pandemic: Infodemiology study. JMIR Infodemiology 2022 Mar 31;2(1):e32449 [FREE Full text] [doi: 10.2196/32449] [Medline: 36406146]

Abbreviations

BERT: Bidirectional Encoder Representations From Transformers
DST: Dempster-Shafer theory
NLP: natural language processing
OR: odds ratio
RoBERTa: Robustly Optimized Bidirectional Encoder Representations From Transformers Pretraining Approach

Edited by G Eysenbach; submitted 28.07.22; peer-reviewed by K Chow, N Bragazzi; comments to author 27.09.22; revised version received 20.10.22; accepted 15.11.22; published 27.12.22.

Please cite as:
Maghsoudi A, Nowakowski S, Agrawal R, Sharafkhaneh A, Kunik ME, Naik AD, Xu H, Razjouyan J
Sentiment Analysis of Insomnia-Related Tweets via a Combination of Transformers Using Dempster-Shafer Theory: Pre– and Peri–COVID-19 Pandemic Retrospective Study
J Med Internet Res 2022;24(12):e41517
URL: https://www.jmir.org/2022/12/e41517
doi:10.2196/41517
PMID:36417585

Review

User Experience of COVID-19 Chatbots: Scoping Review

Becky K White[1,2], PhD; Annegret Martin[2], Grad Dip; James Angus White[2], PhD

[1]School of Population Health, Curtin University, Perth, Australia
[2]Reach Health Promotion Innovations, Perth, Australia

Corresponding Author:
Becky K White, PhD
Reach Health Promotion Innovations
PO Box 372
Mount Hawthorn
Perth, 6915
Australia
Phone: 61 450169891
Email: becky@rhpi.com.au

Abstract

Background: The COVID-19 pandemic has had global impacts and caused some health systems to experience substantial pressure. The need for accurate health information has been felt widely. Chatbots have great potential to reach people with authoritative information, and a number of chatbots have been quickly developed to disseminate information about COVID-19. However, little is known about user experiences of and perspectives on these tools.

Objective: This study aimed to describe what is known about the user experience and user uptake of COVID-19 chatbots.

Methods: A scoping review was carried out in June 2021 using keywords to cover the literature concerning chatbots, user engagement, and COVID-19. The search strategy included databases covering health, communication, marketing, and the COVID-19 pandemic specifically, including MEDLINE Ovid, Embase, CINAHL, ACM Digital Library, Emerald, and EBSCO. Studies that assessed the design, marketing, and user features of COVID-19 chatbots or those that explored user perspectives and experience were included. We excluded papers that were not related to COVID-19; did not include any reporting on user perspectives, experience, or the general use of chatbot features or marketing; or where a version was not available in English. The authors independently screened results for inclusion, using both backward and forward citation checking of the included papers. A thematic analysis was carried out with the included papers.

Results: A total of 517 papers were sourced from the literature, and 10 were included in the final review. Our scoping review identified a number of factors impacting adoption and engagement including content, trust, digital ability, and acceptability. The papers included discussions about chatbots developed for COVID-19 screening and general COVID-19 information, as well as studies investigating user perceptions and opinions on COVID-19 chatbots.

Conclusions: The COVID-19 pandemic presented a unique and specific challenge for digital health interventions. Design and implementation were required at a rapid speed as digital health service adoption accelerated globally. Chatbots for COVID-19 have been developed quickly as the pandemic has challenged health systems. There is a need for more comprehensive and routine reporting of factors impacting adoption and engagement. This paper has shown both the potential of chatbots to reach users in an emergency and the need to better understand how users engage and what they want.

(J Med Internet Res 2022;24(12):e35903) doi:10.2196/35903

KEYWORDS

COVID-19; chatbot; engagement; user experience; pandemic; global health; pandemic; digital health; health information

Introduction

Background

On March 11, 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a pandemic [1]. Since then, the virus has had global impacts, and as of July 2021, there had been over 185 million confirmed cases of COVID-19 and over 4 million deaths [2]. The virus has at times overwhelmed the health systems of different countries, and the need for accurate and timely information has never been higher. The COVID-19 pandemic has seen users around the world turn to digital technology for the sourcing of health information [3]. Health

authorities such as WHO have responded with new and innovative ways of delivering trusted information, including the use of chatbots [4]. Chatbots are software systems that enable users to interact with a program as if they are talking to another person, often using machine learning to achieve the effect of intelligent response. Chatbots are being used across all areas of health to deliver information, promote behavior change, and deliver treatment. Chatbots can be deployed as stand-alone systems or via existing communication platforms such as WhatsApp or Facebook Messenger. This integration into familiar apps can result in a low barrier to entry for many users and a potential for substantial reach.

Chatbots can be rule-based or artificial intelligence–based. Rule-based chatbots use decision trees and defined rules to guide conversations, whereas artificial intelligence–based chatbots uses machine learning and natural language processing to generate and respond to dialogue [5]. As chatbots become more sophisticated, this opens up new and innovative ways to interact with and engage users in digital content. Features such as videos, quizzes, emoji, and style of voice are all used to keep users engaged. As these tools are used more, it is important to understand the marketing and communication strategies used to improve effectiveness and user experience. This enables the design of systems that people want to use and can benefit from.

Relevant Literature

Good user experience design is important for chatbot adoption and acceptance, and this requires expert input [6]. Digital solutions that are not designed with the user in mind can result in low engagement [7]. Working in a co-design model and involving end users throughout development can help in developing user-centered products and increasing engagement [7]. Chatbots have been developed for a wide range of health issues and behaviors, with mental health chatbots having the most dedicated literature [8]. A recent scoping review of user perspectives and opinions about mental health chatbots found that people generally found them easy to use and that enjoyment and trust are key mediators of interaction with chatbots [9]. However, the quality of conversation was identified as a limitation.

A review of technical metrics used to evaluate chatbots found a diversity in approaches and no apparent standardization [8]. The authors identified 27 metrics related to chatbots as a whole, response generation, response understanding, and aesthetics. They identified a range of ways to measure usability, such as a single question in a questionnaire, multiple questions, observation, or the use of a validated scale such as the System Usability Scale. In addition, they found that only 7% of studies included any assessment of aesthetics [8]. A scoping review of the features most commonly used in mental health chatbots showed that most were rule-based and stand-alone software [10]. In most chatbots, the conversations were led by the chatbot, and most included digital representations such as an avatar or digital human characters [10].

COVID-19 Chatbots

Chatbots can have special use in a pandemic for reaching people with information, supporting behavior change, providing mental health support, and identifying and monitoring symptoms [11]. Recognizing these opportunities, chatbots for COVID-19 have been quickly developed and scaled up by health authorities such as the WHO [12] and Centers for Disease Control and Prevention [13]. COVID-19 has overwhelmed health systems around the world [14], and the combination of increased need for health services and the need for social distancing has highlighted the potential for chatbots to ease some of the health system burden [15]. A number of studies describe guidance or frameworks for the design of chatbots for COVID-19 [15-17]. Despite the importance of assessing adoption and user engagement, they are not routinely reported in chatbot evaluations, and measures are not standardized [3,4]. The COVID-19 pandemic has presented a unique situation that has accelerated the need for the implementation of digital solutions. Although scoping reviews have previously examined COVID-19 chatbots in general [18], this is the first paper to our knowledge that focuses on this specific scenario. This paper sought to describe the current knowledge base about the user experience of COVID-19 chatbots.

Methods

Overview

We conducted a scoping review of the user experience of COVID-19–related chatbots, with regard to design, engagement, and communication features. We followed guidance from the Joanna Briggs Institute on conducting systematic scoping reviews [19] and reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for scoping reviews [20]. This guidance outlines the steps, from the development of a protocol and the search strategy to charting the results and reporting the findings.

Review Objective

We sought to provide an overview of the evidence of design, engagement, and communication features used in COVID-19 chatbots and the impact on user engagement, preference, and retention.

Review Questions

Our review questions were as follows:

- What are the best practice approaches to user engagement in COVID-19 chatbots?
- What is known about user experience, preference, and retention in relation to the different engagement strategies, content, and language features of COVID-19 chatbots?
- What marketing and communication strategies have been implemented and evaluated with COVID-19 chatbots?
- What are the gaps in the literature and recommendations for future research?

Search Strategy

We used a standard 3-step search strategy for scoping reviews.

1. Initial search of MEDLINE and Embase databases followed by an analysis of the keywords and results found
2. Second search using a revised keyword list across all databases

3. Backward and forward citation checking of all the included papers and articles

Search Sources

For this review, we searched databases from June 19-20, 2021. The search strategy was initially developed for MEDLINE and then adapted for other databases. We initially searched MEDLINE Ovid and Embase. Search results from these databases were reviewed, and the search strategy was adjusted slightly before being rerun. Other databases used in the search included CINAHL, ACM Digital Library, Cochrane COVID-19 study register, Emerald, Communication abstracts (EBSCO), and the WHO COVID-19 Global literature on coronavirus disease. To source unpublished literature, Google Scholar was searched using the adapted search terms and the first 100 results were scanned. Backward and forward citation searches of all the included articles were conducted, and additional articles identified were sourced individually.

Search Terms

The search terms for this review were first developed with previous knowledge from the authors and then further informed by the current literature about this topic. A comprehensive list of search terms was developed to ensure a broad range of articles. We used search terms related to the technological subject matter of interest (eg, chatbots, conversational agent, and dialogue system), user experience more generally (eg, engagement, features, and user experience), and the health issue at hand (eg, COVID, COVID-19, Corona, and coronavirus). The search terms used for each database can be viewed in Multimedia Appendix 1.

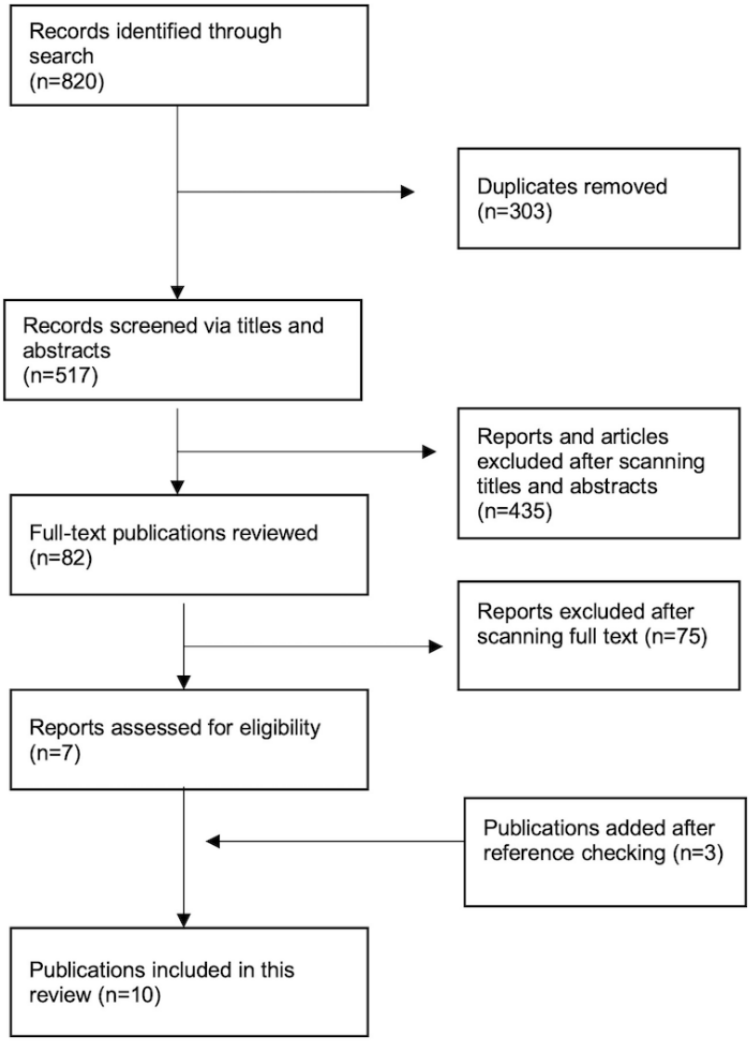
Inclusion and Exclusion Criteria

We included studies that assessed the design, marketing, and user features of COVID-19 chatbots or those that explored user perspectives on features, marketing, and design. As the aim was to map the existing evidence across published and unpublished literature, there was no criteria limitation with regard to geographical location, type of chatbots, or study design. We included papers, articles, and conference proceedings. Any articles or reports pertaining to chatbots that were not related to COVID-19 were excluded. We excluded papers that did not include any reporting on user perspectives, experience, or the general use of chatbot features or marketing, as well as papers where a version was not available in English. We excluded conference abstracts, books, and theses.

Study Selection Process

The selection of articles was defined by the inclusion and exclusion criteria. In line with the iterative nature of scoping reviews, these criteria were refined as the review progressed with alterations made to the criteria following the initial search of 2 databases. Deduplication was carried out in Endnote software (Clarivate), and then Rayyan was used for screening. Rayyan is a web-based program designed for managing the citation-screening process [21]. Authors BW and AM independently screened article titles and abstracts, excluding irrelevant studies. Any discrepancies were resolved via discussion. For articles deemed possibly relevant or where it was difficult to ascertain from the title and abstract, the full text was reviewed. Figure 1 shows the study selection process.

Figure 1. Search strategy and selection process.



Data Extraction and Charting

A data extraction form was developed to ensure the uniform collection of data, and this form was trialed and revised. Data were charted in categories of study characteristics, intervention characteristics, and outcomes. A thematic analysis was then used to synthesize the findings in each paper and identify themes.

Results

Included Studies

Of the 517 papers initially identified, 10 were included in the final analysis, with 2 papers reporting results from the same study population. The authors took a wide view of the inclusion criteria, including studies that included any reporting of the abovementioned criteria. The papers differed widely in their approach to reporting user experience, from empirical studies that evaluated user experience comprehensively to those that reported minimal findings. Several papers were excluded even though they had detailed their use experience plan because they did not report findings. Papers included discussions about chatbots developed for COVID-19 screening and general COVID-19 information, as well as studies investigating user perceptions and opinions on COVID-19 chatbots. A thematic analysis was carried out on the included studies to understand and organize the findings across the papers. Table 1 describes the characteristics of the included studies.

Table 1. Characteristics of included studies.

Author, year	Chatbot type	Country	Study design	Study aim	Key outcomes
Almalki [22], 2021	Chatbots for COVID-19	Saudi Arabia	Survey, distributed via social media and messaging apps	To explore chatbot use and associated challenges and barriers	• 40% were aware of chatbots, yet only 24% had used them before • 56.9% had positive perceptions • 84% expressed willingness to use in future
Almalki [23], 2020	Chatbots for COVID-19	Saudi Arabia	Survey, distributed via social media and messaging apps	To explore user perceptions of health chatbots in Saudi Arabia	• Overall positive perceptions • Users were more willing to use for general information about COVID-19 • Highly educated users were more likely to engage
Dennis et al [24], 2020	COVID-19–screening chatbot	United States	Web-based experiment comparing human or chatbot agents	To understand user response to COVID-19–screening chatbots	• User perception of agent ability (human or chatbots) was the primary factor • User trust in provider is an important factor
Judson et al [25], 2020	COVID-19–screening chatbot	United States	Descriptive process of development	To describe the development of screening chatbots for a hospital setting	• Users need to trust the authority the chatbot is coming from • Saved employee time
Liang et al [26], 2021	Chatbots to discuss movies and COVID-19	Not reported	Tested reactions and use with different levels of chatbot disclosure	To understand chatbot self-disclosure, user engagement, and perception	• Users' self-disclosure increased with chatbot's self-disclosure • Chatbots' self-disclosure also positively impacted engagement and users' perception
McKillop et al [27], 2021	COVID-19 chatbots using the Watson Assistant platform	Across 9 countries[a]	Descriptive study of the Watson Assistant platform as COVID-19 chatbots	To describe COVID-19 conversational agents built using the Watson Assistant	• The average number of conversational turns ranged from 1.9-3.5 • Clinical providers had the highest number of turns
Morse et al [28], 2020	COVID-19 symptom checker	United States	Descriptive study of completed assessments	To describe user demographics and levels of triage acuity provided by a symptom checker	• 30-39 years was most common age group, but a sizable minority were aged 60 years or over • Most users were female
Ollier et al [29], 2021	COVID-19 health promotion and health coaching	United Kingdom and Ireland	Descriptive analysis of the marketing of chatbot using Facebook advertisements	To understand the use of Facebook advertisements to drive chatbot use	• Static images have better conversions than carousel images • Android downloads were higher than iOS • Middle-aged older women were more engaged
Rodsawang et al [30], 2020	COVID-19 general information chatbot	Thailand	Descriptive analysis of chatbots development and user feedback	To describe characteristics and user perspectives with the "COVID-19 Preventable" chatbot	• Government COVID-19 chatbot with good user feedback and uptake • Menu feature became important as users and content increased
Schubel et al [31], 2021	COVID-19 symptom screener and learning module	United States	Survey with people invited to use chatbots over text messaging or email	To understand interaction and feature access by population subgroups	• Demographic differences observed, with women, African American individuals, and those aged 51-90 years interacting more

[a]The 9 countries were the United States, Australia, Canada, the United Kingdom, Pakistan, Germany, Russia, Ireland, and Singapore.

Thematic Analysis

The papers described a range of factors impacting on user experience and chatbot use. The COVID-19 pandemic expedited the development of chatbots with some tools being developed and deployed within 5 days [25,27]. Partnership with technical teams was described as being beneficial [25], with one paper describing the lack of technical expertise on their team as a limitation [30]. Findings are mapped by theme and described in the following section.

Factors Impacting User Experience and Use

Trust

Trust was explicitly reported in 4 papers [24-26,30]. A user's perception of the ability of a chatbot was a primary factor impacting interaction according to one paper, and that perception was impacted by the trust placed in the chatbot's provider [24]. Being clearly marked with the provider's branding was important, with one study reporting users being suspicious about using a tool that was not clearly branded to their hospital [25]. Users evaluating the "COVID-19 Preventable" chatbot recommended that the user interface should represent public health authorities [30]. As disclosure by one chatbot increased, so did the user's likelihood of self-disclosure [26]. Increased disclosure by the chatbots also positively affected the engagement and perceived warmth. When the chatbots showed emotional disclosure, engagement significantly increased [26].

Digital Literacy

Two papers mentioned digital literacy and ability. Almalki [23] reported that perceived IT skills or past use of chatbots did not impact perceptions of chatbots ability. They found that those who reported frequently searching for health information on the web were more likely to use health chatbots to source medical services [23]. Another paper reported that a user's perceived digital ability had a small-to-medium effect on satisfaction, motivation, likelihood of use, and adherence to advice [24].

Design and Usability

Four papers described the process of design of their chatbots, with a focus on enhancing usability. Two papers described a process of chatbot development that involved stages of defining, design, journey mapping, iteration, and evaluation [25,30]. Another paper examining disclosure with COVID-19 chatbots reported their process of conversation, including starting with small talk and then building up to recommendations [26]. Another paper highlighted the importance of ease of use in maintaining user engagement [30].

Demographic Factors

Four studies reported demographics with use and user experience. Schubel et al [31] reported that people aged 51-90 years were the most likely to use their chatbots. However, this differed by feature. Younger users (aged 18-50 years) were more likely to use a symptom screen checker, and older users (years 51-90 years) were more likely to use the learning module. This same study found a greater proportion of African American users than other races or ethnicities and more female users than male users [31].

Conversely, Morse et al [28] reported that although the most common users were aged 30-39 years, there was a sizable minority of older users, with 13.3% being aged 60 years or over. They also found that the majority of completed symptom checker assessments were carried out by female users. Ollier et al [29] reported that female users over the age of 35 years were downloading their chatbots at higher rates than any other group. This was in response to Facebook advertisements promoting the chatbots, and the authors noted that the content of those advertisements may have been more attractive to this group. Almalki [23] reported no significant difference in gender with any of their variables, but that participants aged under 30 years reported more enjoyment with using chatbots [22].

User Perspectives on Content and Features

Four papers reported findings related to chatbot content. One paper found that people were more likely to use a chatbot to seek general information about COVID-19 than information about medical treatments [23]. Another paper describing a chatbot with 2 feature components found no difference in the use of the health screener information compared to the general learning content [31]. McKillop et al [27] reported on the average number of conversational turns per session. These ranged from 1.9 to 3.5. Clinician providers had the highest number of turns, suggesting that these chatbots may have been delivering more complex content. The authors suggested that the lower number of conversational turns may be due to people asking more simple question at the start of the pandemic, which may have increased in complexity and, thus, turns as the pandemic continued. Indeed, they saw a marked increase in conversation turns among employees, possibly related to more complex questioning as they looked to return to work [27]. As the number of users increased with one COVID-19 chatbot, so did the range of questions and content, and thus, new systems were needed to manage content [30].

Managing content and daily updates of information during the pandemic was reported as being challenging and constant by the authors describing the "COVID-19 Preventable" chatbot [30]. Identifying user questions that chatbots were unable to answer was a useful way of prioritizing new information and conversation rules to be added. Another challenge was the daily task of translating complex content into deliverable bites for public consumption. Once the use of this chatbot reached over 100,000 users, the use of the menu significantly increased. Content was filtered into categories, and the menu reportedly helped users to navigate and find answers more easily [30]. Progressive disclosure was used to manage information flow from basic to more complex. Additional features were added to this chatbot based on user demand, such as the ability to report a concern about the potential spread of COVID-19 directly to the authorities [30].

Marketing

One paper detailed the experience of using Facebook advertisements to promote their chatbot. The authors used A/B testing to test *carousel images* with *static images and text*, finding that *static images with texts* were better received [29]. Android users downloaded and engaged with the chatbots more than iOS users. The authors reported that through their increased

engagement with the advertisements, Android users essentially marketed the product to their contacts. Women aged 35 years and over downloaded the chatbots the most. This group also had the highest engagement with the Facebook advertisements themselves, in terms of reacting to and sharing the post [29]. Two other papers, although not reporting specifically on promotion, reported that mass media campaigns, social role model endorsement, or national health authority adoption may help in raising awareness and trust in chatbots [22,23]

User Perspectives and Acceptability

The theme of acceptability included willingness to use COVID-19 chatbots, perceptions on chatbots in general, and adoption. Five papers reported some acceptability findings. An employee-screening chatbot deployed at a large hospital site found that adoption was aided by their chatbot not requiring download or a log-in [25]. Despite many of the respondents to one survey having little experience with chatbots, most (82.5%) reported being willing to use chatbots to seek general information about health care services and how to prevent the spread of COVID-19 [23]. Acceptability appeared to be higher for those who already used the internet to seek health information [22]. Participants reported generally having positive perceptions of health chatbots and being willing to use them. Empathy and emotional sensitively was an important factor in acceptability; participants perceived that chatbots had no emotions, which may impact acceptability. Social norms also impact chatbot acceptability, with participants reporting a tendency to align with the views of others [22].

The "COVID-19 Preventable" chatbot reported that most people (98%) stated that they were likely to continue using the chatbot and 96% said that they would recommend it to others [30]. This acceptability may have been aided by the 64% of people who reported that the chatbots answered their questions appropriately [30]. The perceived ability of the chatbot was rated as being important in Dennis et al [24], another study that compared the user reactions of chatbots with human conversations. Interestingly, when ability was perceived to be the same, chatbots were perceived more positively than humans. The authors surmised that this finding may be due to users feeling more comfortable with chatbots when discussing socially challenging information [24].

Discussion

Principal Findings

This paper describes what is known about the user experience and user uptake of COVID-19 chatbots. The 10 papers in this review included some describing chatbots for COVID-19 screening, some describing chatbots for information dissemination, and others seeking to understand uptake and user perspectives. The papers generally reported good acceptance of chatbots and reported a number of factors that appeared to determine this. Key themes included content, trust, digital ability, and acceptability. Trust in the chatbot, or in the chatbot provider, was commonly reported as being an important factor. Studies that reported on user acceptability generally found that the chatbots were rated highly, regardless of its type. The number of people seeking health information via digital methods throughout the pandemic has increased [3]; this may have played a role in this acceptability. Digital ability or perceived IT skills were not reported to have a large effect on chatbot use. In some studies, women used the chatbots more than men, and although there were differences in chatbot use or perspectives reported by age, it was clear that chatbots are being used and enjoyed by older people as well as those who are younger. There was no clear indication to whether age and gender play a role in openness to using these tools.

Research presenting results from the United States was overrepresented (5/9, 56% of studies), and the results seen here indicate a need for more diverse reporting and evaluation with COVID-19 chatbots. There have been broader calls in the literature for chatbots responding to COVID-19 to take a global view and be developed with researchers and local data scientists from low- and middle-income countries, to enable better access on a global level [32].

The COVID-19 pandemic has presented challenges to health services and authorities globally, with solutions being developed and rolled out to support countries with accurate and timely health information at a rapid speed. This expediated time frame may have resulted in some of the usual design processes, such as user consultations, message testing, and literature reviews, having not been conducted prior to launch, resulting in phases of evaluation and iteration in real-world settings. The challenge of managing the amount of content in a fast-moving landscape was described by one national government chatbot [30], and partnerships with technical teams were reported as being beneficial [25]. New processes were developed and chatbots have helped to quickly disseminate high-quality information in a time of crisis. These processes and learnings from implementation will be key in informing chatbots for future health emergencies. As well as information delivery, chatbots have the potential to inform information provision and pandemic response by analyzing the questions people are asking or looking at the information they are searching for to identify information voids.

The review revealed a lack of standardized reporting on user experience and user preferences. Although this may be partially explained by the need for expedited deployment, it will be important in the future to revisit and reaffirm best practices in user-focused design and acceptance testing to ensure that these tools are effective. Developing standardized measures for reporting on user experiences with chatbots will help to synthesize the evidence and move it forward in a more cohesive manner. The literature suggests that working in a co-design model involving end users can help with engagement [7], and that usability, enjoyment, and trust are also key factors [9].

It is apparent that further research is needed to better understand user experience, engagement, and uptake, particularly more research from countries outside of the United States. The COVID-19 pandemic has had an unprecedented effect on digital health [33]. Lessons learned on how best to reach people with health information and interventions during the COVID-19 pandemic will have broad application to other health emergencies and future pandemics.

Limitations

The scoping review summarized the evidence about user experience with COVID-19 chatbots, but as the goal of a scoping review is to map the literature, study quality was not assessed and data synthesis did not occur. The studies that were included reported diverse aspects of user experience, resulting in limitations in drawing together findings on particular components.

Conclusion

The COVID-19 pandemic presented a unique and specific challenge for digital health interventions in that design and implementation were required at a rapid speed as digital health service adoption was accelerated across the globe. This paper adds to the literature by describing what is known about this rapid implementation process in terms of user experience and user uptake and provides guidance for future tools, as well as directions for future research. This paper has shown both the potential of chatbots to reach users in an emergency and the need to better understand how users engage and what they want.

Acknowledgments

The authors wish to acknowledge the World Health Organization for funding to write this research and technical support in conceptualization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms.
[DOCX File , 16 KB - jmir_v24i12e35903_app1.docx]

References

1. Rolling updates on coronavirus disease (COVID-19). World Health Organization. 2020 Jul 31. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2022-12-16]
2. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2022-12-16]
3. Du H, Yang J, King RB, Yang L, Chi P. COVID-19 increases online searches for emotional and health-related terms. Appl Psychol Health Well Being 2020 Dec;12(4):1039-1053 [FREE Full text] [doi: 10.1111/aphw.12237] [Medline: 33052612]
4. Walwema J. The WHO Health Alert: communicating a global pandemic with WhatsApp. J Bus Tech Commun 2020 Sep 16;35(1):35-40. [doi: 10.1177/1050651920958507]
5. Lo HN, Lee CB. Chatbots and conversational agents: a bibliometric analysis. 2017 Feb 12 Presented at: 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); December 10-13, 2017; Singapore p. 215-219. [doi: 10.1109/ieem.2017.8289883]
6. Kieffer S, Ko A, Ko G, Gruson D. Six case-based recommendations for designing mobile health applications and chatbots. Clin Chim Acta 2019 Jun;493:S29-S30. [doi: 10.1016/j.cca.2019.03.072]
7. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
8. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical metrics used to evaluate health care chatbots: scoping review. J Med Internet Res 2020 Jun 05;22(6):e18301 [FREE Full text] [doi: 10.2196/18301] [Medline: 32442157]
9. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
10. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: a scoping review. Int J Med Inform 2019 Dec;132:103978 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
11. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. NPJ Digit Med 2020;3:65 [FREE Full text] [doi: 10.1038/s41746-020-0280-0] [Medline: 32377576]
12. WHO launches a chatbot on Facebook Messenger to combat COVID-19 misinformation. World Health Organization. 2020 Apr 15. URL: https://www.who.int/news-room/feature-stories/detail/who-launches-a-chatbot-powered-facebook-messenger-to-combat-covid-19-misinformation [accessed 2022-12-16]
13. Coronavirus self-checker. Centers for Disease Control and Prevention. URL: https://covid19healthbot.cdc.gov [accessed 2021-07-17]

14. Tangcharoensathien V, Bassett MT, Meng Q, Mills A. Are overwhelmed health systems an inevitable consequence of COVID-19? experiences from China, Thailand, and New York state. BMJ 2021 Jan 22;372:n83 [FREE Full text] [doi: 10.1136/bmj.n83] [Medline: 33483336]
15. Espinoza J, Crown K, Kulkarni O. A guide to chatbots for COVID-19 screening at pediatric health care facilities. JMIR Public Health Surveill 2020 Apr 30;6(2):e18808 [FREE Full text] [doi: 10.2196/18808] [Medline: 32325425]
16. Battineni G, Chintalapudi N, Amenta F. AI chatbot design during an epidemic like the novel coronavirus. Healthcare (Basel) 2020 Jun 03;8(2):154 [FREE Full text] [doi: 10.3390/healthcare8020154] [Medline: 32503298]
17. Meinert E, Milne-Ives M, Surodina S, Lam C. Agile requirements engineering and software planning for a digital health platform to engage the effects of isolation caused by social distancing: case study. JMIR Public Health Surveill 2020 May 06;6(2):e19297 [FREE Full text] [doi: 10.2196/19297] [Medline: 32348293]
18. Almalki M, Azeez F. Health chatbots for fighting COVID-19: a scoping review. Acta Inform Med 2020 Dec;28(4):241-247 [FREE Full text] [doi: 10.5455/aim.2020.28.241-247] [Medline: 33627924]
19. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
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 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Rayyan—intelligent systematic review. Rayyan. URL: https://www.rayyan.ai [accessed 2022-12-16]
22. Almalki M. Exploring the influential factors of consumers' willingness toward using COVID-19 related chatbots: an empirical study. Med Arch 2021 Feb;75(1):50-55 [FREE Full text] [doi: 10.5455/medarh.2021.75.50-55] [Medline: 34012200]
23. Almalki M. Perceived utilities of COVID-19 related chatbots in Saudi Arabia: a cross-sectional study. Acta Inform Med 2020 Sep;28(3):218-223 [FREE Full text] [doi: 10.5455/aim.2020.28.218-223] [Medline: 33417645]
24. Dennis AR, Kim A, Rahimi M, Ayabakan S. User reactions to COVID-19 screening chatbots from reputable providers. J Am Med Inform Assoc 2020 Nov 01;27(11):1727-1731 [FREE Full text] [doi: 10.1093/jamia/ocaa167] [Medline: 32984890]
25. Judson TJ, Odisho AY, Young JJ, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jul 01;27(9):1450-1455 [FREE Full text] [doi: 10.1093/jamia/ocaa130] [Medline: 32531066]
26. Liang KH, Shi W, Oh Y, Zhang J, Yu Z. Discovering chatbot's self-disclosure's impact on user trust, affinity, and recommendation effectiveness. arXiv. Preprint posted online on June 3, 2021. [doi: 10.48550/arXiv.2106.01666]
27. McKillop M, South BR, Preininger A, Mason M, Jackson GP. Leveraging conversational technology to answer common COVID-19 questions. J Am Med Inform Assoc 2021 Mar 18;28(4):850-855 [FREE Full text] [doi: 10.1093/jamia/ocaa316] [Medline: 33517402]
28. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020 Nov 30;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]
29. Ollier J, Santhanam P, Kowatsch T. Face(book)ing the truth: initial lessons learned using Facebook advertisements for the chatbot-delivered Elena+ Care for COVID-19 intervention. 2021 Presented at: HEALTHINF 2021: 14th International Joint Conference on Biomedical Engineering Systems and Technologies; February 11-13, 2021; Online event p. 781-788. [doi: 10.5220/0010403707810788]
30. Rodsawang C, Thongkliang P, Intawong T, Sonong A, Thitiwatthana Y, Chottanapund S. Designing a competent chatbot to counter the COVID-19 pandemic and empower risk communication in an emergency response system. OSIR 2020 Jun;13(2):71-77 [FREE Full text]
31. Schubel LC, Wesley DB, Booker E, Lock J, Ratwani RM. Population subgroup differences in the use of a COVID-19 chatbot. NPJ Digit Med 2021 Feb 19;4(1):30 [FREE Full text] [doi: 10.1038/s41746-021-00405-8] [Medline: 33608660]
32. Luengo-Oroz M, Hoffmann Pham K, Bullock J, Kirkpatrick R, Luccioni A, Rubel S, et al. Artificial intelligence cooperation to support the global response to COVID-19. Nat Mach Intell 2020 May 22;2(6):295-297. [doi: 10.1038/s42256-020-0184-3]
33. Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. BMJ Health Care Inform 2020 Jun 21;27(1):e100166 [FREE Full text] [doi: 10.1136/bmjhci-2020-100166] [Medline: 32565418]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

Edited by T Leung, V Arnold, H Gouda; submitted 29.12.21; peer-reviewed by A Dennis, A Joseph; comments to author 23.03.22; revised version received 02.06.22; accepted 08.12.22; published 27.12.22.

Please cite as:
White BK, Martin A, White JA
User Experience of COVID-19 Chatbots: Scoping Review
J Med Internet Res 2022;24(12):e35903
URL: https://www.jmir.org/2022/12/e35903
doi:10.2196/35903
PMID:36520624

Research Letter

Designing Emotions for Health Care Chatbots: Text-Based or Icon-Based Approach

Shubin Yu[1], MA, MSc, PhD; Luming Zhao[2], MA

[1]Department of Communication and Culture, BI Norwegian Business School, Oslo, Norway

[2]School of Journalism, Fudan University, Shanghai, China

Corresponding Author:
Shubin Yu, MA, MSc, PhD
Department of Communication and Culture
BI Norwegian Business School
Nydalsveien 37
Oslo, 0484
Norway
Phone: 47 41228055
Email: shubin.yu@bi.no

(J Med Internet Res 2022;24(12):e39573) doi:10.2196/39573

KEYWORDS

chatbot; health care; emotion; psychological distance; perception; human behavior; behavioral intention; predict; emotional intensity; text-based; icon-based; design

Introduction

Health care chatbots, which are being widely adopted by providers, offer many benefits to users [1]. However, the limited communication capabilities of chatbots hinder their interactions with humans [2]. Therefore, text-based (ie, verbal emotional expression, eg, saying “I am so sorry to hear that”) and icon-based (ie, nonverbal emotional expression, eg, using emojis, emoticons, or stickers) approaches are adopted to communicate emotion in chatbot messages. Previous studies have suggested that both emotion design approaches are effective in improving the evaluation of health care chatbots [3,4]. However, the two approaches differ greatly from each other in their presentation, mechanism, and effectiveness. Understanding such differences could help system developers to optimize their health care chatbots. Nevertheless, research comparing these two approaches of emotion designs, to our knowledge, is nonexistent. This study aims to understand the mechanism and the interaction effect of these two approaches to see if the effect of one approach depends on the other one. In general, we proposed the following hypothesis: both text-based and icon-based emotional clues for health care chatbots can increase perceived emotional intensity (H1). To test the interaction effect of the two approaches, we hypothesized that the addition of an icon-based clue would not significantly affect emotional intensity when a text-based clue is already present (H2). Furthermore, emotional intensity will reduce psychological distance and increase behavioral intention (H3). Please refer to Multimedia Appendix 1 for the theoretical framework and hypothesis development.

Methods

In total, 483 respondents were recruited through a web-based panel in China. The mean age of the participants was 28.8 (SD 8.84) years. A majority of participants self-identified as female (n=300, 62.1%). We used a 2 (text-based emotion design: yes vs no) by 2 (icon-based emotion design: yes vs no) between-subjects factorial experimental design. Participants were asked to imagine they had abdominal pain and then consult a prediagnostic chatbot online. The participants were randomly assigned to one of the four conditions. They were shown a screenshot of a conversation with the chatbot. After viewing the screenshot, participants were asked to answer a series of questions about perceived emotional intensity, psychological distance, and behavioral intention (see Multimedia Appendix 1).

Results

The results of a 2-way ANOVA showed that both icon-based ($mean_{icon}$ 4.15, SD 1.41; $mean_{no}$ 3.62, SD 1.44; $F_{1,479}$=14.4; P<.001; η^2=0.03) and text-based designs significantly enhanced the perceived emotional intensity ($mean_{text}$ 4.35, SD 1.38; $mean_{no}$ 3.46, SD 1.40; $F_{1,479}$=51.2; P<.001; η^2=0.10). H1 was therefore supported. Furthermore, we observed an interaction effect between icon- and text-based designs ($F_{1,479}$=7.96; P=.006; η^2=0.02). In particular, when text-based designs were not used, icon-based designs increased the emotional intensity ($mean_{icon}$ 3.87, SD 1.39; $mean_{no}$ 3.05, SD 1.26; $F_{1,243}$=23.2;

P<.001; η^2=0.09). However, when text-based designs were used, the effect of icon-based designs disappeared ($mean_{icon}$ 4.41, SD 1.40; $mean_{no}$ 4.29, SD 1.35; $F_{1,236}$=0.45; P=.50; η^2=0.002). These findings were consistent with H2.

We performed a moderated serial mediation model to further test our hypotheses. The analysis revealed a significant moderated mediation index (effect −0.07, SE 0.03, lower limit CI −0.12, upper limit CI −0.02). When there was no text-based design, icon-based designs significantly increased the emotional intensity, and thus shortened the psychological distance and enhanced the behavioral intention (Figure 1). This indirect effect was not significant when text-based designs were used (Figure 2). Overall, H3 was supported.

Figure 1. The indirect effect of icon-based designs when text-based designs were not used.

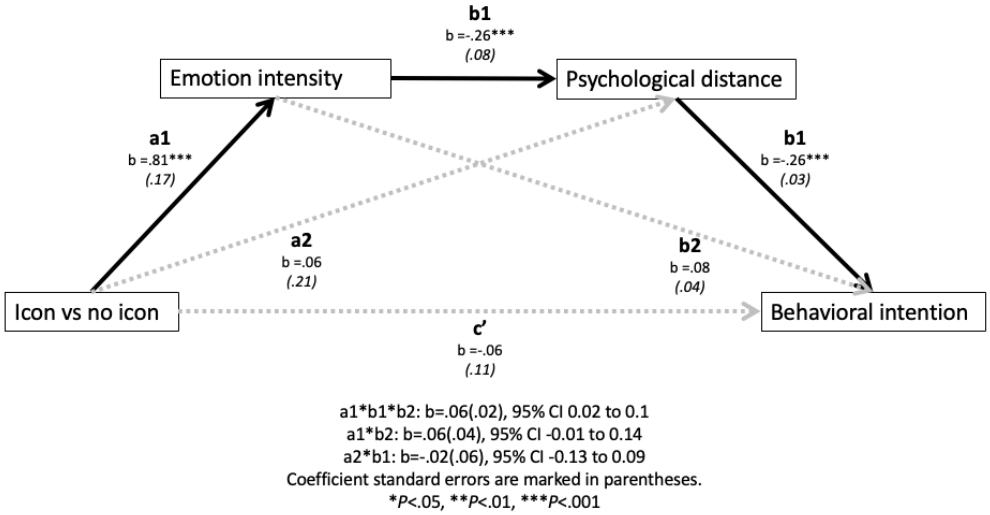
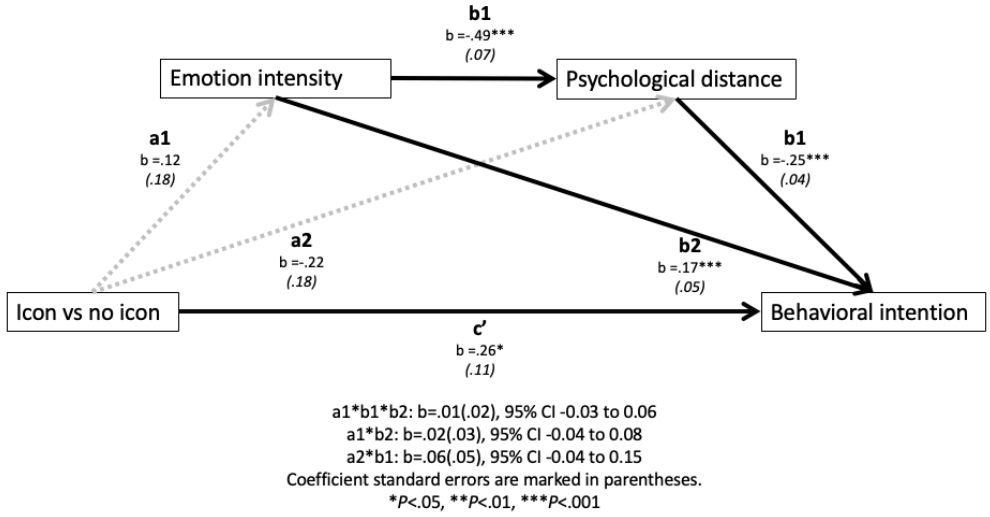


Figure 2. The indirect effect of icon-based designs when text-based designs were used.



Discussion

This study demonstrated that both icon- and text-based emotion designs of health care chatbots increase users' perceived emotional intensity. The effect of the text-based approach on emotional intensity was stronger than the icon-based one. The impact of the icon-based design on the perceived emotional intensity was mitigated when combining text-based designs together. Furthermore, the perceived emotional intensity reduced the psychological distance and enhanced behavioral intention. This is consistent with previous studies on interpersonal emotion disclosure and psychological distance [5].

The findings fill a void in the literature on health care chatbots' emotional design. In particular, we observed an antagonist effect of the two approaches of emotion design on emotional intensity, suggesting that using a single approach is sufficient. Additionally, previous research has examined the psychological distance between physical social robots and humans [6,7]. This study extends our understanding of the role of psychological distance in the effect of emotional expression of health care chatbots. Future studies may also investigate when the use of icon-based designs may backfire (eg, when conveying different emotional valence [8]). For example, the use of emojis may make interactions appear unprofessional in contexts like medical consultation [9]. This implies that the use of icon-based emotional designs may have negative effects under certain

circumstances for health care chatbots, which should be further explored in future research (see Multimedia Appendix 1 for details).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information.
[DOCX File , 1218 KB - jmir_v24i12e39573_app1.docx]

References

1. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: a scoping review. Int J Med Inform 2019 Dec;132:103978 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
2. 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]
3. 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]
4. Beattie A, Edwards AP, Edwards C. A bot and a smile: interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. Commun Stud 2020 Feb 16;71(3):409-427. [doi: 10.1080/10510974.2020.1725082]
5. Laurenceau JP, Rivera LM, Schaffer AR, Pietromonaco PR. Intimacy as an interpersonal process: current status and future directions. In: Mashek DJ, Aron A, editors. Handbook of Closeness and Intimacy. New York, NY: Psychology Press; 2004:61-78.
6. Mumm J, Mutlu B. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In: Proceedings of the 6th International Conference on Human-Robot Interaction. 2011 Mar Presented at: HRI '11; March 6-9, 2011; Lausanne, Switzerland p. 331-338. [doi: 10.1145/1957656.1957786]
7. Hu Y, Hoffman G. Using skin texture change to design emotion expression in social robots. 2019 Presented at: 14th ACM/IEEE International Conference on Human-Robot Interaction; March 11-14, 2019; Daegu, South Korea p. 2-10. [doi: 10.1109/HRI.2019.8673012]
8. Kaye LK, Rodriguez-Cuadrado S, Malone SA, Wall HJ, Gaunt E, Mulvey AL, et al. How emotional are emoji?: exploring the effect of emotional valence on the processing of emoji stimuli. Comput Hum Behav 2021 Mar;116:106648. [doi: 10.1016/j.chb.2020.106648]
9. Fadhil A, Schiavo G, Wang Y, Yilma BA. 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 May Presented at: PervasiveHealth '18; May 21-24, 2018; New York, NY p. 378-383. [doi: 10.1145/3240925.3240965]

Edited by R Kukafka; submitted 16.05.22; peer-reviewed by ER Khalilian, E Sezgin; comments to author 03.06.22; revised version received 24.06.22; accepted 30.11.22; published 08.12.22.

Please cite as:
Yu S, Zhao L
Designing Emotions for Health Care Chatbots: Text-Based or Icon-Based Approach
J Med Internet Res 2022;24(12):e39573
URL: https://www.jmir.org/2022/12/e39573
doi:10.2196/39573
PMID:36454078

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org