Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 24 (2022), Issue 11 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

XSL•FO
RenderX

Original Papers

XSL•FO
RenderX

XSL•FO
RenderX

XSL•FO
RenderX

XSL•FO
RenderX

Viewpoints

Corrigenda and Addendas

Research Letter

XSL•FO
RenderX

Review

The Impact of Artificial Intelligence on Health Equity in Oncology: Scoping Review

Paul Istasy[1,2], BSc, MA; Wen Shen Lee[3], BMedSci; Alla Iansavichene[4], MSc; Ross Upshur[5,6], BA, MA, MSc, MD; Bishal Gyawali[7,8], MD, PhD; Jacquelyn Burkell[9], PhD; Bekim Sadikovic[3], PhD; Alejandro Lazo-Langner[10], MSc, MD; Benjamin Chin-Yee[2,10,11], MA, MD

[1]Schulich School of Medicine and Dentistry, Western University, London, ON, Canada
[2]Rotman Institute of Philosophy, Western University, London, ON, Canada
[3]Department of Pathology & Laboratory Medicine, Schulich School of Medicine, Western University, London, ON, Canada
[4]Library Services, London Health Sciences Centre, London, ON, Canada
[5]Division of Clinical Public Health, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[6]Bridgepoint Collaboratory for Research and Innovation, Lunenfeld Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada
[7]Division of Cancer Care and Epidemiology, Department of Oncology, Queen's University, Kingston, ON, Canada
[8]Division of Cancer Care and Epidemiology, Department of Public Health Sciences, Queen's University, Kingston, ON, Canada
[9]Faculty of Information and Media Studies, Western University, London, ON, Canada
[10]Division of Hematology, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada
[11]Division of Hematology, Department of Medicine, London Health Sciences Centre, London, ON, Canada

Corresponding Author:
Benjamin Chin-Yee, MA, MD
Division of Hematology
Department of Medicine
London Health Sciences Centre
800 Comissioners Rd E
London, ON, N6A 5W9
Canada
Phone: 1 519 685 8475
Email: Benjamin.Chin-Yee@lhsc.on.ca

Abstract

Background: The field of oncology is at the forefront of advances in artificial intelligence (AI) in health care, providing an opportunity to examine the early integration of these technologies in clinical research and patient care. Hope that AI will revolutionize health care delivery and improve clinical outcomes has been accompanied by concerns about the impact of these technologies on health equity.

Objective: We aimed to conduct a scoping review of the literature to address the question, “What are the current and potential impacts of AI technologies on health equity in oncology?”

Methods: Following PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines for scoping reviews, we systematically searched MEDLINE and Embase electronic databases from January 2000 to August 2021 for records engaging with key concepts of AI, health equity, and oncology. We included all English-language articles that engaged with the 3 key concepts. Articles were analyzed qualitatively for themes pertaining to the influence of AI on health equity in oncology.

Results: Of the 14,011 records, 133 (0.95%) identified from our review were included. We identified 3 general themes in the literature: the use of AI to reduce health care disparities (58/133, 43.6%), concerns surrounding AI technologies and bias (16/133, 12.1%), and the use of AI to examine biological and social determinants of health (55/133, 41.4%). A total of 3% (4/133) of articles focused on many of these themes.

Conclusions: Our scoping review revealed 3 main themes on the impact of AI on health equity in oncology, which relate to AI’s ability to help address health disparities, its potential to mitigate or exacerbate bias, and its capability to help elucidate determinants of health. Gaps in the literature included a lack of discussion of ethical challenges with the application of AI

XSL•FO
RenderX

technologies in low- and middle-income countries, lack of discussion of problems of bias in AI algorithms, and a lack of justification for the use of AI technologies over traditional statistical methods to address specific research questions in oncology. Our review highlights a need to address these gaps to ensure a more equitable integration of AI in cancer research and clinical practice. The limitations of our study include its exploratory nature, its focus on oncology as opposed to all health care sectors, and its analysis of solely English-language articles.

(J Med Internet Res 2022;24(11):e39748) doi:10.2196/39748

KEYWORDS

artificial intelligence; eHealth; digital health; machine learning; oncology; cancer; health equity; health disparity; bias; global health; public health; cancer epidemiology; epidemiology; scoping; review; mobile phone

Introduction

Background

Artificial intelligence (AI), a field that aims to create computers that can achieve human-like understanding and perform tasks normally associated with human intelligence, is finding increasing applications in health care and public health [1,2]. Machine learning (ML) is a form of AI that involves algorithms that draw on big data—data sets whose size go beyond the capabilities of standard data analysis software—to *learn* to make predictions [3]. Oncology has been the focus of significant AI research and development and serves as an important area to observe and assess the early integration of AI in health care [4]. AI applications in oncology are expanding to cover a wide range of uses, from pathology and diagnostic imaging to clinical risk prediction and treatment planning for several types of cancer [5-7].

Despite its promise, the use of AI in health care raises several ethical issues, most notably concerns over bias and the potential for AI systems to adversely impact health equity. Health equity has been defined as "the absence of systematic disparities in health between groups with different levels of underlying social advantage/disadvantage" [8]. Studies have demonstrated how the use of biased data sets in training ML algorithms can exacerbate health inequities [9-11]. For example, Obermeyer et al [11] revealed how an ML algorithm trained to predict health risk consistently underestimated the health of Black patients because of the use of health care cost as a proxy for health. However, others have argued that AI systems can help illuminate health inequities and, if used correctly, may help address existing disparities [12-15]; for example, AI has been used to analyze search engine results from 54 African nations to guide resource allocation and improve access to care [15]. It is no wonder that a recent report from the Wellcome Trust on the ethical, social, and political challenges of AI in health care was not able to reach a clear consensus on the impact of AI on health equity [16]. Moreover, despite cancer being a major focus of AI research and development, the impact of AI on health equity in oncology remains underexplored. There is growing literature characterizing the problems of health disparities in oncology, which range from issues of access to high-quality care and research to structural barriers in health promotion and the lack of awareness of existing health inequities [17]. Given the expanding use of AI in oncology, there is an urgent need to assess the interplay between AI technologies and health equity in oncology to better understand the social and ethical dimensions surrounding the integration of AI.

Objective

This scoping review of the literature aimed to address the question, "What are the current and potential impacts of AI applications on health equity in oncology?" We analyzed the literature on contemporary AI applications in oncology with a focus on implications for health equity to identify recurring themes as well as important gaps and areas for future research.

Methods

Overview

Our scoping review protocol followed previously established methods [18] with reporting in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) framework [19].

Search Strategy

We used a sensitive search strategy to identify a representative sample of the available literature on the influence of AI on health equity in oncology. On the basis of a combination of synonymous searches comprising controlled vocabularies, such as Medical Subject Headings in MEDLINE or EMTree descriptors in Embase, and free-text terms using alternative word spellings and endings for the 3 core concepts: AI (*algorithm, machine learning, artificial intelligence, deep learning, and convolutional neural networks*), equity (*health equality, health inequality, health disparity, and socioeconomic factors*), and oncology (*neoplasm, cancer, squamous, and metaplasia*), which informed a comprehensive search strategy developed by the clinical librarian (AI) with experience in conducting electronic literature searches on the recommendations from the review authors (PI, ALL, and BCY). We searched both databases (MEDLINE and Embase, via the OVID platform) from January 2000 to August 2021, and a preliminary search was performed on December 4, 2020. A detailed description of our search strategy is provided Multimedia Appendix 1.

Eligibility Criteria and Article Screening

In addition, the web-based search engine Google Scholar was used to identify additional potentially relevant studies that were not indexed in bibliographic databases. The bibliographies of all relevant retrieved articles were also examined to identify

XSL•FO
RenderX

further relevant studies. To capture the breadth of literature on AI and health equity in oncology, we did not impose limits based on study type and included clinical studies—that is, studies in which AI was applied and evaluated for a specific clinical intervention, whether it be diagnostic, prognostic, screening, or treatment planning—commentaries and opinion articles. Limits were imposed for English-language-only articles, as it was the main language of proficiency for the research team, thus allowing for detailed and critical examination of the selected articles to take place. All identified records from the electronic search were imported into Covidence systematic review software (Veritas Health Innovation) for further analysis and screening.

After duplicate records were removed, 2 reviewers (PI and WSL) independently screened the titles and abstracts of selected records using the inclusion and exclusion criteria, which were defined a priori: records were selected during the title and abstract screening if they mentioned the core concepts (AI, health equity, and oncology) or related terms. Abstracts were excluded if they did not meet the inclusion criteria or if they involved nonhuman participants. All conflicts were resolved by a third reviewer (BCY). The list of selected abstracts was then reassessed by all 3 reviewers (PI, WSL, and BCY) in full-text reviews to identify records related to the research question. Records that generated a unanimous consensus were selected for full-text review, whereas those that did not engage with the 3 key concepts were excluded. Conflicts were resolved through discussion between all 3 reviewers. A further full-text review was conducted by all 3 authors, further applying the eligibility criteria.

Data Extraction and Analysis

Data extraction and analysis involved both descriptive and qualitative components. Descriptively, we extracted data on the year of publication, country of affiliation of the senior author, type of institution of affiliation of the senior author, type of study, type of AI, cancer type, and, when available, the cost of the proposed technology. The country of affiliation of the senior author was classified as high income, low income, and middle income following the most recent United Nations classification [20]. Qualitatively, we analyzed articles for emerging themes related to health equity in oncology, inherent assumptions, and gaps in the literature. Thematic analysis followed the steps outlined by Braun and Clarke, which have been widely applied in scoping reviews of qualitative research, including in health care [21], to generate a comprehensive thematic representation of a given area of research [22]. This process involved familiarization with the data set of included articles, generation of initial codes, collation of codes into provisional themes, review of themes in relation to initial codes, and the entire data set, followed by definition and naming of each theme to generate a comprehensive representation of the data. Steps of data familiarization and initial coding were performed independently by 3 reviewers (PI, WSL, and BCY); steps of collation, review of themes, and definition and naming were performed through discussion between study coauthors. Articles that had insufficient engagement with 3 key concepts, that is, those that mentioned the issues of bias or equity but did not elaborate on specific issues arising from AI, or made insufficient links between the core concepts, that is, those that mentioned all 3 core concepts but had no further exploration of their relationships, were excluded.

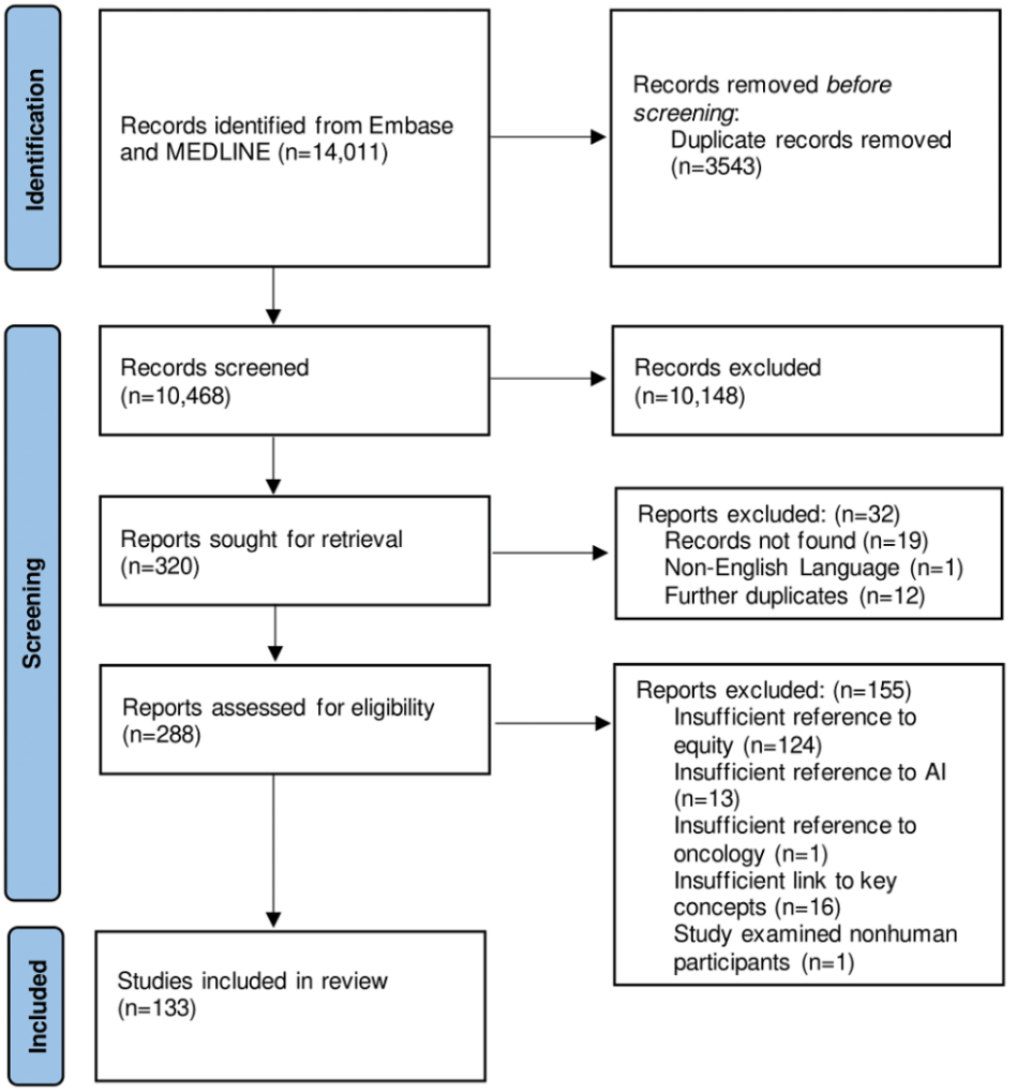
Results

Selection and Characteristics of Sources of Evidence

Our search yielded 14,011 records. After removing duplicates, 10,468 records were screened, and 133 articles met the inclusion criteria [4,23-154] (Figure 1). All the records included in our review were published between 2010 and 2021, with the majority (124/133, 93.2%) published after 2018 (Table 1). Although a range of countries, based on the affiliation of the senior author, were represented in our review (Figure 2), most were from the United States (90/133, 67.7%). The majority were from academic centers (121/133, 90.9%), with a minority from the government, nonprofit organizations, and industry (Table 1). Approximately half of the records involved clinical studies (68/133, 51.1%), whereas the rest were epidemiological studies, commentaries, surveys, and interviews. Most of the records drew on ML techniques to address their research question: 12.8% (17/133) records discussed AI in general; 30.8% (41/133) records did not specify the type of ML used or used multiple ML algorithms; 47.4% (63/133) used supervised ML algorithms; and a smaller subset (4/133, 3%; 6/133, 4.5%; and 2/133, 1.5%) used unsupervised ML, natural language processing, and reinforcement ML, respectively. AI was used for a wide range of applications and often a combination of applications, including epidemiological (28/133, 21.1%), diagnostic (25/133, 18.8%), prognostic (25/133, 18.8%), and screening (25/133, 18.8%; Table 1).

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram for the identification of studies via databases and registers. AI: artificial intelligence.



XSL•FO
RenderX

Table 1. Characteristics of studies included in the scoping review (n=133).

Study characteristics	Studies, n (%)
Year of publication[a]	
2000-2017	9 (7.5)
2018	13 (9.8)
2019	21 (15.8)
2020	48 (36.1)
2021	42 (31.6)
Type of study	
Clinical	62 (46.6)
Epidemiological	40 (30)
Review	15 (11.3)
Commentary	11 (8.3)
Survey and interviews	5 (3.8)
Institution type[b]	
Academic	121 (91)
Governmental and nongovernmental organizations	12 (9)
Type of artificial intelligence application[c]	
Screening	41 (30.8)
Diagnostic	41 (30.8)
Therapeutic	15 (11.3)
Prognostic	45 (33.8)
Epidemiological	45 (33.8)
Type of cancer	
General	28 (21.1)
Gynecologic	19 (14.3)
Breast	16 (12)
Oral	12 (9)
Prostate	12 (9)
Skin	12 (9)
Lung	8 (6)
Hematologic	6 (4.5)
Brain	4 (3)
Liver	4 (3)
Colorectal	3 (2.3)
Esophageal	3 (2.3)
Head and neck	2 (1.5)
Pancreatic	2 (1.5)
Gastrointestinal	1 (0.8)
Thyroid	1 (0.8)

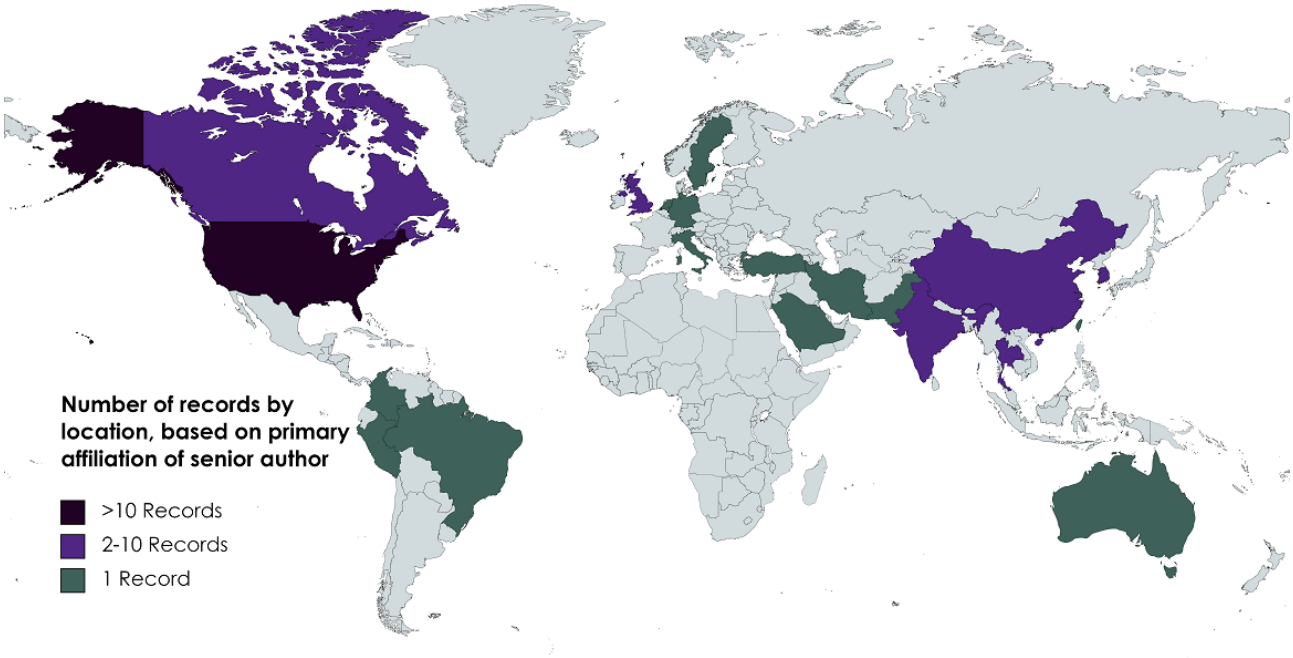
[a]Inclusive.

[b]On the basis of affiliation of the senior author.

[c]Total numbers exceed 133 due to 26 articles falling into multiple categories.

XSL•FO
RenderX

Figure 2. Country of affiliation of senior author (map created with MapChart).



AI Applications in Specific Cancer Types

Studies from our review investigated a wide range of cancers, with general oncological applications being the dominant category (28/133, 21.1%), followed by gynecologic (19/133, 14.3%), breast (16/133, 12%), oral (12/133, 9%), prostate (12/133, 9%), and dermatologic cancers (12/133, 9%). Among the articles on gynecologic cancers, 84% (16/19) were categorized under theme 1, discussing the use of AI technologies to address disparities in gynecologic cancer screening (11/16, 70%) [23,84-93], diagnosis (4/16, 25%) [94-97], and treatment (1/16, 6%) [98]. Of the 16 articles, 15 (94%) developed AI technologies to target gynecologic cancer disparities in low- and middle-income countries (LMICs) [84-98], while 1 (6%) did so for implementation in high-income countries (HICs) [23]. The other 3 (n=19, 16%) articles fell under theme 3, discussing the use of AI to explore the genetic (1/3, 33%) [99] and social (2/3, 67%) determinants of health outcomes in gynecologic cancers [100,101]. Moreover, most of these articles were clinical studies (14/19, 74%) [84-89,93-95,97-101], 16% (3/19) were commentaries [23,90,91], 5% (1/19) was an epidemiological study [96], and 5% (1/19) was a review [92].

Articles examining breast cancer have discussed a broader range of themes relating to health equity. Of the 16 articles, 6 (38%) focused on theme 1 [24,102-106], with all 6 looking at the applications of AI in LMICs. Of the 16 articles, 2 (13%) fell under theme 2: one discussed the use of AI to mitigate bias [107], whereas the other raised the issue of how AI might exacerbate and mitigate biases in breast cancer diagnoses [108]. Of the 16 articles, 7 (44%) fell under theme 3, with 6 (86%) examining the link between social determinants [109-114] and 1 (14%) examining the link between genetic determinants of health and breast cancer [115]. Of the 16 articles, 1 (6%) fell under multiple themes [116]. In addition to touching on a wider variety of themes than gynecologic cancers, articles examining breast cancer were also more varied: 44% (7/16) were clinical studies [24,102,103,110,112-114], 25% (4/16) were epidemiological studies [104,109,111,115], 25% (4/16) were reviews [106-108,116], and 6% (1/16) was a commentary [105].

Critical Appraisal Within Sources of Evidence

We identified three main themes related to the impact of AI on health equity in oncology: (1) the development of AI technologies to reduce health disparities faced by populations in both LMICs and HICs; (2) the concern that biased AI algorithms might exacerbate health inequities counterposed by the hope that AI technologies might help overcome human biases; and (3) the power of AI to uncover biological and social determinants of health in oncology. Themes were further broken down into subthemes, where applicable. A full list of the articles categorized by theme can be found in Multimedia Appendices 2-5.

AI and Health Disparities

Overview

The most prominent theme in our analysis, based on the number of records, was the development of AI technologies to address health disparities in oncology (58/133, 43.6%). This included the use of AI to address disparities in access to screening, diagnostic, and therapeutic technologies for underserved populations in LMICs (53/133, 39.8%) and minority populations in HICs (3/133, 2.3%). Of 133 studies, 2 (1.5%) used AI to address disparities in both LMICs and HICs. A total of 16 articles on this theme were commentaries or reviews discussing multiple applications in cancer care. Of the 58 articles, 17 (29%) were described as pilot studies. We further divided this theme into several subthemes based on the type of AI technology, including AI applications, to analyze the genomic, histological, radiographic, image, and demographic data.

XSL•FO
RenderX

Using AI to Address Disparities in Cancer Screening and Diagnosis

The literature under this theme highlighted how technologies could improve the delivery of health care to disadvantaged populations in both LMICs and HICs. In LMICs, these technologies were aimed at rectifying 2 main problems: addressing health care personnel shortages, thereby reducing the bottleneck effect created by a low ratio of health care professionals to the populations they serve and overcoming constraints resulting from limited medical equipment [117]. For example, point-of-care and smartphone-based technologies for oral cancer screening in low-resource settings aim to address the bottleneck effect created by a low number of health care professionals [118]. One example of AI technology aimed at addressing constraints from limited medical equipment is a mobile-based oral cancer image analysis software for use in rural India [119]. In the absence of a stable internet connection, the AI algorithm can analyze images directly on a smartphone, which are then uploaded to a cloud server and assessed by a remote specialist when internet is available. AI applications to address health disparities in oncology in HICs was a less explored topic, with some articles discussing algorithms to selectively target disadvantaged populations [120,121]. For instance, given the high prevalence of oral cancer in South Asian populations [155], 1 study used ML to develop a quantitative cytology program to selectively improve oral cancer screening among South Asians living in British Columbia, Canada [120].

The development of AI aimed at reducing health disparities drew on a range of data, from genomics and imaging to demographic data, all aimed at reducing demands on underresourced health care systems and improving the available medical equipment. One example is an AI image analysis algorithm for breast cancer detection that improves screening in underserved and low-resource settings by applying deep learning to novel ultrasound techniques [105,106,114]. Finally, AI has also been applied to address disparities in access to diagnostic pathology; these included examples such as decision support systems to assist with histopathological diagnosis of brain tumors in resource-poor settings [122] and image analysis of cervical lesions [97].

Studies have reported a range of outcomes, with screening and diagnostic technologies showing a wide variation in sensitivity (75%-100%), specificity (71%-100%), and accuracy (61%-100%). Most studies on this topic (43/58, 74%) offered no comparison between the performance of the proposed technology and the existing standard of care. When the AI algorithms were directly compared with the standard of care, the results varied. Most of the articles noted no difference between AI algorithms and the standard of care [93,97,123-125], whereas others observed that the accuracy of AI algorithms was lower than that of human physicians [85,126]. One study noted that AI outperformed its human counterpart when detecting and staging prostate cancer in a higher number of patients [117].

Gaps and Challenges With Using AI to Address Health Disparities

Although several articles have highlighted how AI might help address health care shortages in LMICs, a recurrent problem noted in the literature is the lack of consideration for the infrastructure and human resources necessary to implement these AI technologies. To support the use of digital technologies, and specifically AI, LMICs require both health care providers trained to use specific technologies and sufficient technological infrastructure, including buildings where the hardware can be housed and cables to carry digital signals leading to widespread and stable internet access; in other words, the performance of AI algorithms is intertwined with sociotechnical factors [156,157]. Although HICs may have existing technological infrastructure to implement AI technologies more readily, LMICs often lack such infrastructure [158]. Considerations such as the cost of implementation and the need for maintenance and ongoing support once implemented, the need for trained personnel to use AI technologies, and the need for technological support to allow for the integration of the developed AI technologies were rarely discussed by articles in our review. Only select articles mentioned the lack of infrastructural considerations in the development of AI technologies [87,117,119,127,128]. For example, Anirvan et al [129] noted that, “while in developed countries with a well-equipped health care model in place this may not be a problem, in poor, rural, and resource-constrained settings, it may aggravate the burdened health care system in place.”

In addition, our review identified equity issues related to the cost of AI technologies; such technologies can be costly and may not be affordable in many LMICs under existing economic circumstances. Love et al [102] developed an AI device to triage breast lumps in low-resource settings but noted that “the device used in this study is more expensive than most LMICs settings can afford, lower cost devices are becoming more available.” However, others were able to create technologies that may be more affordable for LMICs: a gene expression assay costing US $450 capable of assessing samples for only US $10 [130].

To ensure that AI technologies designed for HICs can be effectively applied in LMICs, collaboration between these 2 settings is invaluable. Of the 42 studies that were conducted in LMICs, 11 (26%) were led by research groups from LMICs in question, and from the remaining 31 records, 27 involved collaboration with coauthors from the specific LMIC. When such a collaboration occurred, AI technologies were primarily designed in HICs and implemented in LMICs. This divide between the location of development and location of the implementation of AI in global oncology can pose a barrier to integration in LMICs due to costs [102] and infrastructural considerations [88], thereby suggesting a need for greater attention to co-design, which refers to the involvement of end users in the design process of AI technologies [159]. Moreover, it is important to recognize that the inclusion of researchers from LMICs in the design of AI technologies alone does not guarantee widespread improvements in health for patients in these countries. Rather, benefits are often limited to select partner sites of HICs; therefore, while these technologies may help address global disparities, they may exacerbate inequities within LMICs [130]. To ensure a more equitable distribution of benefits within LMICs, research should extend beyond specific partner institutions, engaging additional stakeholders from relevant government and nongovernmental organizations

XSL•FO
RenderX

to evaluate and implement technologies. However, as noted in our review, there was only limited involvement of nonacademic institutions in the articles included in our review.

One additional problem that has been raised surrounding the use of AI in LMICs is the issue of data colonialism [160], a practice in which data are extracted from LMICs by institutions in HICs for the purposes of building algorithms whose benefits accrue primarily to stakeholders in HICs [161]. Although articles from our review did not engage directly with these issues, some did discuss important considerations for what collaboration means between HICs and LMICs [85,97,130,131]. However, there was limited acknowledgment of ethical issues arising from the involvement of LMICs as mere resources for data extraction and algorithmic training or as an exploratory ground for novel applications of AI technologies for global health.

AI and Bias

Overview

The second theme identified in our review relates to the issue of bias. Bias in AI is a widely discussed topic and has the potential to exacerbate health disparities across different populations; while bias is an inherent feature of all AI systems, the main types of bias of ethical concern are those biases arising in algorithmic development or data sets [162] that can result in individuals being treated unfairly based on particular characteristics [163]. In a similar vein, 1 article in our review distinguished between concepts of desirable and undesirable biases, whereas desirable biases are those that take group data into consideration to account for base-rate differences and undesirable biases are those that are developed based on inaccurate or incomplete data, which in turn leads to group discrimination [132]. For instance, total melanoma rates are higher in men than in women [164]; thus, a desirable bias would include a training sample for an AI algorithm used to detect melanoma purposefully *biased* (desirably) to contain more men than women, representing the base rates of melanoma incidence. The authors suggest the use and integration of desirable biases to promote gender equity in health care while decreasing undesirable biases.

With rising concerns surrounding bias in AI [9-11], and conversely, the hope that AI algorithms may be able to help mitigate bias in human judgment [12,13,15], we expected to see a much larger number of articles discussing this issue; however, only 12% (16/133) articles directly engaged with the theme of bias. These articles fell into 2 main categories: those that explored how AI algorithms might help mitigate biased judgments in physicians' clinical practice (5/133, 5%) and those that argued that AI trained on biased data sets can exacerbate existing inequities (10/133, 7.5%), while 1 article (1/133, 0.8%) focused on both subthemes.

The Use of AI to Uncover Bias in Clinical Practice

The use of AI technologies in health care can uncover biases in both data sets and physicians' actions. For instance, head and neck cancers may develop spontaneously or in association with human papillomavirus (HPV), and characterization of such cancers as HPV-associated can affect treatment decisions [165]. Patients diagnosed with HPV-positive versus HPV-negative head and neck cancers have different demographic features, with younger individuals and individuals with more sexual partners being overrepresented in the HPV-positive group [166]. D'Souza et al [133] thus used AI to assess the use of clinical and demographic characteristics as diagnostic predictors of HPV-positive and HPV-negative head and neck cancers. However, these authors noted that clinical and demographic characteristics had only *moderate* accuracy in predicting HPV status, leading to a potential bias in treatment if these variables were used to predict HPV status without further investigation. In addition, AI can be used to uncover the biases found in data sets. Howard et al [134] deployed a deep learning model to assess institutional biases in data submitted to The Cancer Genome Atlas. They noted that biased digital histological signatures can stem from specific features of the institutions from which the data originate. AI algorithms may then provide prognostic information based on these institution-specific signatures rather than on the intrinsic histology of the sample.

The Use of AI to Mitigate Bias in Clinical Practice

We also identified articles that discussed the use of AI to mitigate bias in clinician decision-making. In criticizing the Fitzpatrick scale in dermatology, Okoji et al [135] argued that AI-based approaches might lead to a more objective classification system for skin typing. AI systems can identify subtle variations that are not visible to the human eye, thereby leading to more equitable dermatological assessments. However, a major caveat was the lack of discussion surrounding the populations used to train these AI algorithms in dermatology. For instance, several studies included predominantly White populations or did not specify the racial and ethnic makeup of the population used to develop their algorithms [124,136,167]. Only 1 article in our review specifically addressed this problem: to counterbalance the skewed nature of dermatologic data available for AI training, Pangti et al [137] sought selective patient populations to train an AI algorithm to detect skin diseases using locally generated data from India. As medical AI systems are prone to generating biased results that lead to disparities between ethnic groups, some authors proposed that stratification for minority communities that suffer from underrepresentation in training data sets could help rectify this bias [108]. Instead of a one-size-fits-all model, AI programs can be developed to target specific subpopulations. For instance, Gao and Cui [138] suggested the use of transfer learning, an AI training technique whereby knowledge gained from training an AI system on a larger data set, for example, a majority ethnic group, is transferred to be applied to a smaller data set, such as a minority ethnic group [138]. This technique attempts to compensate for missing data from "data-disadvantaged ethnic groups by leveraging knowledge learned from other groups with more abundant data" [138]. Yet, as the authors note, data inequality remains a central issue in training ML algorithms in multiethnic populations, and differential accuracy in performance between ethnic groups is an ongoing challenge.

Biased Data Sets and Biased AI

The final category in this theme was articles discussing the use of biased data sets to train AI algorithms; surprisingly, few articles discussed this topic. For instance, Khor et al [139] used

XSL•FO
RenderX

a data set with racial demographics of 53% non-Hispanic White, 22% Hispanic, and 13% Black or African American to develop a recurrence risk prediction model for adults with prostate cancer. Even with the explicit inclusion of race, they noted that the model had "worse performance in minority subgroups compared to NHW [non-Hispanic White]." Conversely, others argued that bias in training data sets of AI algorithms may not always result in decreased generalizability; for example, Gilson et al [140] suggested that biased gender representation in training data sets did not lead to decreased generalizability in an algorithm to predict survival in non–small cell lung cancer.

Gaps in the Discussion of AI and Bias

Overall, engagement with issues of bias resulting from the use of AI in oncology was limited, an unexpected finding, given that this concern is widely discussed elsewhere in the literature on AI ethics and may act as a mechanism through which AI systems exacerbate health inequities. Our findings suggest that bias remains an underexplored topic in the literature on AI in oncology. It is also worth noting that the few articles that mentioned bias often did so briefly in their limitations section, usually in reference to how biased data sets might impact the validity and generalizability of AI algorithms but without further engagement with how these issues might be mitigated or addressed by future research.

AI and Determinants of Health Outcomes

Overview

The final theme identified in our review was the use of AI to investigate the determinants of health outcomes in oncology. A total of 41.4% (55/133) articles fell under this theme and were divided into subthemes based on the determinants of health examined, ranging from biological variables (9/133, 6.8%) to social determinants of health (43/133, 32.3%), whereas 2.3% (3/133) articles focused on both themes. This category can be understood as the use of AI as an extension of traditional statistical models in clinical and epidemiological research in oncology.

AI and Biological Determinants of Health

Several articles under this theme applied AI to genomic data to predict outcomes in patients with cancer. For instance, Li et al [141] applied AI to genomic analysis across 3 racial groups to identify the impact of differential gene expression on racial disparities in cancer prevalence. They found differential gene expression in several cancers between racial groups, which they interpreted as supporting a genetic basis for racial differences in cancer prevalence.

AI and Social Determinants of Health

Although several studies have similarly applied AI in a reductionist manner, for example, to look for a genetic basis of health disparities [115,142], others have used AI to examine additional individual, environmental, and societal factors contributing to differential health outcomes between populations. Several articles in our review applied AI to shed light on the influence of race and socioeconomic status on health outcomes in oncology. For example, An et al [143] used an ML algorithm to examine the risk factors for the development of hepatocellular carcinoma in a Korean cohort, noting that higher income is associated with a lower risk of developing hepatocellular carcinoma. Bibault et al [144] applied AI to satellite imagery to investigate the relationship between socioeconomic status and cancer prevalence, observing that "satellite features are highly correlated with individual socioeconomic and health measures that are linked to cancer prevalence." Several studies have suggested that applying AI to demographic data could help provide more comprehensive risk stratification models in oncology [112,168,169].

AI has also been used to identify racial disparities in cancer outcome. Tossas et al [101] used AI to predict populations at risk of delayed diagnosis of cervical cancer. They noted that more than half of the patients with a late cancer diagnosis were African American, findings that they argue can be used to target cervical cancer screening. Others have also used AI to examine outcomes following neurosurgery for brain tumors, noting that minority race is an independent risk factor for an extended length of stay and increased cost [145,146].

AI has also been applied to examine the influence of rural and urban residences on cancer prevalence and outcomes. Rural residences are known to influence access to cancer treatment, with novel therapies often concentrated in academic centers located in urban settings [170]. The impact of rural residence on cancer outcomes was investigated by Zhong et al [112], who used AI to create personalized prognostication models for early invasive breast cancer in a Chinese cohort. By incorporating residential status in their algorithm, the group found that despite lower rates of breast cancer in rural populations, the associated mortality risk was significantly higher. Aghdam et al [147] used the AI algorithm to study access to stereotactic body radiation therapy for prostate cancer and noted that travel distance did not prevent access to stereotactic body radiation therapy for rural patients, suggesting that income and race may be more important determinants of access to treatment.

Gaps in Using AI to Investigate Determinants of Health

For most studies in our review, there was a lack of justification for the use of AI and, more specifically, a lack of discussion as to why particular AI algorithms were chosen and their advantages over other statistical methods to address a given research question. AI algorithms are undeniably powerful tools for analyzing large amounts of data and selecting articles that mention the benefits of AI over other statistical methods [143,144,167,168]. However, others have argued that the use of AI has not yielded better risk prediction models compared with traditional statistical methods [169]. In their review on the efficacy of AI as opposed to traditional statistics in medicine, Rajula et al [145] noted that the latter seemed to be more useful when the number of participants significantly outweighed the number of variables in question, whereas the former is more suitable in fields with a large quantity of data, such as omics or radiodiagnostics. In light of this discussion, further justification for the use of AI to address specific research questions in oncology should be undertaken.

XSL•FO
RenderX

Discussion

Principal Findings

In this review, we evaluate the literature on the impact of AI on health equity in oncology. We identified 14,011 records in our search, of which 133 (0.95%) were substantially engaged with the core concepts of AI, health equity, and oncology. Our literature review revealed three main themes related to how AI technologies can (1) help address health disparities, (2) mitigate or exacerbate biased decision-making, and (3) elucidate the biological and social determinants of cancer outcomes. These themes relate to several issues discussed in the literature on AI and health equity in oncology and health care.

The first main theme noted in our review is how AI technologies can help address health disparities, both in LMICs and HICs. Previous scholarship examining the application of AI in global oncology has shed light on numerous practical and ethical challenges that have been discussed in the literature [171]. The existence of a "digital divide," often cited as a key barrier to the implementation of AI technologies in global health, refers to the inequitable distribution of digital technologies, such as computational power, technical infrastructure, and data storage, that is required to use AI technologies [171]. Without prioritizing investment in the basic infrastructure, such as appropriate hardware to run AI programs, buildings where such hardware can be housed, and cables to carry digital signals, the utility of these technologies in the global health context should be questioned [172,173]. A number of articles identified in our review engaged with these voiced concerns, with some researchers creating technologies with the infrastructural capacities of specific LMICs in mind and others highlighting the need for additional infrastructure to support the technology they developed [87,102,129,130,148].

Another barrier to the implementation of AI technologies in LMICs discussed in the literature is the lack of generalizability of algorithms primarily designed in HICs but applied in LMICs [170]. As some researchers have observed, data used for training AI algorithms in HICs are "notorious for their lack of diversity, and concerns have been raised about their applicability even in their home countries" [172]. These data are often skewed toward the populations, diseases, and treatments available in countries training and developing AI technologies, thereby decreasing their generalizability to populations in LMICs. Articles from our review addressed this issue, voicing concerns about the applicability of AI algorithms developed in HICs to LMICs [108,110,134,143,149-151].

Our review also focused on solutions to the challenges posed by the integration of AI technologies in a global health context, which have been proposed elsewhere in the literature, with the predominant one being greater collaboration between HICs and LMICs in the development of AI technologies [171-173]. AI technologies created without appropriate consultation with the populations they are intended to serve may be highly inapplicable, impractical, and unethical. For example, treatment patterns produced by Watson for Oncology, an AI decision support system trained by data and experts from the Memorial Sloan Kettering Cancer Center, may be inapplicable to many LMICs [173]. In previous studies investigating this issue, some researchers have argued for the co-design of AI technologies, which requires the involvement of end users—and specifically marginalized groups—in AI research and development to ensure the equitable distribution of the benefits of these technologies [159,174].

To improve collaboration in global health research, others have proposed that journals publishing research conducted in LMICs have the responsibility of ensuring that at least one author involved in the study is from the countries in question [175]. We observed that this standard was met in most studies conducted in LMICs included in our review (27/31, 87%). However, further steps are required to ensure meaningful collaboration with investigators and stakeholders in LMICs, beyond simple inclusion in authorship, which risks fostering tokenism. As discussed earlier, this is especially important in AI research and development focused on addressing global health inequities in oncology, which needs to engage additional stakeholders beyond select partner sites to ensure fair distribution of benefits throughout populations [130,175]. This lacuna identified by our review reflects a broader lack of global coordination in AI research to set priorities and ensure fair distribution of research opportunities and resources, which is essential to prevent AI research from perpetuating existing global health inequities.

Finally, a balance must be struck between the global dissemination of existing diagnostic and treatment technologies and the development of new technologies for global health. Our review revealed how pilot studies of AI in global oncology are particularly common. Although pilot studies can provide an important starting point, if not followed by a robust evaluation to measure the clinical effectiveness of these technologies, which occurs in only a minority of cases [176], these applications will remain an ineffective means of addressing global health disparities in cancer care. Moreover, it has been noted that most cancer deaths occurring in LMICs are due to a lack of access to already present and cost-effective diagnostic and treatment strategies, as opposed to the latest cutting-edge technology [177,178]. Exploratory research into novel technologies in global oncology may detract from the need to develop cost-effective ways to disseminate existing evidence-based technologies in cancer care.

The second major theme noted in our review was the use of biased AI algorithms in clinical decision-making, which may impact the quality and accuracy of decisions and consequently lead to adverse health outcomes for patients [179]. One theme identified in our review was the use of AI algorithms to standardize and reduce bias in clinical decision-making in oncology. One high-profile example is Watson for Oncology, an AI decision support system that has been proposed as a method of standardizing clinical decisions. Watson for Oncology uses natural language processing to provide treatment recommendations in oncology based on the latest scientific literature. Select studies have shown high concordance between treatment plans produced by Watson for Oncology and recommendations from multidisciplinary tumor boards [180-182]. Previous criticisms of this technology have pointed toward problems using concordance to assess the capability of

XSL•FO
RenderX

AI technologies, such as Watson for Oncology, because it simply assesses its ability to reproduce specific expert knowledge while not evaluating the validity of applying this knowledge in different contexts [183,184]. Treatment recommendations are based on the current literature as opposed to novel findings produced by the AI system, and preexisting biases found in data sets will be exacerbated rather than mitigated in an automation process. As Murphy et al [185] note, concerns regarding implicit bias becoming embedded in AI algorithms have been widely voiced. The authors noted that implicit biases often reflect preexisting societal values that may exacerbate already-existing health inequities for marginalized populations. Moreover, concerns surrounding lack of transparency in how Watson for Oncology integrates data from heterogeneous sources to arrive at decisions, including the influence of implicit value judgments found in different oncology guidelines, require further attention, specifically focusing on how this might impact the application of Watson for Oncology in different global contexts and its effects on health equity.

Despite the pressing nature of these concerns, the paucity of studies on biased AI algorithms in our search was surprising. Many AI applications identified in our study were trained on selecting data sets from single institutions, creating a high risk of bias, which should be a pressing concern, given that algorithmic bias can exacerbate health inequities [140,186]. A prominent cause of bias is the lack of consideration of the different contexts in which an algorithm is developed and subsequently deployed. Academics weary of these concerns have argued that a generalizable AI model should be developed from data reflecting the diversity of patients on whom it will be applied, yet "most health organizations lack the data infrastructure required to collect the data needed to optimally train these algorithms" [186,187]. Patterns detected when these algorithms are trained on majority groups may result in decreased accuracy when applied to minority groups [188]. For instance, most AI algorithms for diagnosing melanoma are trained on white-skinned individuals and thus may underperform in diagnosing lesions on persons of color [189]. Panch et al [186] note that solutions to these contextual problems involve establishing the appropriate context for which the algorithms will be used. Our literature review identified some proposed solutions, such as the application of transfer learning to improve outcomes for populations with data sparsity; stratification of groups based on race and ethnicity to mitigate bias; and the need for multidisciplinary collaboration between clinicians, engineers, social scientists, and ethicists to aid in the contextual design and development of AI algorithms to mitigate biases [108,135,138,190].

The final theme identified in our review was the use of AI to examine determinants of health outcomes in oncology. Social determinants of health such as education, neighborhood, social community, and socioeconomic status impact health outcomes in oncology [191], and the complex interactions between these variables suggest a potential area for AI applications. Several studies in our review applied AI to analyze large volumes of data to help elucidate the social determinants of cancer outcomes. The identification of social determinants of health can help support more comprehensive strategies to improve health equity in underserved populations [192].

However, as noted by several researchers comparing the use of AI with traditional statistical methods to analyze large amounts of data, it is not always clear what benefits the former provides over the latter to investigate the social determinants of health [145]. A systematic review compared the performance of logistic regression and ML in clinical prediction models and found no evidence that ML performs better than logistic regression [193]. Moreover, traditional statistical models are often easier to interpret than complex, multilayered ML models. Trade-offs between accuracy and transparency have been widely discussed in the literature on AI [194,195] and should be considered when deciding the method of analysis for a given research question. Appropriate and sufficient justification for the use of ML models in clinical and epidemiological oncology research is imperative.

Ethical concerns regarding the use of AI to analyze large amounts of health care data have also been raised in the literature. In establishing a research ethics framework for health care ML, McCraden et al [196,197] note how AI can influence 2 phases of health care research: hypothesis generation and hypothesis testing. AI research focused on hypothesis generation applies computational techniques to large data sets to explore models with potential clinical applicability [197]. This type of exploratory research raises important ethical issues, such as the protection of data privacy and tensions between enabling ready access to data and the requirements of informed consent [197]. Most articles from our review under this theme fit into the hypothesis generation phase and used AI for exploratory research on the determinants of health outcomes in oncology. In our review, the discussion of ethical issues in data privacy versus the need to enable ready access to data was sparse, despite the importance of such considerations in exploratory AI research on social determinants of health, which often requires large amounts of personal health information and other sensitive data. Moreover, as previously emphasized by advocates for equity in AI, exploratory AI research also entails an ethical commitment to ensure representative data sets, including minorities and "data-impoverished" groups, to avoid biased and misleading findings [198]. Few articles from our review addressed these ethical concerns [91,128,199-201].

Finally, it is important to note that the use of AI in health care research lends itself to the analysis of quantitative and categorical data, limiting its ability to understand and explain many social and health-related phenomena. The use of race and other contested social categories in AI algorithms often relies on third-party classification in a way that risks misrepresentation [202]. Therefore, although AI may offer insights into the social determinants of health in oncology, such tools do not obviate the need for other methods, including qualitative methods, in cancer research.

Limitations

Our study has several limitations. First, the application of AI in oncology is a rapidly evolving field, and as such, the themes and gaps identified in our scoping review are necessarily provisional. To help mitigate this, we conducted a secondary search 9 months after our initial search, which yielded an

XSL•FO
RenderX

additional 949 abstracts, of which 21 (2.2%) met the inclusion criteria. Despite this rapid evolution, our findings provide insights into the current state of the literature on the impact of AI on health equity in oncology and may also provide a lens for the early integration of AI technologies in health care more generally. Second, we focused our search strategy in the field of oncology and contemporary cancer research; while the themes and gaps highlighted may be illustrative of more general health equity issues arising from the integration of novel technologies in health care at large, there are likely additional themes pertaining to other areas of health care not covered by our review. Finally, our search was limited to records written in English; we were unable to include articles published in other languages, which may bias our findings toward research conducted in and themes prevalent in the English-speaking world; further work could involve a team of multilingual researchers to shed light on themes from non–English-language research literature.

Conclusions

In conclusion, we conducted a scoping review to characterize and assess the literature on the impact of AI on health equity in oncology. Our analysis identified 3 general themes related to how AI can be used to address health disparities, how bias might be mitigated or exacerbated by AI algorithms, and how AI can help investigate the social determinants of health. Our review also identified several gaps and areas in need of further research. These include fostering greater collaboration between HICs and LMICs in the design of AI technologies, ensuring representation in training data sets, considering the context of algorithmic development and application to mitigate bias, and recognizing ethical and methodological issues arising from the use of AI to investigate the determinants of cancer outcomes. As AI applications in oncology continue to expand, attention to these issues will be critical to prevent harm and ensure equitable distribution of the potential benefits of these technologies.

Acknowledgments

The authors would like to thank AMS Healthcare for their funding of this project, and Dr Luke Stark for his valuable comments on the earlier draft of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy.
[DOCX File , 15 KB - jmir_v24i11e39748_app1.docx]

Multimedia Appendix 2
Theme 1 articles: artificial intelligence to address health disparities.
[DOCX File , 23 KB - jmir_v24i11e39748_app2.docx]

Multimedia Appendix 3
Theme 2 articles: artificial intelligence and bias.
[DOCX File , 16 KB - jmir_v24i11e39748_app3.docx]

Multimedia Appendix 4
Theme 3 articles: artificial intelligence and determinants of health.
[DOCX File , 23 KB - jmir_v24i11e39748_app4.docx]

Multimedia Appendix 5
Multiple-theme articles.
[DOCX File , 14 KB - jmir_v24i11e39748_app5.docx]

References

1. Bianchi P, Labory S. The fourth industrial revolution. In: Industrial Policy for the Manufacturing Revolution. Cheltenham, United Kingdom: Edward Elgar Publishing; Jun 29, 2018.
2. Chin-Yee B, Upshur R. The impact of artificial intelligence on clinical judgment: a briefing document. AMS Healthcare. 2019. URL: https://www.ams-inc.on.ca/wp-content/uploads/2020/02/The-Impact-of-AI-on-clinical-judgement.pdf [accessed 2022-03-06]
3. Chui M, Manyika J, Bughin J. Big Data's Potential for Businesses. McKinsey Global Institute. 2011. URL: https://www.mckinsey.com/mgi/overview/in-the-news/big-data-potential-for-businesses [accessed 2022-03-05]

XSL•FO
RenderX

4. Kann BH, Hosny A, Aerts HJ. Artificial intelligence for clinical oncology. Cancer Cell 2021 Jul 12;39(7):916-927 [FREE Full text] [doi: 10.1016/j.ccell.2021.04.002] [Medline: 33930310]
5. Luchini C, Pea A, Scarpa A. Artificial intelligence in oncology: current applications and future perspectives. Br J Cancer 2022 Jan 26;126(1):4-9 [FREE Full text] [doi: 10.1038/s41416-021-01633-1] [Medline: 34837074]
6. Niazi MK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol 2019 May;20(5):e253-e261 [FREE Full text] [doi: 10.1016/S1470-2045(19)30154-8] [Medline: 31044723]
7. Dlamini Z, Francies FZ, Hull R, Marima R. Artificial intelligence (AI) and big data in cancer and precision oncology. Comput Struct Biotechnol J 2020;18:2300-2311 [FREE Full text] [doi: 10.1016/j.csbj.2020.08.019] [Medline: 32994889]
8. Braveman P, Gruskin S. Defining equity in health. J Epidemiol Community Health 2003 Apr;57(4):254-258 [FREE Full text] [doi: 10.1136/jech.57.4.254] [Medline: 12646539]
9. O'Neil C. Weapons of Math Destruction How Big Data Increases Inequality and Threatens Democracy. New York, United States: Penguin Books Limited; 2016.
10. Topol E. Deep Medicine How Artificial Intelligence Can Make Healthcare Human Again. New York, United States: Basic Books; 2019.
11. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
12. Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. Nat Med 2020 Jan 13;26(1):16-17. [doi: 10.1038/s41591-019-0649-2] [Medline: 31932779]
13. Chen I, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? AMA J Ethics 2019 Feb 01;21(2):E167-E179 [FREE Full text] [doi: 10.1001/amajethics.2019.167] [Medline: 30794127]
14. Pfohl S, Marafino B, Coulet A, Rodriguez F, Palaniappan L, Shah N. Creating fair models of atherosclerotic cardiovascular disease risk. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. 2019 Presented at: AIES '19: AAAI/ACM Conference on AI, Ethics, and Society; Jan 27 - 28, 2019; Honolulu HI USA. [doi: 10.1145/3306618.3314278]
15. Abebe R, Hill S, Vaughan JW, Small P, Schwartz HA. Using search queries to understand health information needs in Africa. arXiv 2019. [doi: 10.48550/arXiv.1806.05740]
16. Fenech M, Strukelj N, Buston O. Ethical, social, and political challenges of artificial intelligence in health. Future Advocacy. 2018. URL: https://wellcome.org/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf [accessed 2021-12-17]
17. Patel MI, Lopez AM, Blackstock W, Reeder-Hayes K, Moushey EA, Phillips J, et al. Cancer disparities and health equity: a policy statement from the American society of clinical oncology. J Clin Oncol 2020 Oct 10;38(29):3439-3448. [doi: 10.1200/jco.20.00642]
18. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Oct 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
19. 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]
20. United Nations Department of Economic and Social Affairs. Statistical annex. In: World Economic Situation and Prospects. New York, United States: United Nations; 2014.
21. Hussain M, Figueiredo M, Tran B, Su Z, Molldrem S, Eikey EV, et al. A scoping review of qualitative research in JAMIA: past contributions and opportunities for future work. J Am Med Inform Assoc 2021 Feb 15;28(2):402-413 [FREE Full text] [doi: 10.1093/jamia/ocaa179] [Medline: 33225361]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. DeStephano C, Bakkum-Gamez J, Kaunitz A, Ridgeway J, Sherman M. Intercepting endometrial cancer: opportunities to expand access using new technology. Cancer Prev Res (Phila) 2020 Jul;13(7):563-568. [doi: 10.1158/1940-6207.CAPR-19-0556] [Medline: 32047026]
24. Choudhary T, Mishra V, Goswami A, Sarangapani J. A transfer learning with structured filter pruning approach for improved breast cancer classification on point-of-care devices. Comput Biol Med 2021 Jul;134:104432. [doi: 10.1016/j.compbiomed.2021.104432] [Medline: 33964737]
25. Zhang Y. Ep1.17-02 artificial intelligence in the qualitative study of pulmonary nodules by analyzing the genetic map and imaging data of lung adenocarcinoma. J Thorac Oncol 2019 Oct;14(10):S1083-S1084. [doi: 10.1016/j.jtho.2019.08.2412]
26. Snuderl M. Abstract IA-19: machine learning and AI in molecular pathology diagnostics and clinical management of cancer. Clin Cancer Res 2021;27(5_Supplement):IA-19. [doi: 10.1158/1557-3265.ADI21-IA-19]
27. Tunthanathip T, Oearsakul T. Machine learning approaches for prognostication of newly diagnosed glioblastoma. Int J Nutr Pharmacol Neurol Dis 2021;11(1):57. [doi: 10.4103/ijnpnd.ijnpnd_93_20]
28. Rocha HA, Emani S, Arruda CA, Rizvi R, Garabedian P, Machado de Aquino C, et al. Non-user physician perspectives about an oncology clinical decision-support system: a qualitative study. J Clin Oncol 2020 May 20;38(15_suppl):e14061. [doi: 10.1200/jco.2020.38.15_suppl.e14061]

XSL•FO
RenderX

29. Mohamed N, van de Goor R, El-Sheikh M, Elrayah O, Osman T, Nginamau ES, et al. Feasibility of a portable electronic nose for detection of oral squamous cell carcinoma in Sudan. Healthcare (Basel) 2021 May 03;9(5):534 [FREE Full text] [doi: 10.3390/healthcare9050534] [Medline: 34063592]
30. James BL, Sunny SP, Heidari AE, Ramanjinappa RD, Lam T, Tran AV, et al. Validation of a point-of-care optical coherence tomography device with machine learning algorithm for detection of oral potentially malignant and malignant lesions. Cancers (Basel) 2021 Jul 17;13(14):3583 [FREE Full text] [doi: 10.3390/cancers13143583] [Medline: 34298796]
31. Ningrum DN, Yuan S, Kung W, Wu C, Tzeng I, Huang C, et al. Deep learning classifier with patient's metadata of dermoscopic images in malignant melanoma detection. J Multidisciplinary Healthcare 2021 Apr;14:877-885. [doi: 10.2147/jmdh.s306284]
32. Tanriver G, Soluk Tekkesin M, Ergen O. Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders. Cancers (Basel) 2021 Jul 02;13(11):2766 [FREE Full text] [doi: 10.3390/cancers13112766] [Medline: 34199471]
33. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. J Oral Pathol Med 2021 Oct 16;50(9):911-918. [doi: 10.1111/jop.13227] [Medline: 34358372]
34. Chen B, Lu M, Lipkova J, Mahmood F. Abstract PR-01: real-time, point-of-care pathology diagnosis via embedded deep learning. Clin Cancer Res 2021;27((5_Supplement)):PR-01. [doi: 10.1158/1557-3265.ADI21-PR-01]
35. Jin L, Tang Y, Wu Y, Coole JB, Tan MT, Zhao X, et al. Deep learning extended depth-of-field microscope for fast and slide-free histology. Proc Natl Acad Sci U S A 2020 Dec 29;117(52):33051-33060 [FREE Full text] [doi: 10.1073/pnas.2013571117] [Medline: 33318169]
36. Wadhawan T, Situ N, Lancaster K, Yuan X, Zouridakis G. SkinScan©: a portable library for melanoma detection on handheld devices. In: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2011 Presented at: IEEE International Symposium on Biomedical Imaging: From Nano to Macro; Mar 30 - Apr 02, 2011; Chicago, IL, USA. [doi: 10.1109/isbi.2011.5872372]
37. Tan MC, Bhushan S, Quang T, Schwarz R, Patel KH, Yu X, et al. Automated software-assisted diagnosis of esophageal squamous cell neoplasia using high-resolution microendoscopy. Gastrointest Endosc 2021 Apr;93(4):831-8.e2 [FREE Full text] [doi: 10.1016/j.gie.2020.07.007] [Medline: 32682812]
38. Cerrato TR. Use of artificial intelligence to improve access to initial leukemia diagnosis in low- and middle-income countries. J Clin Oncol 2020 May 20;38(15_suppl):e14117. [doi: 10.1200/jco.2020.38.15_suppl.e14117]
39. Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanapragasam VJ. Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the Surveillance, Epidemiology, and End Results (SEER) database. Lancet Digital Health 2021 Mar;3(3):e158-e165. [doi: 10.1016/s2589-7500(20)30314-9]
40. Matin RN, Dinnes J. AI-based smartphone apps for risk assessment of skin cancer need more evaluation and better regulation. Br J Cancer 2021 May 19;124(11):1749-1750 [FREE Full text] [doi: 10.1038/s41416-021-01302-3] [Medline: 33742148]
41. Khoury M, Engelgau M, Chambers D, Mensah G. Beyond public health genomics: can big data and predictive analytics deliver precision public health? Public Health Genomics 2018 Jul 17;21(5-6):244-250 [FREE Full text] [doi: 10.1159/000501465] [Medline: 31315115]
42. Guo LN, Lee MS, Kassamali B, Mita C, Nambudiri VE. Bias in, bias out: underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection-A scoping review. J Am Acad Dermatol 2022 Jul;87(1):157-159. [doi: 10.1016/j.jaad.2021.06.884] [Medline: 34252465]
43. Chen Z, Chen S, Liang R, Peng Z, Shen J, Zhu W, et al. Can artificial intelligence support the clinical decision making for Barcelona clinic liver cancer stage 0/a hepatocellular carcinoma in China? J Clin Oncol 2019 May 20;37(15_suppl):e15634. [doi: 10.1200/jco.2019.37.15_suppl.e15634]
44. Bradley A, van der Meer R, McKay C. Personalized pancreatic cancer management: a systematic review of how machine learning is supporting decision-making. Pancreas 2019;48(5):598-604. [doi: 10.1097/MPA.0000000000001312] [Medline: 31090660]
45. Pramesh CS, Sirohi B, Nag SM, Gupta S, Anderson BO, Feldman NR, et al. Prospective study of an AI enabled online intervention to increase delivery of guideline compliant cancer care, on the ground. J Clin Oncol 2020 May 20;38(15_suppl):2011. [doi: 10.1200/jco.2020.38.15_suppl.2011]
46. Chen Z, Cao B, Edwards A, Deng H, Zhang K. A deep imputation and inference framework for estimating personalized and race-specific causal effects of genomic alterations on PSA. J Bioinform Comput Biol 2021 Jul 02;19(04):2150016. [doi: 10.1142/s0219720021500165]
47. Pino L, Triana I, Mejia J, Camelo M, Galvez-Nino M, Ruiz R, et al. P09.14 predictive analytics in real-world data from Peru: the new models for personalized oncology. J Thoracic Oncol 2021 Mar;16(3):S294. [doi: 10.1016/j.jtho.2021.01.442]
48. Das AK, Gopalan SS. Prevalence and predictors of routine prostate-specific antigen screening in Medicare beneficiaries in the USA: retrospective cohort analysis using machine learning. Open Public Health J 2019 Dec 31;12(1):521-531. [doi: 10.2174/1874944501912010521]
49. Sultana N. Predicting sun protection measures against skin diseases using machine learning approaches. J Cosmet Dermatol 2022 Feb 13;21(2):758-769. [doi: 10.1111/jocd.14120] [Medline: 33786953]

XSL•FO
RenderX

50. Sim J, Kim YA, Kim JH, Lee JM, Kim MS, Shim YM, et al. The major effects of health-related quality of life on 5-year survival prediction among lung cancer survivors: applications of machine learning. Sci Rep 2020 Jul 01;10(1):10693 [FREE Full text] [doi: 10.1038/s41598-020-67604-3] [Medline: 32612283]
51. Mahmood N, Shahid S, Bakhshi T, Riaz S, Ghufran H, Yaqoob M. Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. Med Biol Eng Comput 2020 Nov;58(11):2631-2640 [FREE Full text] [doi: 10.1007/s11517-020-02245-2] [Medline: 32840766]
52. Hu C, Li Q, Shou J, Zhang F, Li X, Wu M, et al. Constructing a predictive model of depression in chemotherapy patients with non-Hodgkin's lymphoma to improve medical staffs' psychiatric care. Biomed Res Int 2021 Jul 17;2021:9201235-9201212 [FREE Full text] [doi: 10.1155/2021/9201235] [Medline: 34337060]
53. Zhu VJ, Lenert LA, Bunnell BE, Obeid JS, Jefferson M, Halbert CH. Automatically identifying social isolation from clinical narratives for patients with prostate Cancer. BMC Med Inform Decis Mak 2019 Mar 14;19(1):43 [FREE Full text] [doi: 10.1186/s12911-019-0795-y] [Medline: 30871518]
54. Cirstea D, Fillmore N, Yameen H, Yellapragada S, Ifeorah C, Do N, et al. Abstract 1569: racial differences in incidence and impact of TP53 deletion on outcome in African American and Caucasian veterans with multiple myeloma. Cancer Res 2019;79(13_Supplement):1569. [doi: 10.1158/1538-7445.AM2019-1569]
55. Hassoon A, Baig Y, Naimann D, Celentano D, Lansey D, Stearns V, et al. Abstract 54: addressing cardiovascular health using artificial intelligence: randomized clinical trial to increase physical activity in cancer survivors using intelligent voice assist (amazon alexa) for patient coaching. Circulation 2020 Mar 03;141(Suppl_1):A54. [doi: 10.1161/circ.141.suppl_1.54]
56. Kaplan B, Halmos P, Newberg J, Sokol E, Montesion M, Albacker L, et al. Abstract 2337: pan-cancer analysis of sex differences and their associations with ancestry and genomic biomarkers in a large comprehensive genomic profiling dataset. Cancer Res 2020;80(16_Supplement):2337. [doi: 10.1158/1538-7445.AM2020-2337]
57. Shew M, New J, Bur AM. Machine learning to predict delays in adjuvant radiation following surgery for head and neck cancer. Otolaryngol Head Neck Surg 2019 Jun 29;160(6):1058-1064. [doi: 10.1177/0194599818823200] [Medline: 30691352]
58. Sha S, Du W, Parkinson A, Glasgow N. Relative importance of clinical and sociodemographic factors in association with post-operative in-hospital deaths in colorectal cancer patients in New South Wales: an artificial neural network approach. J Eval Clin Pract 2020 Oct 16;26(5):1389-1398. [doi: 10.1111/jep.13318] [Medline: 31733029]
59. Galadima H, Adunlin G, Blando J. Abstract A006: multi-modal estimation of causal influences of environmental agents on colorectal cancer in an understudied population. Cancer Epidemiol Biomarkers Prev 2020;29(6_Supplement_2):A006. [doi: 10.1158/1538-7755.DISP19-A006]
60. Cherry DR, Chen Q, Murphy JD. A novel prediction model to identify patients with early-stage pancreatic cancer. J Clin Oncol 2020 May 20;38(15_suppl):e16801. [doi: 10.1200/jco.2020.38.15_suppl.e16801]
61. Dillon ST, Bhasin MK, Feng X, Koh DW, Daoud SS. Quantitative proteomic analysis in HCV-induced HCC reveals sets of proteins with potential significance for racial disparity. J Transl Med 2013 Oct 01;11(1):239 [FREE Full text] [doi: 10.1186/1479-5876-11-239] [Medline: 24283668]
62. Ding X, Tsang S, Ng S, Xue H. Application of machine learning to development of copy number variation-based prediction of cancer risk. Genomics Insights 2014 Jun 26;7:GEI.S15002. [doi: 10.4137/gei.s15002]
63. Tran P, Monlezun D, De Sirkar S, Iliescu G, Kim P, Lopez-Mattei J, et al. Takotsubo cardiomyopathy mortality and costs in cancer, pci, and racial disparities: propensity score adjusted machine learning guided analysis of 6+ million inpatient admissions. J Cardiac Failure 2019 Aug;25(8):S60. [doi: 10.1016/j.cardfail.2019.07.171]
64. He J, Zhang J, Chen C, Ma Y, De Guzman R, Meng J, et al. The relative importance of clinical and socio-demographic variables in prognostic prediction in non-small cell lung cancer: a variable importance approach. Med Care 2020 May;58(5):461-467. [doi: 10.1097/MLR.0000000000001288] [Medline: 31985586]
65. Benci JL, Vachani CC, Hampshire MK, Bach C, Arnold-Korzeniowski K, Metz JM, et al. Factors influencing delivery of cancer survivorship care plans: a national patterns of care study. Front Oncol 2019 Jan 31;9:1577 [FREE Full text] [doi: 10.3389/fonc.2019.01577] [Medline: 32083018]
66. Liao Y, Li C, Xia C, Zheng R, Xu B, Zeng H, et al. Spatial distribution of esophageal cancer mortality in China: a machine learning approach. Int Health 2021 Jan 14;13(1):70-79 [FREE Full text] [doi: 10.1093/inthealth/ihaa022] [Medline: 32478387]
67. Fiano R, Merrick G, Innes K, LeMasters T, Mattes M, Shen C, et al. PCN159 prediction of low-VALUE cancer care among older men with low-risk prostate cancer: a machine learning approach. Value in Health 2021 Jun;24:S49. [doi: 10.1016/j.jval.2021.04.251]
68. Gajra A, Zettler ME, Miller KA, Blau S, Venkateshwaran SS, Sridharan S, et al. Augmented intelligence to predict 30-day mortality in patients with cancer. Future Oncol 2021 Oct;17(29):3797-3807 [FREE Full text] [doi: 10.2217/fon-2021-0302] [Medline: 34189965]
69. Raffenaud A, Gurupur V, Fernandes SL, Yeung T. Utilizing telemedicine in oncology settings: patient favourability rates and perceptions of use analysis using Chi-Square and neural networks. Technol Health Care 2019 Mar 06;27(2):115-127. [doi: 10.3233/thc-181293]
70. Muhammad W, Hart G, Nartowt B, Deng J. In silico simulation to quantify liver cancer risk with smoking. Int J Radiation Oncol Biol Physics 2019 Sep;105(1):E137. [doi: 10.1016/j.ijrobp.2019.06.2175]

XSL•FO
RenderX

71. Pino LE, Large E, Mejía J, Triana IC. MAIA (Medical Artificial Intelligence Assistant) as interface for a new cancer healthcare integrative platform. J Global Oncol 2019 Oct 07;5(suppl):25. [doi: 10.1200/jgo.2019.5.suppl.25]
72. Hanson HA, Martin C, O'Neil B, Leiser CL, Mayer EN, Smith KR, et al. The relative importance of race compared to health care and social factors in predicting prostate cancer mortality: a random forest approach. J Urol 2019 Dec;202(6):1209-1216. [doi: 10.1097/ju.0000000000000416]
73. Aghdam N, Arab A, Kumar D, Suy S, Dritschilo A, Lynch JJ, et al. Accessibility and utilization of stereotactic body radiation therapy for localized prostate cancer: an analysis of geodemographic clusters. J Clin Oncol 2018 May 20;36(15_suppl):e18636. [doi: 10.1200/jco.2018.36.15_suppl.e18636]
74. Ramakrishnan S, Xuan P, Qi Q, Hu Q, Ellman E, Azabdaftari G, et al. Abstract A003: epigenetic alterations as potential biologic determinants of racial health disparities. Cancer Res 2018;78(16_Supplement):A003. [doi: 10.1158/1538-7445.PRCA2017-A003]
75. Juacaba S, Rocha HA, Meneleu P, Hekmat R, Felix W, Arriaga YE, et al. A retrospective evaluation of treatment decision making for thyroid cancer using clinical decision support in Brazil. J Clin Oncol 2020 May 20;38(15_suppl):e19193. [doi: 10.1200/jco.2020.38.15_suppl.e19193]
76. Cheng B, Joe Stanley R, Stoecker WV, Stricklin SM, Hinton KA, Nguyen TK, et al. Analysis of clinical and dermoscopic features for basal cell carcinoma neural network classification. Skin Res Technol 2013 Feb 22;19(1):e217-e222 [FREE Full text] [doi: 10.1111/j.1600-0846.2012.00630.x] [Medline: 22724561]
77. Urman A, Wang C, Dankwa-Mullan I, Scheinberg E, Young MJ. Harnessing AI for health equity in oncology research and practice. J Clin Oncol 2018 Oct 20;36(30_suppl):67. [doi: 10.1200/jco.2018.36.30_suppl.67]
78. Greatbatch O, Garrett A, Snape K. The impact of artificial intelligence on the current and future practice of clinical cancer genomics. Genet Res 2019 Oct 31;101:E9. [doi: 10.1017/s0016672319000089]
79. Lee J, Estevez M, Segal B, Sondhi A, Cohen A, Cherng S. Ai1 quantifying bias in ML-extracted variables for inference in clinical oncology. Value Health 2021 Jun;24:S1. [doi: 10.1016/j.jval.2021.04.006]
80. Agrawal N, Monlezun D, Grable C, Graham J, Marmagkiolis K, Chauhan S, et al. Abstract 13212: chronic total occlusion racial and income inequities by mortality and cost: propensity score and machine learning augmented nationally representative case-control study of 30 million hospitalizations. Circulation 2021 Nov 16;144(Suppl_1):A14107. [doi: 10.1161/circ.144.suppl_1.13212]
81. Kim J, Monlezun D, Palaskas N, Cilingirolu M, Marmagkiolis K, Iliescu C. Abstract A-8: Sex disparities in cardio-oncology treatment and mortality: Propensity score nationally representative case-control analysis with machine learning augmentation of over 30 million hospitalizations. https://onlinelibrary.wiley.com/doi/10.1002/ccd.29644 2021 Apr 28:S9-S111. [doi: 10.21203/rs.3.rs-1705689/v1]
82. Giannitrapani K, Walling A, Gamboa R, O'Hanlon C, Canning M, Lindvall C, et al. Getting ahead of bias: qualitative pre-work for developing AI capture of palliative and end-of life quality measures. Harvard Medical School. 2020. URL: https://tinyurl.com/y6sytvkv [accessed 2022-09-13]
83. Schlemmer HP. A40 Radiomics and AI: future of cancer imaging without radiologists? In: Proceedings of the International Cancer Imaging Society (ICIS) 18th Annual Teaching Course. 2018 Sep 3 Presented at: International Cancer Imaging Society (ICIS) 18th Annual Teaching Course; 7-9 October 2018; Menton, France.
84. Asiedu M, Guillermo S, Ramanujam N. Low-cost, speculum-free, automated cervical cancer screening: bringing expert colposcopy assessment to community health. Annals Global Health 2017;83(1):199. [doi: 10.1016/j.aogh.2017.03.494]
85. Holmström O, Linder N, Kaingu H, Mbuuko N, Mbete J, Kinyua F, et al. Point-of-care digital cytology with artificial intelligence for cervical cancer screening in a resource-limited setting. JAMA Netw Open 2021 Mar 01;4(3):e211740 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.1740] [Medline: 33729503]
86. Kudva V, Prasad K, Guruvare S. Andriod device-based cervical cancer screening for resource-poor settings. J Digit Imaging 2018 Oct 18;31(5):646-654 [FREE Full text] [doi: 10.1007/s10278-018-0083-x] [Medline: 29777323]
87. Hu L, Horning M, Banik D, Ajenifuja OK, Adepiti CA, Yeates K, Mehanian. Deep learning-based image evaluation for cervical precancer screening with a smartphone targeting low resource settings – Engineering approach. In: Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020 Presented at: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 20-24, 2020; Montreal, QC, Canada. [doi: 10.1109/embc44109.2020.9175863]
88. Xue Z, Novetsky AP, Einstein MH, Marcus JZ, Befano B, Guo P, et al. A demonstration of automated visual evaluation of cervical images taken with a smartphone camera. Int J Cancer 2020 Nov 01;147(9):2416-2423 [FREE Full text] [doi: 10.1002/ijc.33029] [Medline: 32356305]
89. Bae JK, Roh H, You JS, Kim K, Ahn Y, Askaruly S, et al. Quantitative screening of cervical cancers for low-resource settings: pilot study of smartphone-based endoscopic visual inspection after acetic acid using machine learning techniques. JMIR Mhealth Uhealth 2020 Mar 11;8(3):e16467 [FREE Full text] [doi: 10.2196/16467] [Medline: 32159521]
90. Ajenifuja KO, Belinson J, Goldstein A, Desai KT, de Sanjose S, Schiffman M. Designing low-cost, accurate cervical screening strategies that take into account COVID-19: a role for self-sampled HPV typing2. Infect Agent Cancer 2020 Oct 14;15(1):61 [FREE Full text] [doi: 10.1186/s13027-020-00325-4] [Medline: 33072179]

XSL•FO
RenderX

91. Xue P, Ng MT, Qiao Y. The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence. BMC Med 2020 Jun 03;18(1):169 [FREE Full text] [doi: 10.1186/s12916-020-01613-x] [Medline: 32493320]
92. Yang Z, Francisco J, Reese AS, Spriggs DR, Im H, Castro CM. Addressing cervical cancer screening disparities through advances in artificial intelligence and nanotechnologies for cellular profiling. Biophys Rev (Melville) 2021 Mar;2(1):011303 [FREE Full text] [doi: 10.1063/5.0043089] [Medline: 33842926]
93. Castro C, Im H, Lee H, Avila-Wallace M, Weissleder R, Randall T. Harnessing artificial intelligence and digital diffraction to advance point-of-care HPV 16 and 18 detection. Gynecologic Oncol 2019 Jun;154:38. [doi: 10.1016/j.ygyno.2019.04.090]
94. Asiedu M, Skerrett E, Sapiro G, Ramanujam N. Combining multiple contrasts for improving machine learning-based classification of cervical cancers with a low-cost point-of-care Pocket colposcope. In: Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020 Presented at: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 20-24, 2020; Montreal, QC, Canada. [doi: 10.1109/embc44109.2020.9175858]
95. Parra S, Carranza E, Coole J, Hunt B, Smith C, Keahey P, et al. Development of low-cost point-of-care technologies for cervical cancer prevention based on a single-board computer. IEEE J Transl Eng Health Med 2020;8:1-10. [doi: 10.1109/jtehm.2020.2970694]
96. Asiedu MN, Simhal A, Chaudhary U, Mueller JL, Lam CT, Schmitt JW, et al. Development of algorithms for automated detection of cervical pre-cancers with a low-cost, point-of-care, pocket colposcope. IEEE Trans Biomed Eng 2019 Aug;66(8):2306-2318. [doi: 10.1109/tbme.2018.2887208]
97. Hunt B, Fregnani JH, Brenes D, Schwarz RA, Salcedo MP, Possati-Resende JC, et al. Cervical lesion assessment using real-time microendoscopy image analysis in Brazil: the CLARA study. Int J Cancer 2021 Jul 15;149(2):431-441 [FREE Full text] [doi: 10.1002/ijc.33543] [Medline: 33811763]
98. Kisling K, Zhang L, Simonds H, Fakie N, Yang J, McCarroll R, et al. Fully automatic treatment planning for external-beam radiation therapy of locally advanced cervical cancer: a tool for low-resource clinics. J Global Oncol 2019 Dec(5):1-9. [doi: 10.1200/jgo.18.00107]
99. Azarianpour Esfahani S, Fu P, Mahdi H, Madabhushi A. Computational features of TIL architecture are differentially prognostic of uterine cancer between African and Caucasian American women. J Clin Oncol 2021 May 20;39(15_suppl):5585. [doi: 10.1200/jco.2021.39.15_suppl.5585]
100. Asadi F, Salehnasab C, Ajori L. Supervised algorithms of machine learning for the prediction of cervical cancer. J Biomed Phys Eng 2020 Aug;10(4):513-522 [FREE Full text] [doi: 10.31661/jbpe.v0i0.1912-1027] [Medline: 32802799]
101. Tossas K, Khan J, Winn R. Abstract A010: Hidden figures - an example of using machine learning to prioritize cervical cancer screening outreach. Cancer Epidemiol Biomarkers Prev 2020;29((6_Supplement_2)):A010. [doi: 10.1158/1538-7755.DISP19-A010]
102. Love SM, Berg WA, Podilchuk C, López Aldrete AL, Gaxiola Mascareño AP, Pathicherikollamparambil K, et al. Palpable breast lump triage by minimally trained operators in Mexico using computer-assisted diagnosis and low-cost ultrasound. J Global Oncol 2018 Dec(4):1-9. [doi: 10.1200/jgo.17.00222]
103. Min J, Im H, Allen M, McFarland PJ, Degani I, Yu H, et al. Computational optics enables breast cancer profiling in point-of-care settings. ACS Nano 2018 Sep 25;12(9):9081-9090 [FREE Full text] [doi: 10.1021/acsnano.8b03029] [Medline: 30113824]
104. Bakre M, Ramkumar C, Basavaraj C, Attuluri A, Madhav L, Prakash C, et al. Abstract P3-08-10: development and validation of a broad-based second generation multi marker “Morphometric IHC” test for optimal treatment planning of stage 1 and 2 breast cancer patients in low resource settings. Cancer Res 2018;78((4_Supplement)):P3-08-10. [doi: 10.1158/1538-7445.SABCS17-P3-08-10]
105. Lehman C, Yala A, Lamb L, Barzilay R. Abstract SP080: hidden clues in the mammogram: how AI can improve early breast cancer detection. Cancer Res 2021;81((4_Supplement)):SP080. [doi: 10.1158/1538-7445.sabcs20-sp80]
106. Lehman C. Abstract IS-3: breast imaging in resource constrained regions: lessons from Uganda. Cancer Res 2018;78((4_Supplement)):IS-3. [doi: 10.1158/1538-7445.SABCS17-IS-3]
107. Cobb AN, Janjua HM, Kuo PC. Big data solutions for controversies in breast cancer treatment. Clin Breast Cancer 2021 Jun;21(3):e199-e203. [doi: 10.1016/j.clbc.2020.08.003] [Medline: 32933862]
108. Thrall JH, Fessell D, Pandharipande PV. Rethinking the approach to artificial intelligence for medical image analysis: the case for precision diagnosis. J Am Coll Radiol 2021 Jan;18(1 Pt B):174-179. [doi: 10.1016/j.jacr.2020.07.010] [Medline: 33413896]
109. Hou C, Zhong X, He P, Xu B, Diao S, Yi F, et al. Predicting breast cancer in Chinese women using machine learning techniques: algorithm development. JMIR Med Inform 2020 Jun 08;8(6):e17364 [FREE Full text] [doi: 10.2196/17364] [Medline: 32510459]
110. Sidey-Gibbons C, Pfob A, Asaad M, Boukovalas S, Lin Y, Selber JC, et al. Development of machine learning algorithms for the prediction of financial toxicity in localized breast cancer following surgical treatment. JCO Clin Cancer Inform 2021 Dec(5):338-347. [doi: 10.1200/cci.20.00088]

XSL•FO
RenderX

111. Sidey-Gibbons C, Asaad M, Pfob A, Boukovalas S, Lin Y, Offodile A. Machine learning algorithms to predict financial toxicity associated with breast cancer treatment. J Clin Oncol 2020 May 20;38(15_suppl):2047. [doi: 10.1200/jco.2020.38.15_suppl.2047]
112. Zhong X, Luo T, Deng L, Liu P, Hu K, Lu D, et al. Multidimensional machine learning personalized prognostic model in an early invasive breast cancer population-based cohort in china: algorithm validation study. JMIR Med Inform 2020 Nov 09;8(11):e19069 [FREE Full text] [doi: 10.2196/19069] [Medline: 33164899]
113. Wheeler SB, Spees L, Biddell CB, Rotter J, Trogdon JG, Birken SA, et al. Development of a personalized follow-up care algorithm for Medicare breast cancer survivors. J Clin Oncol 2020 Oct 10;38(29_suppl):204. [doi: 10.1200/jco.2020.38.29_suppl.204]
114. Lehman C. Abstract IA-21: AI in an imaging center: challenges and opportunities. Clin Cancer Res 2021;27((5_Supplement)):IA-21. [doi: 10.1158/1557-3265.ADI21-IA-21]
115. Yang X, Amgad M, Cooper LA, Du Y, Fu H, Ivanov AA. High expression of MKK3 is associated with worse clinical outcomes in African American breast cancer patients. J Transl Med 2020 Sep 01;18(1):334 [FREE Full text] [doi: 10.1186/s12967-020-02502-w] [Medline: 32873298]
116. Mema E, McGinty G. The role of artificial intelligence in understanding and addressing disparities in breast cancer outcomes. Curr Breast Cancer Rep 2020 May 18;12(3):168-174. [doi: 10.1007/s12609-020-00368-x]
117. Zhang J, Chen Z, Wu J, Liu K. An intelligent decision-making support system for the detection and staging of prostate cancer in developing countries. Comput Math Methods Med 2020 Aug 17;2020:5363549-5363518 [FREE Full text] [doi: 10.1155/2020/5363549] [Medline: 32879636]
118. Uthoff RD, Song B, Sunny S, Patrick S, Suresh A, Kolur T, et al. Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities. PLoS One 2018 Dec 5;13(12):e0207493 [FREE Full text] [doi: 10.1371/journal.pone.0207493] [Medline: 30517120]
119. Song B, Sunny S, Li S, Gurushanth K, Mendonca P, Mukhia N, et al. Mobile-based oral cancer classification for point-of-care screening. J Biomed Opt 2021 Jun 1;26(06):065003. [doi: 10.1117/1.jbo.26.6.065003]
120. Rock L, Datta M, Laronde D, Carraro A, Korbelik J, Harrison A, et al. Abstract 4223: conducting community oral cancer screening among South Asians in British Columbia. Cancer Res 2019;79((13_Supplement)):4223. [doi: 10.1158/1538-7445.AM2019-4223]
121. Adams SJ, Mondal P, Penz E, Tyan C, Lim H, Babyn P. Development and cost analysis of a lung nodule management strategy combining artificial intelligence and lung-RADS for baseline lung cancer screening. J Am Coll Radiol 2021 May;18(5):741-751. [doi: 10.1016/j.jacr.2020.11.014] [Medline: 33482120]
122. Otero J, Garagorry F, Cabell-Izquierdo Y, Thomas D. Neuropathology Decision Support Systems for Resource-Poor Pathologists. Berlin, Heidelberg: Springer Nature; 2020.
123. Quang T, Schwarz RA, Dawsey SM, Tan MC, Patel K, Yu X, et al. A tablet-interfaced high-resolution microendoscope with automated image interpretation for real-time evaluation of esophageal squamous cell neoplasia. Gastrointest Endosc 2016 Dec;84(5):834-841 [FREE Full text] [doi: 10.1016/j.gie.2016.03.1472] [Medline: 27036635]
124. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056]
125. Wu J, Zhuang Q, Tan Y. Auxiliary medical decision system for prostate cancer based on ensemble method. Comput Math Methods Med 2020 May 18;2020:6509596-6509511 [FREE Full text] [doi: 10.1155/2020/6509596] [Medline: 32508976]
126. Briercheck E, Valvert F, Solorzano E, Silva O, Puligandla M, Tala MM, et al. High accuracy, low-cost transcriptional diagnostic to transform lymphoma care in low- and middle-income countries. Blood 2019;134((Supplement_1)):409. [doi: 10.1182/blood-2019-121397]
127. Ilhan B, Lin K, Guneri P, Wilder-Smith P. Improving oral cancer outcomes with imaging and artificial intelligence. J Dent Res 2020 Mar 20;99(3):241-248 [FREE Full text] [doi: 10.1177/0022034520902128] [Medline: 32077795]
128. Ngwa W, Olver I, Schmeler KM. The use of health-related technology to reduce the gap between developed and undeveloped regions around the globe. Am Soc Clin Oncol Educ Book 2020 Mar;40:1-10 [FREE Full text] [doi: 10.1200/EDBK_288613] [Medline: 32223667]
129. Anirvan P, Meher D, Singh SP. Artificial intelligence in gastrointestinal endoscopy in a resource-constrained setting: a reality check. Euroasian J Hepatogastroenterol 2020;10(2):92-97 [FREE Full text] [doi: 10.5005/jp-journals-10018-1322] [Medline: 33511071]
130. Valvert F, Silva O, Solórzano-Ortiz E, Puligandla M, Siliézar Tala MM, Guyon T, et al. Low-cost transcriptional diagnostic to accurately categorize lymphomas in low- and middle-income countries. Blood Adv 2021 May 25;5(10):2447-2455 [FREE Full text] [doi: 10.1182/bloodadvances.2021004347] [Medline: 33988700]
131. Timmerman R, Li B, Sarria G, Zhang Y, Perez T, Jiang S. Covering gaps in radiation oncology through artificial intelligence in low-resource countries: a survey-based analysis. Int J Radiation Oncol Biol Physics 2020 Nov;108(3):e420. [doi: 10.1016/j.ijrobp.2020.07.2489]
132. Lee M, Guo L, Nambudiri V. Towards gender equity in artificial intelligence and machine learning applications in dermatology. J Am Med Inform Assoc 2022 Jan 12;29(2):400-403 [FREE Full text] [doi: 10.1093/jamia/ocab113] [Medline: 34151976]

XSL•FO
RenderX

133. D'Souza G, Zhang HH, D'Souza WD, Meyer RR, Gillison ML. Moderate predictive value of demographic and behavioral characteristics for a diagnosis of HPV16-positive and HPV16-negative head and neck cancer. Oral Oncol 2010 Mar;46(2):100-104 [FREE Full text] [doi: 10.1016/j.oraloncology.2009.11.004] [Medline: 20036610]
134. Howard FM, Dolezal J, Kochanny S, Schulte J, Chen H, Heij L, et al. The impact of site-specific digital histology signatures on deep learning model accuracy and bias. Nat Commun 2021 Jul 20;12(1):4423 [FREE Full text] [doi: 10.1038/s41467-021-24698-1] [Medline: 34285218]
135. Okoji U, Taylor S, Lipoff J. Equity in skin typing: why it is time to replace the Fitzpatrick scale. Br J Dermatol 2021 Jul 22;185(1):198-199. [doi: 10.1111/bjd.19932] [Medline: 33666245]
136. Veronese F, Branciforti F, Zavattaro E, Tarantino V, Romano V, Meiburger KM, et al. The role in teledermoscopy of an inexpensive and easy-to-use smartphone device for the classification of three types of skin lesions using convolutional neural networks. Diagnostics (Basel) 2021 Mar 05;11(3):451 [FREE Full text] [doi: 10.3390/diagnostics11030451] [Medline: 33807976]
137. Pangti R, Mathur J, Chouhan V, Kumar S, Rajput L, Shah S, et al. A machine learning-based, decision support, mobile phone application for diagnosis of common dermatological diseases. J Eur Acad Dermatol Venereol 2021 Mar 12;35(2):536-545. [doi: 10.1111/jdv.16967] [Medline: 32991767]
138. Gao Y, Cui Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. Nat Commun 2020 Oct 12;11(1):5131 [FREE Full text] [doi: 10.1038/s41467-020-18918-3] [Medline: 33046699]
139. Khor S, Hahn E, Haupt E, Shankaran V, Clark S, Rodriguez P, et al. AI2 the impact of including race and ethnicity in risk prediction models on racial bias. Value Health 2021 Jun;24:S1. [doi: 10.1016/j.jval.2021.04.007]
140. Gilson A, Du J, Janda G, Umrao S, Joel M, Choi R, et al. Abstract PO-074: the impact of phenotypic bias in the generalizability of deep learning models in non-small cell lung cancer. Clin Cancer Res 2021;27(5_Supplement):PO-074. [doi: 10.1158/1557-3265.ADI21-PO-074]
141. Li Y, Pang X, Cui Z, Zhou Y, Mao F, Lin Y, et al. Genetic factors associated with cancer racial disparity - an integrative study across twenty-one cancer types. Mol Oncol 2020 Nov 24;14(11):2775-2786 [FREE Full text] [doi: 10.1002/1878-0261.12799] [Medline: 32920960]
142. van Dams R, Kishan A, Nickols N, Raldow A, King C, Chang A, et al. Racial disparity in the genomic basis of radiosensitivity – an exploration of whole-transcriptome sequencing data via a machine-learning approach. Int J Radiation Oncol Biol Physics 2019 Sep;105(1):E138-E139. [doi: 10.1016/j.ijrobp.2019.06.2180]
143. An C, Choi JW, Lee HS, Lim H, Ryu SJ, Chang JH, et al. Prediction of the risk of developing hepatocellular carcinoma in health screening examinees: a Korean cohort study. BMC Cancer 2021 Jul 29;21(1):755 [FREE Full text] [doi: 10.1186/s12885-021-08498-w] [Medline: 34187409]
144. Bibault J, Bassenne M, Ren H, Xing L. Deep learning prediction of cancer prevalence from satellite imagery. Cancers (Basel) 2020 Dec 19;12(12):3844 [FREE Full text] [doi: 10.3390/cancers12123844] [Medline: 33352801]
145. Muhlestein WE, Akagi DS, Chotai S, Chambless LB. The impact of race on discharge disposition and length of hospitalization after craniotomy for brain tumor. World Neurosurg 2017 Aug;104:24-38 [FREE Full text] [doi: 10.1016/j.wneu.2017.04.061] [Medline: 28478245]
146. Muhlestein W, Akagi D, McManus A, Chambless L. Machine learning ensemble models predict total charges and drivers of cost for transsphenoidal surgery for pituitary tumor. J Neurosurg 2018 Sep 21;131(2):507-516. [doi: 10.3171/2018.4.JNS18306] [Medline: 30239321]
147. Aghdam N, Carrasquilla M, Wang E, Pepin AN, Danner M, Ayoob M, et al. Ten-year single institutional analysis of geographic and demographic characteristics of patients treated with stereotactic body radiation therapy for localized prostate cancer. Front Oncol 2020 Feb 25;10:616286 [FREE Full text] [doi: 10.3389/fonc.2020.616286] [Medline: 33718117]
148. Im H, Pathania D, McFarland PJ, Sohani AR, Degani I, Allen M, et al. Design and clinical validation of a point-of-care device for the diagnosis of lymphoma via contrast-enhanced microholography and machine learning. Nat Biomed Eng 2018 Oct 23;2(9):666-674 [FREE Full text] [doi: 10.1038/s41551-018-0265-3] [Medline: 30555750]
149. Kar A, Wreesmann VB, Shwetha V, Thakur S, Rao VU, Arakeri G, et al. Improvement of oral cancer screening quality and reach: the promise of artificial intelligence. J Oral Pathol Med 2020 Oct 28;49(8):727-730. [doi: 10.1111/jop.13013] [Medline: 32162398]
150. Lynch SM, Handorf E, Sorice KA, Blackman E, Bealin L, Giri VN, et al. The effect of neighborhood social environment on prostate cancer development in black and white men at high risk for prostate cancer. PLoS One 2020 Aug 13;15(8):e0237332 [FREE Full text] [doi: 10.1371/journal.pone.0237332] [Medline: 32790761]
151. Manz CR, Chen J, Liu M, Chivers C, Regli SH, Braun J, et al. Validation of a machine learning algorithm to predict 180-day mortality for outpatients with cancer. JAMA Oncol 2020 Dec 01;6(11):1723-1730 [FREE Full text] [doi: 10.1001/jamaoncol.2020.4331] [Medline: 32970131]
152. Uthoff RD, Song B, Sunny S, Patrick S, Suresh A, Kolur T, et al. Small form factor, flexible, dualmodality handheld probe for smartphone-based, point-of-care oral and oropharyngeal cancer screening. J Biomed Opt ;24(06). [doi: 10.1117/1.JBO.24.10.106003]

XSL•FO
RenderX

153. Song B, Sunny S, Uthoff RD, Patrick S, Suresh A, Kolur T, et al. Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning. Biomed Opt Express 2018 Oct 10;9(11):5318-5329. [doi: 10.1364/BOE.9.005318]
154. Ilhan B, Guneri P, Wilder-Smith P. The contribution of artificial intelligence to reducing the diagnostic delay in oral cancer. Oral Oncol 2021 May;116. [doi: 10.1016/j.oraloncology.2021.105254]
155. Ahluwalia KP. Assessing the oral cancer risk of South-Asian immigrants in New York City. Cancer 2005 Dec 15;104(12 Suppl):2959-2961 [FREE Full text] [doi: 10.1002/cncr.21502] [Medline: 16247776]
156. Donia J, Shaw JA. Co-design and ethical artificial intelligence for health: an agenda for critical research and practice. Big Data Soc 2021 Dec 17;8(2):205395172110652. [doi: 10.1177/20539517211065248]
157. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3313831.3376718]
158. Mollura DJ, Culp MP, Pollack E, Battino G, Scheel JR, Mango VL, et al. Artificial intelligence in low- and middle-income countries: innovating global health radiology. Radiology 2020 Dec;297(3):513-520. [doi: 10.1148/radiol.2020201434] [Medline: 33021895]
159. Shaw JA, Donia J. The sociotechnical ethics of digital health: a critique and extension of approaches from bioethics. Front Digit Health 2021 Sep 23;3:725088 [FREE Full text] [doi: 10.3389/fdgth.2021.725088] [Medline: 34713196]
160. Couldry N, Mejias UA. Data colonialism: rethinking big data's relation to the contemporary subject. Television New Media 2018 Sep 02;20(4):336-349. [doi: 10.1177/1527476418796632]
161. Ferryman K. The dangers of data colonialism in precision public health. Glob Policy 2021 Aug 05;12(S6):90-92. [doi: 10.1111/1758-5899.12953]
162. Veale M, Binns R. Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. Big Data Soc 2017 Nov 20;4(2):205395171774353. [doi: 10.1177/2053951717743530]
163. Boddington P. Normative modes: codes and standards. In: The Oxford Handbook of Ethics of AI. Oxford, United Kingdom: Oxford University Press; 2020.
164. Olsen CM, Thompson JF, Pandeya N, Whiteman DC. Evaluation of sex-specific incidence of Melanoma. JAMA Dermatol 2020 May 01;156(5):553-560 [FREE Full text] [doi: 10.1001/jamadermatol.2020.0470] [Medline: 32211827]
165. Kobayashi K, Hisamatsu K, Suzui N, Hara A, Tomita H, Miyazaki T. A review of HPV-related head and neck cancer. J Clin Med 2018 Aug 27;7(9):241 [FREE Full text] [doi: 10.3390/jcm7090241] [Medline: 30150513]
166. Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: a virus-related cancer epidemic. Lancet Oncol 2010 Aug;11(8):781-789. [doi: 10.1016/s1470-2045(10)70017-6]
167. Tognetti L, Bonechi S, Andreini P, Bianchini M, Scarselli F, Cevenini G, et al. A new deep learning approach integrated with clinical data for the dermoscopic differentiation of early melanomas from atypical nevi. J Dermatol Sci 2021 Mar;101(2):115-122. [doi: 10.1016/j.jdermsci.2020.11.009] [Medline: 33358096]
168. Hart GR, Roffman DA, Decker R, Deng J. A multi-parameterized artificial neural network for lung cancer risk prediction. PLoS One 2018 Oct 24;13(10):e0205264 [FREE Full text] [doi: 10.1371/journal.pone.0205264] [Medline: 30356283]
169. Hart GR, Yan V, Huang GS, Liang Y, Nartowt BJ, Muhammad W, et al. Population-based screening for endometrial cancer: human vs machine intelligence. Front Artif Intell 2020 Nov 24;3:539879 [FREE Full text] [doi: 10.3389/frai.2020.539879] [Medline: 33733200]
170. Yaemsiri S, Alfier JM, Moy E, Rossen LM, Bastian B, Bolin J, et al. Healthy people 2020: rural areas lag in achieving targets for major causes of death. Health Aff (Millwood) 2019 Dec 01;38(12):2027-2031 [FREE Full text] [doi: 10.1377/hlthaff.2019.00915] [Medline: 31794308]
171. Vayena E, Feretti A. Big data and artificial intelligence for global health: ethical challenges and opportunities. In: Global Health: Ethical Challenges. Cambridge: Cambridge University Press; 2021.
172. Hosny A, Aerts HJ. Artificial intelligence for global health. Science 2019 Nov 22;366(6468):955-956 [FREE Full text] [doi: 10.1126/science.aay5189] [Medline: 31753987]
173. Gyawali B. Does global oncology need artificial intelligence? Lancet Oncol 2018 May;19(5):599-600. [doi: 10.1016/s1470-2045(18)30269-9]
174. Harrington C, Erete S, Piper AM. Deconstructing community-based collaborative design. Proc ACM Hum Comput Interact 2019 Nov 07;3(CSCW):1-25. [doi: 10.1145/3359318]
175. Olusanya BO, Opoka RO. Obligations under global health partnerships in LMICs should be contractual. Lancet Global Health 2017 Sep;5(9):e869. [doi: 10.1016/s2214-109x(17)30260-7]
176. Kaur N, Figueiredo S, Bouchard V, Moriello C, Mayo N. Where have all the pilot studies gone? A follow-up on 30 years of pilot studies in Clinical Rehabilitation. Clin Rehabil 2017 Oct 01;31(9):1238-1248 [FREE Full text] [doi: 10.1177/0269215517692129] [Medline: 28786333]
177. Mutebi M, Dehar N, Nogueira LM, Shi K, Yabroff KR, Gyawali B. Cancer groundshot: building a robust cancer control platform in addition to launching the cancer moonshot. Am Soc Clin Oncol Educ Book 2022 Jul(42):100-115. [doi: 10.1200/edbk_359521]

XSL•FO
RenderX

178. Gyawali B, Sullivan R, Booth CM. Cancer groundshot: going global before going to the moon. Lancet Oncol 2018 Mar;19(3):288-290. [doi: 10.1016/s1470-2045(18)30076-7]
179. Featherston R, Downie LE, Vogel AP, Galvin KL. Decision making biases in the allied health professions: a systematic scoping review. PLoS One 2020 Oct 20;15(10):e0240716 [FREE Full text] [doi: 10.1371/journal.pone.0240716] [Medline: 33079949]
180. Somashekhar S, Bakre M, Ramkumar C, Basavaraj C, Arun Kumar A, Madhav L, et al. Risk of recurrence prediction and optimum treatment planning for early stage breast cancer patients: a cost-effective, accurate and broad based solution for Asia. Annals Oncol 2017 Sep;28:v587. [doi: 10.1093/annonc/mdx390.047]
181. Tian Y, Liu X, Wang Z, Cao S, Liu Z, Ji Q, et al. Concordance between Watson for oncology and a multidisciplinary clinical decision-making team for gastric cancer and the prognostic implications: retrospective study. J Med Internet Res 2020 Feb 20;22(2):e14122 [FREE Full text] [doi: 10.2196/14122] [Medline: 32130123]
182. Jie Z, Zhiying Z, Li L. A meta-analysis of Watson for Oncology in clinical application. Sci Rep 2021 Mar 11;11(1):5792 [FREE Full text] [doi: 10.1038/s41598-021-84973-5] [Medline: 33707577]
183. Tupasela A, Di Nucci E. Concordance as evidence in the Watson for oncology decision-support system. AI Soc 2020 Feb 01;35(4):811-818. [doi: 10.1007/s00146-020-00945-9]
184. Wajcman J. Automation: is it really different this time? Br J Sociol 2017 Mar 21;68(1):119-127. [doi: 10.1111/1468-4446.12239] [Medline: 28321856]
185. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Med Ethics 2021 Feb 15;22(1):14 [FREE Full text] [doi: 10.1186/s12910-021-00577-8] [Medline: 33588803]
186. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019 Dec;9(2):010318 [FREE Full text] [doi: 10.7189/jogh.09.020318] [Medline: 31788229]
187. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;2:77 [FREE Full text] [doi: 10.1038/s41746-019-0155-4] [Medline: 31453372]
188. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. Patterns (N Y) 2021 Oct 08;2(10):100347 [FREE Full text] [doi: 10.1016/j.patter.2021.100347] [Medline: 34693373]
189. Adamson AS, Smith A. Machine learning and health care disparities in dermatology. JAMA Dermatol 2018 Nov 01;154(11):1247-1248. [doi: 10.1001/jamadermatol.2018.2348] [Medline: 30073260]
190. Silberg J, Manyika J. Notes from the AI frontier: Tackling bias in AI (and in humans). McKinsey Global Institute. 2019. URL: https://tinyurl.com/yc8mxxk4 [accessed 2022-03-09]
191. McDaniel JT, Nuhu K, Ruiz J, Alorbi G. Social determinants of cancer incidence and mortality around the world: an ecological study. Glob Health Promot 2019 Mar 16;26(1):41-49. [doi: 10.1177/1757975916686913] [Medline: 29187122]
192. Bompelli A, Wang Y, Wan R, Singh E, Zhou Y, Xu L, et al. Social and behavioral determinants of health in the era of artificial intelligence with electronic health records: a scoping review. Health Data Sci 2021 Aug 24;2021:1-19. [doi: 10.34133/2021/9759016]
193. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019 Jun;110:12-22. [doi: 10.1016/j.jclinepi.2019.02.004] [Medline: 30763612]
194. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 01;27(3):491-497 [FREE Full text] [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
195. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep 2019 Jan 21;49(1):15-21. [doi: 10.1002/hast.973] [Medline: 30790315]
196. McCradden MD, Stephenson EA, Anderson JA. Clinical research underlies ethical integration of healthcare artificial intelligence. Nat Med 2020 Sep 09;26(9):1325-1326. [doi: 10.1038/s41591-020-1035-9] [Medline: 32908273]
197. McCradden MD, Anderson JA, A Stephenson E, Drysdale E, Erdman L, Goldenberg A, et al. A research ethics framework for the clinical translation of healthcare machine learning. Am J Bioeth 2022 May 20;22(5):8-22. [doi: 10.1080/15265161.2021.2013977] [Medline: 35048782]
198. Ibrahim SA, Charlson ME, Neill DB. Big data analytics and the struggle for equity in health care: the promise and perils. Health Equity 2020 Apr 01;4(1):99-101 [FREE Full text] [doi: 10.1089/heq.2019.0112] [Medline: 32258961]
199. Catlow J, Bray B, Morris E, Rutter M. Power of big data to improve patient care in gastroenterology. Frontline Gastroenterol 2022 May 28;13(3):237-244 [FREE Full text] [doi: 10.1136/flgastro-2019-101239] [Medline: 35493622]
200. Chambers DA, Amir E, Saleh RR, Rodin D, Keating NL, Osterman TJ, et al. The impact of big data research on practice, policy, and cancer care. Am Soc Clin Oncol Educ Book 2019 Jan;39:e167-e175 [FREE Full text] [doi: 10.1200/EDBK_238057] [Medline: 31099675]
201. Xu J, Yang P, Xue S, Sharma B, Sanchez-Martin M, Wang F, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. Hum Genet 2019 Mar 22;138(2):109-124 [FREE Full text] [doi: 10.1007/s00439-019-01970-5] [Medline: 30671672]
202. Moscou S, Anderson MR, Kaplan JB, Valencia L. Validity of racial/ethnic classifications in medical records data: an exploratory study. Am J Public Health 2003 Jul;93(7):1084-1086. [doi: 10.2105/ajph.93.7.1084] [Medline: 12835189]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
HIC: high-income country
HPV: human papillomavirus
LMIC: low- and middle-income country
ML: machine learning
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by T Leung; submitted 21.05.22; peer-reviewed by J Lau, SJC Soerensen , S Pesälä; comments to author 21.07.22; revised version received 11.08.22; accepted 24.08.22; published 01.11.22.

Please cite as:
Istasy P, Lee WS, Iansavichene A, Upshur R, Gyawali B, Burkell J, Sadikovic B, Lazo-Langner A, Chin-Yee B
The Impact of Artificial Intelligence on Health Equity in Oncology: Scoping Review
J Med Internet Res 2022;24(11):e39748
URL: https://www.jmir.org/2022/11/e39748
doi:10.2196/39748
PMID:36005841

©Paul Istasy, Wen Shen Lee, Alla Iansavichene, Ross Upshur, Bishal Gyawali, Jacquelyn Burkell, Bekim Sadikovic, Alejandro Lazo-Langner, Benjamin Chin-Yee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 01.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Nurse-led Telehealth Intervention for Rehabilitation (Telerehabilitation) Among Community-Dwelling Patients With Chronic Diseases: Systematic Review and Meta-analysis

Athena Yin Lam Lee[1], BSc; Arkers Kwan Ching Wong[1], PhD; Tommy Tsz Man Hung[1], PhD; Jing Yan[2], MD; Shulan Yang[2], MSc

[1]School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

[2]Zhejiang Hospital, Zhejiang, China

Corresponding Author:
Arkers Kwan Ching Wong, PhD
School of Nursing
The Hong Kong Polytechnic University
Room, 502, Block GH, 1 Yuk Choi Road
Hung Hom, 000
Hong Kong
Phone: 852 34003805
Email: arkers.wong@polyu.edu.hk

Abstract

Background: Chronic diseases are putting huge pressure on health care systems. Nurses are widely recognized as one of the competent health care providers who offer comprehensive care to patients during rehabilitation after hospitalization. In recent years, telerehabilitation has opened a new pathway for nurses to manage chronic diseases at a distance; however, it remains unclear which chronic disease patients benefit the most from this innovative delivery mode.

Objective: This study aims to summarize current components of community-based, nurse-led telerehabilitation programs using the chronic care model; evaluate the effectiveness of nurse-led telerehabilitation programs compared with traditional face-to-face rehabilitation programs; and compare the effects of telerehabilitation on patients with different chronic diseases.

Methods: A systematic review and meta-analysis were performed using 6 databases for articles published from 2015 to 2021. Studies comparing the effectiveness of telehealth rehabilitation with face-to-face rehabilitation for people with hypertension, cardiac diseases, chronic respiratory diseases, diabetes, cancer, or stroke were included. Quality of life was the primary outcome. Secondary outcomes included physical indicators, self-care, psychological impacts, and health-resource use. The revised Cochrane risk of bias tool for randomized trials was employed to assess the methodological quality of the included studies. A meta-analysis was conducted using a random-effects model and illustrated with forest plots.

Results: A total of 26 studies were included in the meta-analysis. Telephone follow-ups were the most commonly used telerehabilitation delivery approach. Chronic care model components, such as nurses-patient communication, self-management support, and regular follow-up, were involved in all telerehabilitation programs. Compared with traditional face-to-face rehabilitation groups, statistically significant improvements in quality of life (cardiac diseases: standard mean difference [SMD] 0.45; 95% CI 0.09 to 0.81; P=.01; heterogeneity: X^2_1=1.9; I^2=48%; P=.16; chronic respiratory diseases: SMD 0.18; 95% CI 0.05 to 0.31; P=.007; heterogeneity: X^2_2=1.7; I^2=0%; P=.43) and self-care (cardiac diseases: MD 5.49; 95% CI 2.95 to 8.03; P<.001; heterogeneity: X^2_5=6.5; I^2=23%; P=.26; diabetes: SMD 1.20; 95% CI 0.55 to 1.84; P<.001; heterogeneity: X^2_4=46.3; I^2=91%; P<.001) were observed in the groups that used telerehabilitation. For patients with any of the 6 targeted chronic diseases, those with hypertension and diabetes experienced significant improvements in their blood pressure (systolic blood pressure: MD 10.48; 95% CI 2.68 to 18.28; P=.008; heterogeneity: X^2_1=2.2; I^2=54%; P=0.14; diastolic blood pressure: MD 1.52; 95% CI –10.08 to 13.11, P=.80; heterogeneity: $X^2$1=11.5; I^2=91%; P<.001), and hemoglobin A1c (MD 0.19; 95% CI –0.19 to 0.57 P=.32; heterogeneity: X^2_4=12.4; I^2=68%; P=.01) levels. Despite these positive findings, telerehabilitation was found to have no statistically significant effect on improving patients' anxiety level, depression level, or hospital admission rate.

XSL•FO
RenderX

Conclusions: This review showed that telerehabilitation programs could be beneficial to patients with chronic disease in the community. However, better designed nurse-led telerehabilitation programs are needed, such as those involving the transfer of nurse-patient clinical data. The heterogeneity between studies was moderate to high. Future research could integrate the chronic care model with telerehabilitation to maximize its benefits for community-dwelling patients with chronic diseases.

Trial Registration: International Prospective Register of Systematic Reviews CRD42022324676; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=324676

(J Med Internet Res 2022;24(11):e40364) doi:10.2196/40364

KEYWORDS

chronic disease; telerehabilitation; meta-analysis; telemedicine; nurses; outpatients

Introduction

"Telehealth" refers to the delivery of health services through technology when health care providers and patients are separated by distance [1]. One of the branches of telehealth, telerehabilitation, is defined as the use of a telehealth approach to provide rehabilitation care to people with long-term chronic diseases [2]. Telerehabilitation programs employ communication and information technology, such as telephones and videoconferencing, as a delivery channel to provide not only exercise training, but also self-management education and health behavior modifications to patients with chronic disease who are not receiving hospital care [3,4]. Despite offering convenience, telerehabilitation also has well-known disadvantages, such as technical issues, limitations on carrying out procedures that require physical contact, and security breaches.

Many systematic reviews have been published in recent years on the effectiveness of telerehabilitation programs for those with one specific chronic disease (eg, cardiac diseases, respiratory diseases, stroke, or neurological diseases). A previous systematic review suggested that there is controversy over the effectiveness of telerehabilitation and that its impacts could differ depending on which chronic disease a person has [5]. However, to our knowledge, no reviews have been published on which chronic disease patients would benefit most from telerehabilitation programs. The aim in this present review is to address these research and service gaps by comparing the effects of telerehabilitation programs on people with different chronic diseases. If proven successful, the findings can aid the government and policymakers in better allocating health care resources, foster the development of telerehabilitation programs during and beyond the COVID-19 pandemic, and improve the quality of community care services.

The objectives of the review are to identify the intervention components of current nurse-led telerehabilitation programs for community-dwelling people with chronic diseases, to evaluate the effectiveness of nurse-led telerehabilitation programs compared with traditional face-to-face rehabilitation programs, and to compare the effects of telerehabilitation on patients with different chronic diseases.

Methods

Overview

A systematic review and meta-analysis was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [6]. This study was registered in the International Prospective Register of Systematic Reviews (PROSPERO; CRD42022324676).

Literature Search

The literature search was performed by 2 independent reviewers (AYLL and AKCW) without the involvement of librarians. PubMed, MEDLINE, CINAHL, Embase, PsycInfo, and the Cochrane Central Register of Controlled Trials were searched for articles published from 2015 to 2021, with the aim of capturing the most updated telerehabilitation approaches under rapid technological development. Handsearching was performed using Google Scholar and the reference lists of included papers. Gray literature, such as abstracts and editorials, were excluded as most of these articles are not peer-reviewed and their inclusion would have lowered the quality of evidence. Search strategies for all databases were constructed based on the key search terms, which included "telerehabilitation," "chronic disease," "nursing," "multi-disciplinary," and "randomized controlled trial." The search was further expanded by the inclusion of different chronic diseases and medical subject headings mesh terms (Multimedia Appendix 1).

Eligibility Criteria

Overview

Studies that included people with hypertension, cardiac diseases (coronary artery diseases, heart failure), chronic respiratory diseases (asthma, chronic obstructive pulmonary disease [COPD]), diabetes, cancer, or stroke were the target of this review because these are common diseases among people in the community that require the provision of long-term nursing rehabilitation care. Articles were screened using the following eligibility criteria constructed using the patient, intervention, comparison, and outcome (PICO) strategy.

Inclusion Criteria

The inclusion criteria for articles were the following: participants aged 18 years or above, diagnosed with one of the targeted chronic diseases, and living independently in the community outside health care facilities; telerehabilitation employed as the intervention delivery channel in 1 arm of the intervention (the

XSL•FO
RenderX

channel could include telephone calls, smartphone apps, videoconferencing, or SMS text messaging), with nurses providing of at least 50% of the program in terms of the frequency or duration of the provision of care; comparison to conventional face-to-face center-based consultations or a waitlist control; outcomes of quality of life, disease-specific physical indicators, self-care ability, psychological outcomes (depression, anxiety), and health-resource use; and a randomized controlled trial study design.

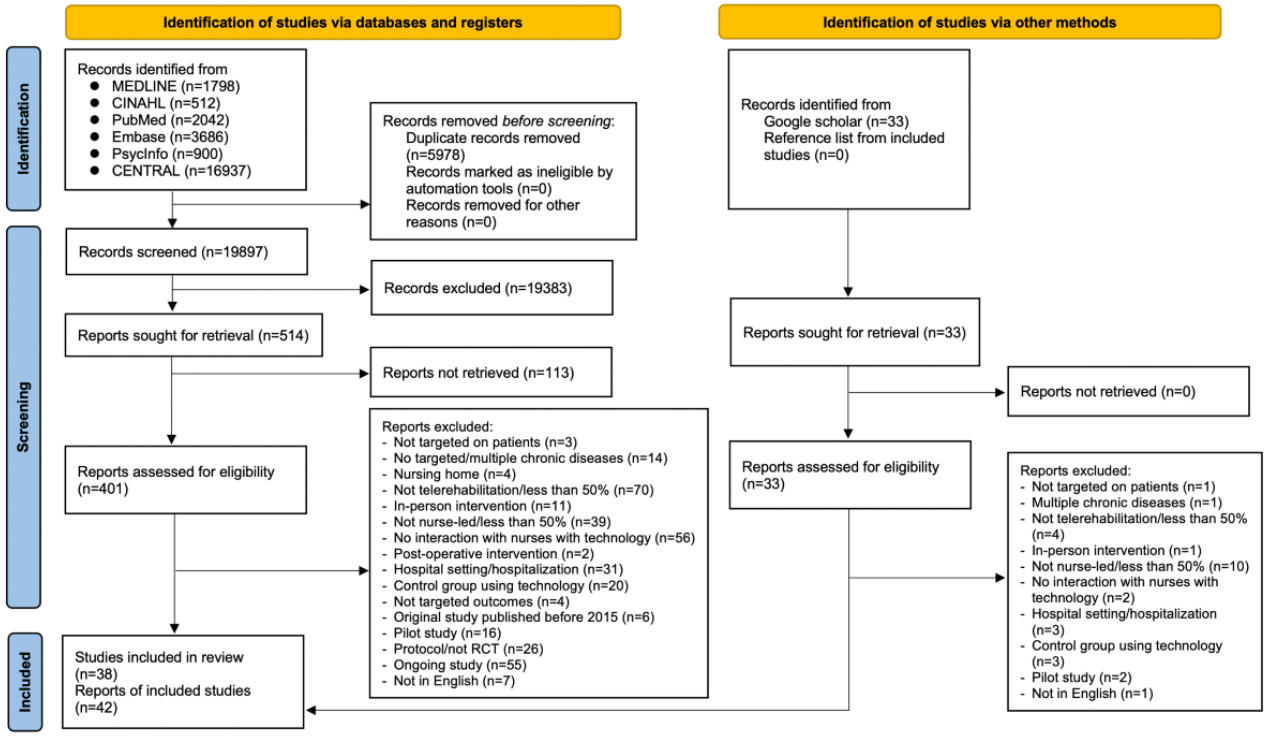
Exclusion Criteria

The exclusion criteria for articles were the following: participants were patients living in assisted residential care facilities (ie, a nursing home) and interventions were telerehabilitation programs conducted in hospital settings where the purpose of the program was to provide education or training only to health care professionals.

Study Selection

The literature screening process is reported using the PRISMA flowchart (Figure 1). The search results were retrieved and imported into EndNote X9 (Clarivate) for the removal of duplicates after the literature search. Articles were screened by title and abstract, which was followed by an examination of the full text by 2 reviewers (AYLL and AKCW) working independently. For the handsearching, the same 2 reviewers independently screened the full text of articles. Any disagreements among the reviewers were resolved through discussion.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flowchart (adapted from Page et al [6], which is published under Creative Commons Attribution 4.0 International License [7]). CENTRAL: The Cochrane Central Register of Controlled Trials; RCT: randomized controlled trial.



Data Extraction

The following variables were extracted and are listed in Multimedia Appendix 1: author, year of publication, study characteristics (study location, study population), intervention characteristics (providers, study duration, intervention group, control group), data collection timepoint, outcome variables, outcome measures, and results.

The interventions of all included studies were extracted according to the chronic care model. The six components of the model are as follows: (1) active, two-way interactions between an informed patient and proactive health care providers; (2) effective self-management support during communication; (3) an intact delivery system design with regular follow-ups for evaluation; (4) proper decision support from expertise, protocols, or training; (5) a clinical information system between patients and health care providers for managing the patients' clinical data; and (6) community resources [8].

Quality Assessment

The Cochrane risk of bias tool was used to identify the potential risk of bias in the included studies [9]. Two reviewers (AYLL and AKCW) performed the quality appraisal independently and resolved any disagreements through discussion.

Statistical Analysis

A meta-analysis was conducted using Review Manager version 5.4.1 (The Cochrane Collaboration) and illustrated using a forest plot when at least 2 studies were measured for the same outcomes for a chronic disease at the longest follow-up timepoint [10]. Under the random-effects model, a mean difference (MD; using the same measurement tool) or a standardized MD (SMD; using different measurement tools)

XSL•FO
RenderX

and a 95% CI were calculated for continuous variables, while odds ratios (ORs) and the 95% CIs were computed using the Mantel-Haenszel method for dichotomous variables. The heterogeneity and significance of the results were assessed using a chi-squared test, I^2 statistics, and a P value (<.05). The value of I^2 could be interpreted as indicating unimportant (0%-40%), moderate (30%-60%), substantial (50%-90%), or considerable (75%-100%) heterogeneity [10]. Publication bias was checked by visualization of a funnel plot [10].

Results

Screening Process

The screening process is reported in Figure 1. A total of 434 papers were included for a full-text screening, and eventually, 38 studies met the criteria for inclusion in this review, 26 of which had data available for a meta-analysis. The characteristics of all of the included studies are presented in the data extraction table (Multimedia Appendix 2).

Study Characteristics

A total of 9677 participants with a mean age of 63.75 years were included in the 38 studies. Of these, 5105 received telerehabilitation and 4572 received conventional face-to-face consultations; 6 studies were 3-armed randomized controlled trials [11-16], and the remaining studies were 2-armed studies. Among the 38 included studies, 9 targeted patients with cardiac disease [15-23], 9 targeted chronic patients with respiratory disease [24-32], 9 targeted patients with diabetes [11,12,33-39], 4 targeted patients with hypertension [14,40-42], 4 targeted patients with cancer [43-46], and 3 targeted patients with stroke [13,47,48].

Intervention Characteristics

The programs were performed by registered nurses (n=20), specialty nurses (n=8), advanced practice nurses (n=3), community nurses (n=1), a nurse case manager (n=1), or people involved in more than one health-related discipline (n=5). The study periods ranged from 4 weeks to 36 months, with 8 weeks (n=8), 12 weeks (n=7), and 24 weeks (n=7) being the most common durations.

The technologies used in telerehabilitation programs are summarized in Table 1. Generally, the nurse telephone follow-up (n=26) was the most commonly adopted nurse-led telerehabilitation delivery channel for all chronic disease patients, followed by telemonitoring (n=9) and videoconferencing (n=4). In addition, most of the studies involved nontechnological components in addition to telerehabilitation, such as distributing written educational materials (n=20), attending an in-person educational session (n=16), regular face-to-face training (n=10), and home visits by nurses (n=3). For the control groups, regular nursing consultations (n=30), paper-based educational materials (n=14), and an in-person educational session (n=6) were used.

Table 1. Telerehabilitation interventions for different chronic diseases.

	Hypertension, n	Cardiac diseases, n	Chronic respiratory diseases, n	Diabetes, n	Cancer, n	Stroke, n
Nurse follow-ups by telephone/video	3	6	5	8	2	3
Nurse follow-ups by SMS texts	2	0	0	0	0	0
Telemonitoring	0	3	4	1	0	0
Smartphone apps	0	0	0	2	2	0
Website	0	1	0	0	0	0
Exercise training	0	1	0	0	0	0

Nurse Follow-Ups

A total of 32 studies conducted telerehabilitation programs through telephone (n=26) [11-13,15,17,19-22,24-28,33,35,37-43,46-48], videoconferencing (n=4) [21,30,32,34], SMS text messaging (n=2) [14,41], or WhatsApp (n=1) [45] (Multimedia Appendix 3). Nurse-led counseling was mostly implemented weekly (n=4), monthly (n=5), or a combination of both (n=7). Seven studies did not report the frequency of their interventions. The contents of nursing follow-up included providing education on disease-specific knowledge (eg, COPD exacerbation, hypoglycemia) and self-care behavior (eg, medication adherence, lifestyle modification; n=14), addressing patients' enquiries on disease self-management (n=8), monitoring patients' signs and symptoms (n=6), conducting motivational interviewing (n=5), performing medication titration with collaboration of physicians (n=4), empowering goal-setting and personal plan implementation (n=4), and providing psychological support (n=3).

There were 2 studies framed by problem-solving theory, which supported chronic disease rehabilitation by developing behavioral plans and providing positive reinforcement during nurse-led phone counseling [13,28]. Another 2 studies used noninteractive information SMS texts and/or interactive SMS texts with nurses to provide education on chronic disease management and support on disease monitoring [14,41].

Telemonitoring

Nine studies integrated telemonitoring in their telerehabilitation program for patients with heart failure, asthma, COPD, or diabetes [16,18,21,29-32,34,35] (Multimedia Appendix 4). Patients were instructed to measure their disease-specific physical indicators (eg, blood pressure, spirometry, oxygen saturation, respiratory rate, blood glucose level) and record their signs and symptoms daily (n=5), weekly (n=1), or from daily to weekly after the first few weeks of interventions (n=3). The data were transmitted to a shared platform by manual recording in tablet and mobile apps (n=3), auto-transmission from

XSL•FO
RenderX

measurement tools to tablet (n=2), or SMS texts (n=1). In 7 studies, alerts were sent automatically to nurses if abnormal data were detected by decision-support systems. These decision-support systems were constructed according to research protocol (n=5) or through shared decision-making with patients (n=2). After receiving the alerts, the nurses would support these patients through continuing telephone follow-up (n=6), videoconferencing (n=2), or referring to physicians (n=1).

Other Telerehabilitation Interventions

Apart from nurse follow-ups and telemonitoring, there were 4 studies that used smartphone apps to support chronic disease rehabilitation in the community [23,35,44,45]. The functions of these apps generally included provision of multimedia educational materials, monitoring of health behaviors, psychological support, chat functions, and discussion forums for nurse counseling. Two studies designed a website to provide education information on cardiac self-management [16] and monitor patients' health status with the use of online health questionnaires [46]. In addition, one study provided online exercise training for patients with health failure through videoconferencing [23].

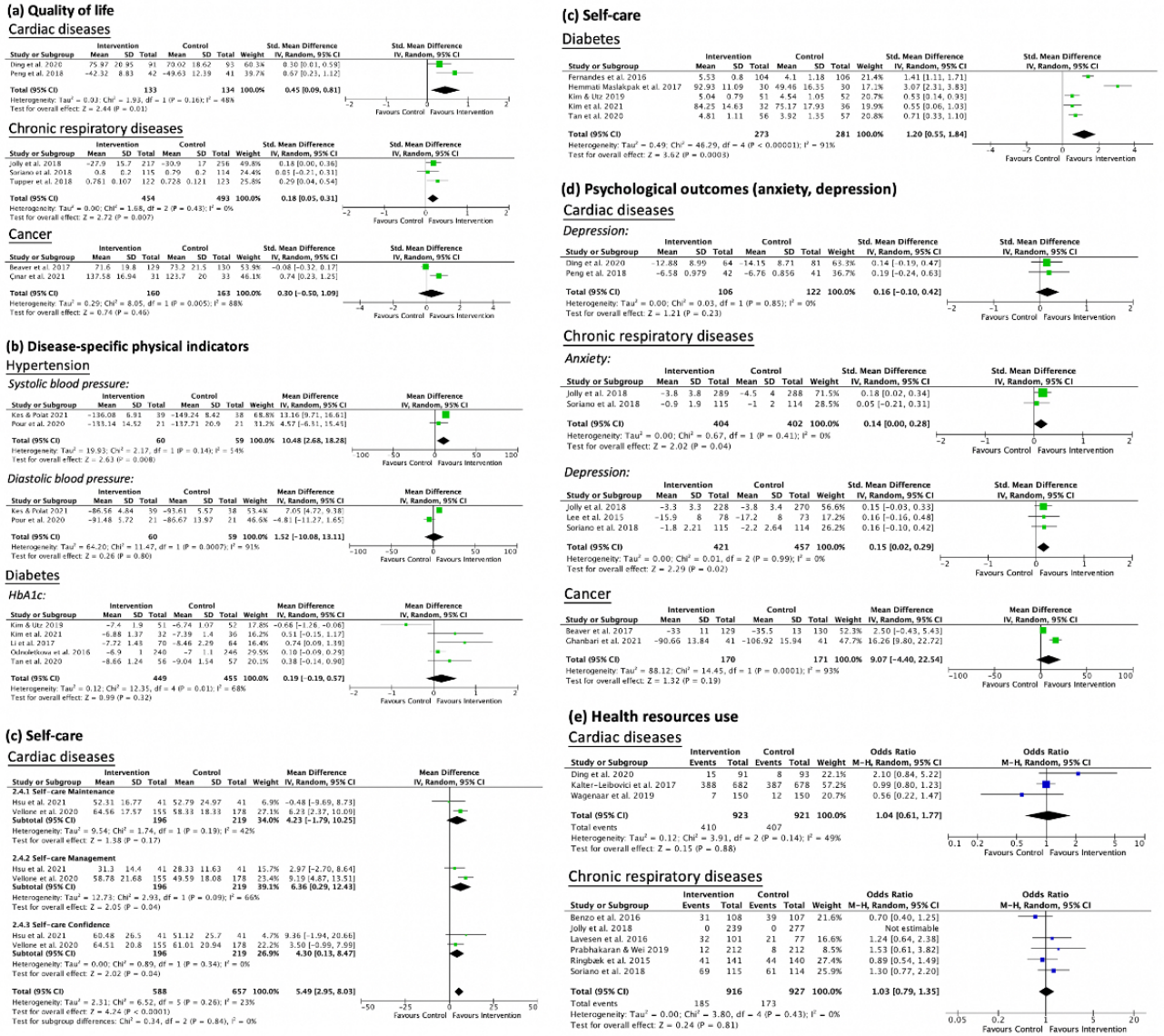
Chronic Care Model

The overall chronic care model elements among all included studies are illustrated in Multimedia Appendix 5. The interventions in all of the studies are shown as being aligned with at least a part of the model. In all of the 38 included studies, regular interactions between patients and nurses, self-management support, and regular follow-ups were provided. In a total of 26 studies, decision support for nurses was provided through preintervention training, guidelines, or protocols. A clinical information system was involved in 10 studies with a telemonitoring component. There were 8 studies in which referrals were provided to available community resources.

Quantitative Synthesis

Forest plots for all outcomes are shown in Figure 2.

Figure 2. Forest plot on the effectiveness of community-based nurse-led telerehabilitation programs on (a) quality of life, (b) disease-specific physical indicators, (c) self-care ability, (d) psychological outcomes, and (e) health-resource use. std: standard.



XSL•FO
RenderX

Quality of Life

Hypertension

One study reported that telerehabilitation programs had a positive effect on the quality of life of people with hypertension [40].

Cardiac Diseases

A significant improvement in the quality of life of cardiac disease patients who had received telerehabilitation was observed when compared with those who had received a nontechnological intervention (SMD 0.45; 95% CI 0.09-0.81; P=.01), with moderate heterogeneity (X^2_1=1.9; I^2=48%; P=.16).

Chronic Respiratory Diseases

Pooled analyses of 3 studies showed that patients with COPD who received telerehabilitation had a significantly higher quality of life than did those who received conventional face-to-face rehabilitation (SMD 0.18; 95% CI 0.05-0.31; P=.007; heterogeneity: X^2_2=1.7; I^2=0%; P=.43).

Diabetes (Type II)

For diabetes, 1 study [34] showed no significant difference, while 2 studies [38,39] reported an improvement in the quality of life of patients after receiving nurse-led telerehabilitation.

Cancer

From a meta-analysis of 2 studies, it was revealed that there was no significant difference in quality of life between the intervention and control groups of patients with cancer (SMD 0.30; 95% CI –0.50 to 1.09; heterogeneity: P=.46; X^2_1=8.1; I^2=88%; P=.005).

Disease-Specific Physical Indicators

Hypertension

Pooled intervention effects from 2 studies showed a significant improvement in the systolic blood pressure of patients through telerehabilitation (MD 10.48; 95% CI 2.68 to 18.28; P=.008), with moderate heterogeneity (X^2_1=2.8; I^2=54%; P=.14). However, no significant difference was observed in their diastolic blood pressure (MD 1.52; 95% CI –10.08 to 13.11; P=.80; heterogeneity: X^2_1=11.5; I^2=91%; P<.001).

Cardiac Diseases

Among those with heart failure, no significant differences between the telerehabilitation and control groups were observed in physical symptoms [17,21,23].

Chronic Respiratory Diseases

One included study reported that telerehabilitation had no effect on reducing the number of instances of COPD exacerbation or COPD symptom levels [31]. Another study also found no significant difference in dyspneic levels between those who received telerehabilitation and those who received conventional in-person follow-ups [26].

Diabetes (Type II)

A meta-analysis of 5 studies found that telerehabilitation had no significant effect on improving the hemoglobin A1c levels of patients (MD 0.19; 95% CI –0.19 to 0.57; P=.32). However, the above result might have been affected by an outlier since the findings showed substantial heterogeneity (X^2_4=12.4; I^2=68%; P=.01).

Stroke

One study showed improved systolic blood pressure, diastolic blood pressure, and low-density lipoprotein levels in stroke survivors who had received a telerehabilitation program [47].

Self-Care Ability

Hypertension

Only 1 study assessed the effect of telerehabilitation on self-care among patients with hypertension, and in that study, no significant difference was found between the groups [42].

Cardiac Diseases

A pooled analysis indicated that telerehabilitation could have a beneficial effect on the self-care ability of patients with cardiac diseases (MD 5.49; 95% CI 2.95 to 8.03; P<.001), with mild heterogeneity (X^2_5=6.5; I^2=23%; P=.26). A subgroup analysis showed that participation in telerehabilitation led to a significant improvement in the participants' self-care management (MD 6.36; 95% CI 0.29 to 12.43; P=.04) and self-care confidence (MD 4.30; 95% CI 0.13 to 8.47; P=.04) but not in their self-care maintenance (MD 4.23; 95% CI –1.79 to 10.25; P=.17).

Chronic Respiratory Diseases

One included study revealed that patients demonstrated a significant improvement in disease self-management after receiving telerehabilitation [27]. Nevertheless, another study showed no significant difference in self-management health behaviors between the telerehabilitation and onsite out-patient follow-up groups [25].

Diabetes (Type II)

The pooled SMD indicated that telerehabilitation had a significant positive effect on enhancing the self-care behavior of patients with diabetes when compared with conventional face-to-face nursing consultations (SMD 1.20; 95% CI 0.55-0.84; P<.001; heterogeneity: X^2_4=46.3; I^2=91%; P<.001).

Psychological Outcomes (Depression, Anxiety)

Cardiac Diseases

A meta-analysis showed that telerehabilitation had no significant effect on reducing the depression levels of patients who experience heart failure (SMD 0.16; 95% CI –0.10 to 0.42; P=.23; heterogeneity: X^2_1=0.03; I^2=0%; P=.85).

Chronic Respiratory Diseases

Pooled analyses in 3 studies found there to be a significant reduction in anxiety (SMD 0.14; 95% CI 0.00-0.28; P=.04) and depression levels (SMD 0.15; 95% CI 0.02-0.29; P=.02) in patients with COPD.

XSL•FO
RenderX

Diabetes (Type II)

The only study that evaluated the effects of a telerehabilitation program on patients with diabetes showed an improvement in depression [36].

Cancer

The pooled MD showed no significant effect between groups on relieving anxiety (MD 9.07; 95% CI –4.40 to 22.54; *P*=.19; heterogeneity: X^2_1=14.5; I^2=93%; *P*<.001).

Stroke

One study reported no significant differences in depression levels between stroke survivors who received telerehabilitation and those who received conventional face-to-face nurse consultations [13].

Health-Resource Use

Cardiac Diseases

A pooled intervention effect of 3 studies showed that telerehabilitation had no significant effect on reducing hospitalizations of patients with heart failure (OR=1.04, 95% CI 0.61-1.77; *P*=.88; heterogeneity: X^2_2=3.91; I^2=49%; *P*=.14).

Chronic Respiratory Diseases

Telerehabilitation had no significant effect on reducing respiratory-related hospitalizations (OR 1.03, 95% CI 0.79-1.35; *P*=.81), with no heterogeneity observed (X^2_4=3.8; I^2=0%; *P*=.43).

Diabetes (Type II)

One included study reported a greater reduction in unplanned health care services usage among patients with diabetes in the telerehabilitation group compared to those in the control group [39]. However, another study found no significant differences [38].

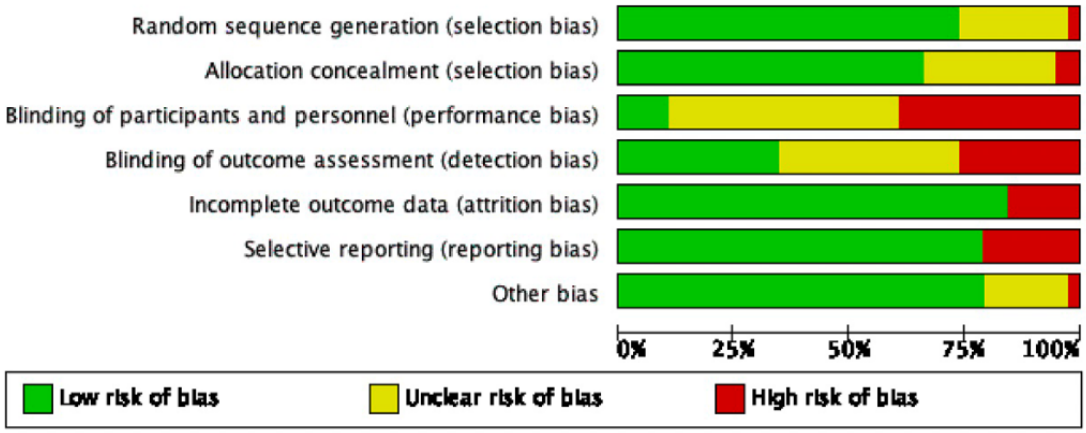
Cancer

One study reported no significant difference in health-resource use between patients with cancer who received telerehabilitation and those in the control group [46].

Risk of Bias

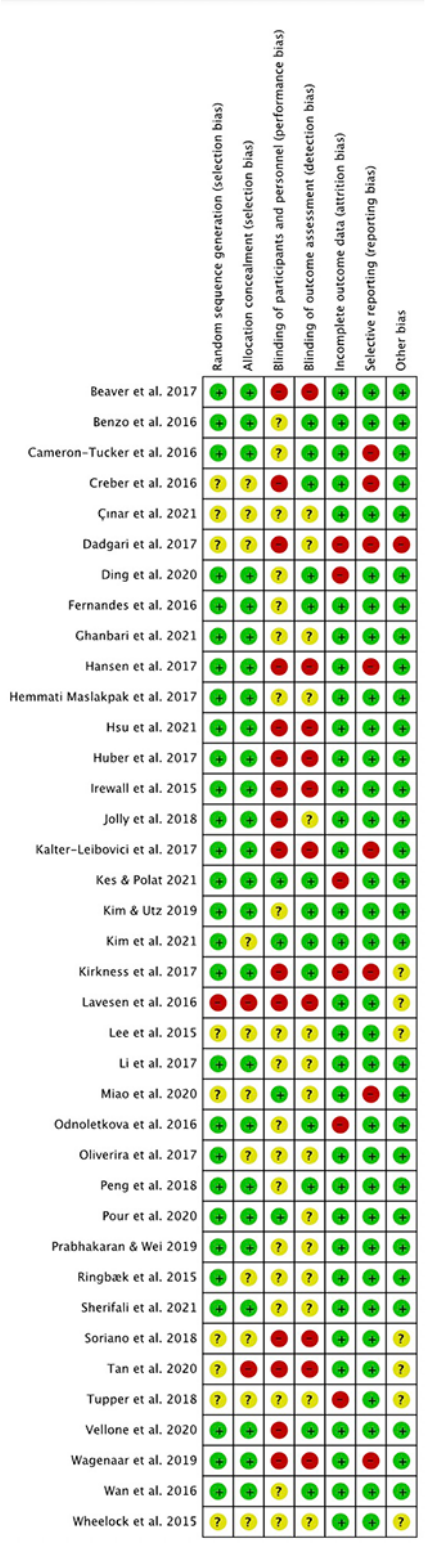
A summary of the risk of bias in the studies is shown in Figure 3 and 4. The quality of the randomization and allocation concealment in most of the studies was good. However, due to the nature of telerehabilitation, blinding of participants and interventionists was difficult. There were 34 out of 42 studies rated as high or unclear risk regarding to blinding of participants and personnel (90%). In addition, 25 out of 42 studies were rated with high or unclear risk on blinding to outcome assessment (66%), while 12 studies had high risk on outcome reporting (29%).

Figure 3. Risk of bias.



XSL•FO
RenderX

Figure 4. Risk of bias table.



Discussion

Principal Results

Given the limited resources in hospital settings and the ongoing COVID-19 pandemic, telerehabilitation seems to be a promising long-term approach to delivering continuous care from health care professionals to people with chronic diseases [49]. In this meta-analysis, it was found that patients with chronic disease experienced a significant improvement in their quality of life and self-care ability after receiving nurse-led telerehabilitation when compared with those who received a conventional in-person rehabilitation service. These improvements might have resulted from an increase in people's knowledge of how to monitor their symptoms and in their ability to perform clinical assessments on their own after participating in a nurse-led telerehabilitation program [30,46]. Similar results were seen in previous reviews targeting community-dwelling patients [50]

XSL•FO
RenderX

with heart failure [51], COPD [52], or cancer [53]. Thus, with these findings, telerehabilitation programs can be applied in community-based rehabilitation services, especially during the current COVID-19 pandemic. Although telerehabilitation is beneficial to the quality of life and self-care ability of patients with chronic disease, its effect on their psychological health and hospital admission is less certain. A previous review analyzing the effects of nurse-driven telerehabilitation programs found significant improvement in anxiety and depression level among patients with COPD [52]. In contrast, a few studies reported that telerehabilitation had no significant impact on the psychological health of those with chronic disease [18,23]. In addition, the effectiveness of telerehabilitation on reducing health resource use was also varied among previous reviews. A previous integrated review was not able to prove the effectiveness of telerehabilitation on reducing hospitalization rate among patients with heart failure [54]. Some reviews found significant reduction of nonplanned hospital admission [55] and emergency department visits [56]. In contrast, a few studies reported no significant differences on health-resource use for community-dwelling older adults [50], patients with diabetes [57], or those with heart failure [58]. These mixed findings might have resulted from differences in the characteristics of the patients and interventions in their respective nurse-led rehabilitation programs. Therefore, future reviews are needed to compare the effectiveness of telerehabilitation programs according to their different intervention characteristics, such as duration, delivery mode, and dosage.

Nurse telephone follow-ups were found to be the most common intervention component in nurse-led telerehabilitation programs, which was consistent with the finding in a previous study [59]. Telephone follow-ups were perceived to be by far the easiest way to ask health care providers questions about disease self-management, while not requiring any sophisticated devices [12,43,46]. By contrast, telemonitoring, another intervention component frequently used in telerehabilitation programs, was regarded as the least favorable by patients because of the frequent technical issues that arose during the transmission of data using wireless devices [18,60,61]. The patients were also concerned about the accuracy of the tools used in the in-home monitoring of vital signs and the wearable sensor [62]. In addition, the inability to use the monitoring tools and interpret their own health data were also common reasons for noncompliance in self-monitoring [18]. Telemonitoring is thus better implemented with adequate preintervention nursing education or training sessions for patients to familiarize themselves with the technological devices. Future research should also improve the quality of the telemonitoring system, including stability, accuracy, and security to increase patients' confidence towards telemonitoring.

Despite their benefits, rehabilitation programs should not be provided solely via a telecare delivery mode. The lack of physical interaction can lead to difficulty in building a trusting nurse-patient relationship and hence lower the satisfaction of patients [26,42]. In addition, telerehabilitation may indeed increase the anxiety and depression levels of patients due to inexperience in using technology [23,63]. Patients with chronic disease may need regular face-to-face nurse consultations to solve the problems that they encounter during telerehabilitation. Supplementing telerehabilitation with face-to-face consultations allows for more comprehensive nursing assessments and physical examinations to be conducted [64].

Guided by the chronic care model, this review found that all included studies provided regular two-way interactions between patients and proactive health care providers. However, when abnormal findings or acute problems were identified, some studies did not provide evidence-based protocols for nurses to follow, which might have led to inaccuracy in clinical judgement and an increase in unnecessary hospital admission [65-67]. Therefore, a reliable guideline should be given to health care providers before the implementation of telerehabilitation programs.

Given the current pace of technological development, more advanced decision support systems can be improved with the aid of artificial intelligence (AI) [68]. Different decision support systems were developed in recent research for chronic disease management, most commonly for diagnosis, follow-up management, and treatment [69]. A previous study created an AI-based decision support system for enhancing shared decision-making in a pharmacotherapy regimen for patients with diabetes [70]. Based on patients' clinical data, this AI-driven system can generate medication regimens with comprehensive information, including predicted success rate, risks and benefits, and medication costs. Another study adopted a machine-learning decision support system in telemonitoring to predict the risk of acute asthma exacerbation according to patients' self-report symptoms, with timely alert notification to nurses when abnormalities were detected [71].

In addition to decision support systems, clinical information systems are another important component in the chronic care model that need to be considered in nurse-led telerehabilitation programs. The lack of a shared clinical information system among health care professionals has shown to increase the risk of medical errors [72]. With the use of technology, health records of patients can be shared electronically between patients and health care providers or among different health care disciplines. Evidence suggests that integrating electronic health records in community-based chronic disease care can effectively improve patients' health outcomes and quality of health care services [73]. Although the use of electronic health records has been widely used and tested, it is limited to showing only objective physical indicators, such as blood glucose level and radiology reports. Recent studies have begun to allow patients to impute their subjective health complaints, such as symptoms and physical activity, into the electronic health records [74]. Nevertheless, concerns have been raised concerning data privacy issues when patients' personal information were uploaded and stored on the internet [62,75]. Therefore, future research is needed to develop a cloud platform with a more advanced security system so as to prevent breaching of patient health data [76]. In addition, policy makers should regulate the storage and sharing of patient health information to third parties for medical follow-up and referral to ensure data privacy [77].

Chronic diseases are usually associated with functional impairment, which can reduce the ability of patients to adapt

XSL•FO
RenderX

telerehabilitation. The evidence shows that patients with greater physical disabilities are less likely than their counterparts to comply with a telerehabilitation program [18,27,78]. To improve compliance, adequate preintervention training is needed on disease self-management and on the use of technological devices [79]. Among the 6 targeted chronic diseases, previous research suggested that cardiac diseases, chronic respiratory diseases, and stroke would cause higher functional disability in patients than would hypertension, diabetes, and cancer and thus compromise their ability to participate in telerehabilitation programs [80]. This may explain why improvement in the physical indicators examined in this review, including COPD exacerbation and physical disability level, was not significant among patients with cardiac diseases, chronic respiratory diseases, or stroke after participating in a nurse-led telerehabilitation program. In view of this, it is suggested that caregivers should be involved in assisting such patients to become engaged in telerehabilitation programs.

Future Directions

Most older adults not only suffer from a single chronic disease, but also face the problem of multimorbidity. The prevalence of multimorbidity in China and the United States has been reported to be 49.4% and 59.6%, respectively [81,82]. Rehabilitation services for patients with multiple chronic diseases are more complex in nature than are those for patients with a single disease due to the interrelated pathophysiological pathways of chronic diseases [83]. The difficulty in interpreting symptoms and managing multiple medical regimens increases due to the overlapping signs and symptoms of these complex and interrelated chronic diseases [84]. It is thus crucial to conduct future studies to evaluate the effectiveness of nurse-led telerehabilitation programs among patients with multiple chronic conditions, as there are currently few such studies.

Limitations

This review has several limitations. First, it only included papers written in English, so relevant studies reported in different languages were missed. Second, this review only included nurse-led telerehabilitation programs for the 6 most common chronic diseases. There was no coverage of telerehabilitation programs for those with other chronic diseases resulting in high functional disability, such as arthritis and neurological diseases. Third, the heterogeneity between studies was from moderate to high due to the differences in intervention characteristics, such as study duration, sample size, and technological devices used. Fourth, not all studies were included in the meta-analysis due to data incompleteness despite reviewers' attempts to contact corresponding authors for relevant data.

Conclusions

Although the meta-analysis showed that the programs led to a significant improvement in the quality of life and self-care ability of patients with various chronic diseases, it did not have an advantage over traditional face-to-face consultations with regard to anxiety, depression, or the number of hospital admissions. Guided by the chronic care model, the review showed that the usage of decision support and clinical information systems may facilitate the work of nurses in telerehabilitation programs. In addition, despite the commonality of multimorbidity, limited studies regarding the effectiveness of telerehabilitation programs targeting patients with multiple chronic diseases are available. Future research could focus on the use of telerehabilitation among these patients.

Acknowledgments

This project was funded by the Health and Medical Research Fund (reference #18191221). The funder had no role in the conduct of the study, collection of data, analysis and interpretation of the results, or in the preparation of the manuscript.

Data Availability

All data sets generated or analyzed during the current study are included in this published article and the supplementary materials.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 19 KB - jmir_v24i11e40364_app1.docx]

Multimedia Appendix 2
Data extraction table.
[DOCX File , 144 KB - jmir_v24i11e40364_app2.docx]

Multimedia Appendix 3
Nurse follow-ups.
[DOCX File , 6711 KB - jmir_v24i11e40364_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Telemonitoring.
[DOCX File , 2765 KB - jmir_v24i11e40364_app4.docx]

Multimedia Appendix 5
Chronic care model.
[DOCX File , 27 KB - jmir_v24i11e40364_app5.docx]

Multimedia Appendix 6
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - jmir_v24i11e40364_app6.docx]

References

1. World Health Organization. 2016. Telehealth. URL: https://www.who.int/gho/goe/telehealth/en/ [accessed 2022-02-18]
2. Duruturk N. Telerehabilitation intervention for type 2 diabetes. World J Diabetes 2020 Jun 15;11(6):218-226 [FREE Full text] [doi: 10.4239/wjd.v11.i6.218] [Medline: 32547696]
3. Peretti A, Amenta F, Tayebati SK, Nittari G, Mahdi SS. Telerehabilitation: review of the state-of-the-art and areas of application. JMIR Rehabil Assist Technol 2017 Jul 21;4(2):e7 [FREE Full text] [doi: 10.2196/rehab.7511] [Medline: 28733271]
4. Wong AKC, Wong FKY, Chow KKS, Wong SM, Lee PH. Effect of a Telecare Case Management Program for Older Adults Who Are Homebound During the COVID-19 Pandemic: A Pilot Randomized Clinical Trial. JAMA Netw Open 2021 Sep 01;4(9):e2123453 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.23453] [Medline: 34499135]
5. Bitar H, Alismail S. The role of eHealth, telehealth, and telemedicine for chronic disease patients during COVID-19 pandemic: A rapid systematic review. Digit Health 2021;7:20552076211009396 [FREE Full text] [doi: 10.1177/20552076211009396] [Medline: 33959378]
6. 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]
7. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2022-10-01]
8. Wagner EH, Davis C, Schaefer J, Von Korff M, Austin B. A survey of leading chronic disease management programs: are they consistent with the literature? Journal of Nursing Care Quality 2002;16(2):67-80. [doi: 10.1097/00001786-200201000-00008]
9. Higgins JPT, 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:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
10. Higgins J, Thomas J, Chandler J. 2021. Cochrane Handbook for Systematic Reviews of Interventions version 6. URL: https://training.cochrane.org/handbook/archive/v6.2 [accessed 2021-12-22]
11. Hemmati Maslakpak M, Razmara S, Niazkhani Z. Effects of face-to-face and telephone-based family-oriented education on self-care behavior and patient outcomes in type 2 diabetes: a randomized controlled trial. J Diabetes Res 2017;2017:8404328 [FREE Full text] [doi: 10.1155/2017/8404328] [Medline: 29359166]
12. Kim SH, Utz S. Effectiveness of a social media–based, health literacy–sensitive diabetes self‐management intervention: a randomized controlled trial. J Nurs Scholarsh 2019 Nov;51(6):661-669. [doi: 10.1111/jnu.12521] [Medline: 31622033]
13. Kirkness CJ, Cain KC, Becker KJ, Tirschwell DL, Buzaitis AM, Weisman PL, et al. Randomized trial of telephone versus in-person delivery of a brief psychosocial intervention in post-stroke depression. BMC Res Notes 2017 Oct 10;10(1):500 [FREE Full text] [doi: 10.1186/s13104-017-2819-y] [Medline: 29017589]
14. Pour E, Aliyari S, Farsi Z, Ghelich Y. Comparing the effects of interactive and noninteractive education using short message service on treatment adherence and blood pressure among patients with hypertension. Nurs Midwifery Stud 2020;9(2):68. [doi: 10.4103/nms.nms_82_19]
15. Vellone E, Rebora P, Ausili D, Zeffiro V, Pucciarelli G, Caggianelli G, et al. Motivational interviewing to improve self‐care in heart failure patients (MOTIVATE‐HF): a randomized controlled trial. ESC Heart Failure 2020 Apr 28;7(3):1309-1318. [doi: 10.1002/ehf2.12733]
16. Wagenaar KP, Broekhuizen BDL, Jaarsma T, Kok I, Mosterd A, Willems FF, et al. Effectiveness of the European Society of Cardiology/Heart Failure Association website 'heartfailurematters.org' and an e-health adjusted care pathway in patients with stable heart failure: results of the 'e-Vita HF' randomized controlled trial. Eur J Heart Fail 2019 Feb;21(2):238-246 [FREE Full text] [doi: 10.1002/ejhf.1354] [Medline: 30485612]

XSL•FO
RenderX

17. Masterson Creber R, Patey M, Lee CS, Kuan A, Jurgens C, Riegel B. Motivational interviewing to improve self-care for patients with chronic heart failure: MITI-HF randomized controlled trial. Patient Educ Couns 2016 Feb;99(2):256-264 [FREE Full text] [doi: 10.1016/j.pec.2015.08.031] [Medline: 26358533]
18. Ding H, Jayasena R, Chen SH, Maiorana A, Dowling A, Layland J, et al. The effects of telemonitoring on patient compliance with self-management recommendations and outcomes of the innovative telemonitoring enhanced care program for chronic heart failure: randomized controlled trial. J Med Internet Res 2020 Jul 08;22(7):e17559 [FREE Full text] [doi: 10.2196/17559] [Medline: 32673222]
19. Hsu M, Chiang C, Chiou A. The effects of a self-regulation programme on self-care behaviour in patients with heart failure: A randomized controlled trial. International Journal of Nursing Studies 2021 Apr;116:103778. [doi: 10.1016/j.ijnurstu.2020.103778]
20. Huber D, Henriksson R, Jakobsson S, Mooe T. Nurse-led telephone-based follow-up of secondary prevention after acute coronary syndrome: One-year results from the randomized controlled NAILED-ACS trial. PLoS One 2017;12(9):e0183963 [FREE Full text] [doi: 10.1371/journal.pone.0183963] [Medline: 28886083]
21. Kalter-Leibovici O, Freimark D, Freedman LS, Kaufman G, Ziv A, Murad H, Israel Heart Failure Disease Management Study (IHF-DMS) investigators. Disease management in the treatment of patients with chronic heart failure who have universal access to health care: a randomized controlled trial. BMC Med 2017 May 01;15(1):90 [FREE Full text] [doi: 10.1186/s12916-017-0855-z] [Medline: 28457231]
22. Oliveira JAD, Cordeiro RG, Rocha RG, Guimarães TCF, Albuquerque DCD. Impacto do monitoramento telefônico em pacientes com insuficiência cardíaca: ensaio clínico randomizado. Acta paul. enferm 2017 Aug;30(4):333-342. [doi: 10.1590/1982-0194201700050]
23. Peng X, Su Y, Hu Z. Home-based telehealth exercise training program in Chinese patients with heart failure: a randomized controlled trial. Med 2018;97(35):a. [doi: 10.1097/md.0000000000012069]
24. Benzo R, McEvoy C. Effect of Health Coaching Delivered by a Respiratory Therapist or Nurse on Self-Management Abilities in Severe COPD: Analysis of a Large Randomized Study. Respir Care 2019 Sep;64(9):1065-1072 [FREE Full text] [doi: 10.4187/respcare.05927] [Medline: 30914491]
25. Cameron-Tucker HL, Wood-Baker R, Joseph L, Walters JA, Schüz N, Walters EH. A randomized controlled trial of telephone-mentoring with home-based walking preceding rehabilitation in COPD. Int J Chron Obstruct Pulmon Dis 2016;11:1991-2000 [FREE Full text] [doi: 10.2147/COPD.S109820] [Medline: 27601892]
26. Jolly K, Sidhu MS, Hewitt CA, Coventry PA, Daley A, Jordan R, et al. Self management of patients with mild COPD in primary care: randomised controlled trial. BMJ 2018 Jun 13;361:k2241 [FREE Full text] [doi: 10.1136/bmj.k2241] [Medline: 29899047]
27. Lavesen M, Ladelund S, Frederiksen AJ, Lindhardt, Overgaard D. Nurse-initiated telephone follow-up on patients with chronic obstructive pulmonary disease improves patient empowerment, but cannot prevent readmissions. Dan Med J 2016 Oct;63(10):1-5 [FREE Full text] [Medline: 27697128]
28. Lee H, Yoon J, Lim Y, Jung H, Kim S, Yoo Y, et al. The effect of nurse-led problem-solving therapy on coping, self-efficacy and depressive symptoms for patients with chronic obstructive pulmonary disease: a randomised controlled trial. Age Ageing 2015 May;44(3):397-403. [doi: 10.1093/ageing/afu201] [Medline: 25548124]
29. Prabhakaran L, Chun Wei Y. Effectiveness of the eCARE programme: a short message service for asthma monitoring. BMJ Health Care Inform 2019 Jun 14;26(1):1-7. [doi: 10.1136/bmjhci-2019-100007]
30. Ulrik CS, Ringbæk T, Laursen LC, Green A, Brøndum E, Frausing E. Effect of tele health care on exacerbations and hospital admissions in patients with chronic obstructive pulmonary disease: a randomized clinical trial. COPD 2015 Sep:1801. [doi: 10.2147/copd.s85596]
31. Soriano JB, García-Río F, Vázquez-Espinosa E, Conforto JI, Hernando-Sanz A, López-Yepes L, et al. A multicentre, randomized controlled trial of telehealth for the management of COPD. Respiratory Medicine 2018 Nov;144:74-81. [doi: 10.1016/j.rmed.2018.10.008]
32. Tupper OD, Gregersen T, Ringbaek T, Brøndum E, Frausing E, Green A, et al. Effect of tele–health care on quality of life in patients with severe COPD: a randomized clinical trial. COPD 2018 Aug;Volume 13:2657-2662. [doi: 10.2147/copd.s164121]
33. Fernandes BSM, Reis IA, Torres HDC. Evaluation of the telephone intervention in the promotion of diabetes self-care: a randomized clinical trial. Rev Lat Am Enfermagem 2016 Aug 29;24:e2719 [FREE Full text] [doi: 10.1590/1518-8345.0632.2719] [Medline: 27579926]
34. Hansen CR, Perrild H, Koefoed BG, Zander M. Video consultations as add-on to standard care among patients with type 2 diabetes not responding to standard regimens: a randomized controlled trial. Eur J Endocrinol 2017 Jun;176(6):727-736. [doi: 10.1530/EJE-16-0811] [Medline: 28325823]
35. Kim Y, Lee H, Seo JM. Integrated Diabetes Self-Management Program Using Smartphone Application: A Randomized Controlled Trial. West J Nurs Res 2022 Apr;44(4):383-394. [doi: 10.1177/0193945921994912] [Medline: 33655794]
36. Li D, Elliott T, Klein G, Ur E, Tang TS. Diabetes Nurse Case Management in a Canadian Tertiary Care Setting: Results of a Randomized Controlled Trial. Can J Diabetes 2017 Jun;41(3):297-304. [doi: 10.1016/j.jcjd.2016.10.012] [Medline: 28318938]

XSL•FO
RenderX

37. Odnoletkova I, Goderis G, Nobels F, Fieuws S, Aertgeerts B, Annemans L, et al. Optimizing diabetes control in people with Type 2 diabetes through nurse-led telecoaching. Diabet Med 2016 Jun;33(6):777-785. [doi: 10.1111/dme.13092] [Medline: 26872105]
38. Sherifali D, Brozic A, Agema P, Punthakee Z, McInnes N, O'Reilly D, et al. Effect of Diabetes Health Coaching on Glycemic Control and Quality of Life in Adults Living With Type 2 Diabetes: A Community-Based, Randomized, Controlled Trial. Can J Diabetes 2021 Oct;45(7):594-600. [doi: 10.1016/j.jcjd.2020.11.012] [Medline: 33582039]
39. Tan CCL, Cheng KKF, Hwang SW, Zhang N, Holroyd E, Wang W. Effect of a Diabetes Self-Efficacy Enhancing Program on Older Adults With Type 2 Diabetes: A Randomized Controlled Trial. Clin Nurs Res 2020 Jun;29(5):293-303. [doi: 10.1177/1054773818792480] [Medline: 30079775]
40. Dadgari F, Hoseini S, Aliyari S, Masoudi S. The effect of sustained nursing consulting via telephone (Tele Nursing) on the quality of life in hypertensive patients. Appl Nurs Res 2017 Jun;35:106-111. [doi: 10.1016/j.apnr.2017.02.023] [Medline: 28532716]
41. Kes D, Polat U. The effect of nurse-led telephone support on adherence to blood pressure control and drug treatment in individuals with primary hypertension: A randomized controlled study. Int J Nurs Pract 2022 Jun;28(3):e12995. [doi: 10.1111/ijn.12995] [Medline: 34318542]
42. Miao J, Wang H, Liu N. The evaluation of a nurse-led hypertension management model in an urban community health care: A randomized controlled trial. Med 2020;99(27):A. [doi: 10.1097/md.0000000000020967]
43. Beaver K, Williamson S, Sutton C, Hollingworth W, Gardner A, Allton B, et al. Comparing hospital and telephone follow-up for patients treated for stage-I endometrial cancer (ENDCAT trial): a randomised, multicentre, non-inferiority trial. BJOG 2017 Jan;124(1):150-160. [doi: 10.1111/1471-0528.14000] [Medline: 27062690]
44. 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]
45. 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]
46. Wheelock AE, Bock MA, Martin EL, Hwang J, Ernest ML, Rugo HS, et al. SIS.NET: a randomized controlled trial evaluating a web-based system for symptom management after treatment of breast cancer. Cancer 2015 Mar 15;121(6):893-899 [FREE Full text] [doi: 10.1002/cncr.29088] [Medline: 25469673]
47. Irewall A, Ögren J, Bergström L, Laurell K, Söderström L, Mooe T. Nurse-Led, Telephone-Based, Secondary Preventive Follow-Up after Stroke or Transient Ischemic Attack Improves Blood Pressure and LDL Cholesterol: Results from the First 12 Months of the Randomized, Controlled NAILED Stroke Risk Factor Trial. PLoS One 2015;10(10):e0139997 [FREE Full text] [doi: 10.1371/journal.pone.0139997] [Medline: 26474055]
48. Wan L, Zhang X, Mo M, Xiong X, Ou C, You L, et al. Effectiveness of Goal-Setting Telephone Follow-Up on Health Behaviors of Patients with Ischemic Stroke: A Randomized Controlled Trial. J Stroke Cerebrovasc Dis 2016 Sep;25(9):2259-2270. [doi: 10.1016/j.jstrokecerebrovasdis.2016.05.010] [Medline: 27371106]
49. Scherrenberg M, Wilhelm M, Hansen D, Völler H, Cornelissen V, Frederix I, et al. The future is now: a call for action for cardiac telerehabilitation in the COVID-19 pandemic from the secondary prevention and rehabilitation section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2020 Jul 02:2047487320939671. [doi: 10.1177/2047487320939671] [Medline: 32615796]
50. Wong AKC, Bayuo J, Wong FKY, Yuen WS, Lee AYL, Chang PK, et al. Effects of a Nurse-Led Telehealth Self-care Promotion Program on the Quality of Life of Community-Dwelling Older Adults: Systematic Review and Meta-analysis. J Med Internet Res 2022 Mar 21;24(3):e31912 [FREE Full text] [doi: 10.2196/31912] [Medline: 35311680]
51. Imanuel Tonapa S, Inayati A, Sithichoksakulchai S, Daryanti Saragih I, Efendi F, Chou F. Outcomes of nurse-led telecoaching intervention for patients with heart failure: A systematic review and meta-analysis of randomised controlled trials. J Clin Nurs 2022 May;31(9-10):1125-1135. [doi: 10.1111/jocn.16025] [Medline: 34535943]
52. Helvaci A, Gok Metin Z. The effects of nurse-driven self-management programs on chronic obstructive pulmonary disease: A systematic review and meta-analysis. J Adv Nurs 2020 Nov;76(11):2849-2871. [doi: 10.1111/jan.14505] [Medline: 32857432]
53. Suh S, Lee MK. Effects of Nurse-Led Telephone-Based Supportive Interventions for Patients With Cancer: A Meta-Analysis. ONF 2017 Jul 1;44(4):E168-E184. [doi: 10.1188/17.onf.e168-e184]
54. Ledwin KM, Lorenz R. The impact of nurse-led community-based models of care on hospital admission rates in heart failure patients: An integrative review. Heart Lung 2021;50(5):685-692. [doi: 10.1016/j.hrtlng.2021.03.079] [Medline: 34107392]
55. Joo JY, Liu MF. Effectiveness of Nurse-Led Case Management in Cancer Care: Systematic Review. Clin Nurs Res 2019 Nov;28(8):968-991. [doi: 10.1177/1054773818773285] [Medline: 29726271]
56. Liang HY, Hann Lin L, Yu Chang C, Mei Wu F, Yu S. Effectiveness of a Nurse-Led Tele-Homecare Program for Patients With Multiple Chronic Illnesses and a High Risk for Readmission: A Randomized Controlled Trial. J Nurs Scholarsh 2021 Mar;53(2):161-170. [doi: 10.1111/jnu.12622] [Medline: 33507626]

XSL•FO
RenderX

57. Jiang Y, Ramachandran HJ, Teo JYC, Leong FL, Lim ST, Nguyen HD, et al. Effectiveness of a nurse-led smartphone-based self-management programme for people with poorly controlled type 2 diabetes: A randomized controlled trial. J Adv Nurs 2022 Apr;78(4):1154-1165. [doi: 10.1111/jan.15178] [Medline: 35170786]
58. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, et al. Effectiveness of Remote Patient Monitoring After Discharge of Hospitalized Patients With Heart Failure: The Better Effectiveness After Transition-Heart Failure (BEAT-HF) Randomized Clinical Trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
59. Kotb A, Cameron C, Hsieh S, Wells G. Comparative effectiveness of different forms of telemedicine for individuals with heart failure (HF): a systematic review and network meta-analysis. PLoS One 2015;10(2):e0118681 [FREE Full text] [doi: 10.1371/journal.pone.0118681] [Medline: 25714962]
60. Richards R, Kinnersley P, Brain K, McCutchan G, Staffurth J, Wood F. Use of Mobile Devices to Help Cancer Patients Meet Their Information Needs in Non-Inpatient Settings: Systematic Review. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e10026 [FREE Full text] [doi: 10.2196/10026] [Medline: 30552082]
61. Toma T, Athanasiou T, Harling L, Darzi A, Ashrafian H. Online social networking services in the management of patients with diabetes mellitus: systematic review and meta-analysis of randomised controlled trials. Diabetes Res Clin Pract 2014 Nov;106(2):200-211. [doi: 10.1016/j.diabres.2014.06.008] [Medline: 25043399]
62. Batalik L, Filakova K, Batalikova K, Dosbaba F. Remotely monitored telerehabilitation for cardiac patients: A review of the current situation. World J Clin Cases 2020 May 26;8(10):1818-1831 [FREE Full text] [doi: 10.12998/wjcc.v8.i10.1818] [Medline: 32518772]
63. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: Systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.013] [Medline: 30784430]
64. Ahmad F, Wysocki RW, Fernandez JJ, Cohen MS, Simcock XC. Patient Perspectives on Telemedicine During the COVID-19 Pandemic. Hand (N Y) 2021 Sep 13:15589447211030692. [doi: 10.1177/15589447211030692] [Medline: 34515550]
65. Martínez-González NA, Tandjung R, Djalali S, Rosemann T. The impact of physician-nurse task shifting in primary care on the course of disease: a systematic review. Hum Resour Health 2015 Jul 07;13:55 [FREE Full text] [doi: 10.1186/s12960-015-0049-8] [Medline: 26149447]
66. Wallace E, Salisbury C, Guthrie B, Lewis C, Fahey T, Smith SM. Managing patients with multimorbidity in primary care. BMJ 2015 Jan 20;350:h176. [doi: 10.1136/bmj.h176] [Medline: 25646760]
67. Nightingale G, Pizzi LT, Barlow A, Barlow B, Jacisin T, McGuire M, et al. The prevalence of major drug-drug interactions in older adults with cancer and the role of clinical decision support software. J Geriatr Oncol 2018 Sep;9(5):526-533. [doi: 10.1016/j.jgo.2018.02.001] [Medline: 29510896]
68. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
69. Souza-Pereira L, Pombo N, Ouhbi S, Felizardo V, Garcia N. Clinical decision support systems for chronic diseases: A Systematic literature review. Comput Methods Programs Biomed 2020 Oct;195:105565. [doi: 10.1016/j.cmpb.2020.105565] [Medline: 32480191]
70. Tarumi S, Takeuchi W, Chalkidis G, Rodriguez-Loya S, Kuwata J, Flynn M, et al. Leveraging Artificial Intelligence to Improve Chronic Disease Care: Methods and Application to Pharmacotherapy Decision Support for Type-2 Diabetes Mellitus. Methods Inf Med 2021 Jun;60(S 01):e32-e43 [FREE Full text] [doi: 10.1055/s-0041-1728757] [Medline: 33975376]
71. Finkelstein J, Jeong IC. Machine learning approaches to personalize early prediction of asthma exacerbations. Ann N Y Acad Sci 2017 Jan;1387(1):153-165 [FREE Full text] [doi: 10.1111/nyas.13218] [Medline: 27627195]
72. Sukumar R, Natarajan RR. Quality of big data in health care. Int J Health Care Qual Assur 2015;28(6):621-634. [doi: 10.1108/ijhcqa-07-2014-0080]
73. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. Eur J Public Health 2016 Feb;26(1):60-64. [doi: 10.1093/eurpub/ckv122] [Medline: 26136462]
74. Gandrup J, Ali S, McBeth J, van der Veer SN, Dixon WG. Remote symptom monitoring integrated into electronic health records: A systematic review. J Am Med Inform Assoc 2020 Nov 01;27(11):1752-1763 [FREE Full text] [doi: 10.1093/jamia/ocaa177] [Medline: 32968785]
75. Lee PA, Greenfield G, Pappas Y. Patients' perception of using telehealth for type 2 diabetes management: a phenomenological study. BMC Health Serv Res 2018 Jul 13;18(1):549 [FREE Full text] [doi: 10.1186/s12913-018-3353-x] [Medline: 30005696]
76. Gajarawala SN, Pelkowski JN. Telehealth Benefits and Barriers. J Nurse Pract 2021 Feb;17(2):218-221 [FREE Full text] [doi: 10.1016/j.nurpra.2020.09.013] [Medline: 33106751]
77. Watzlaf VJM, Zhou L, Dealmeida DR, Hartman LM. A Systematic Review of Research Studies Examining Telehealth Privacy and Security Practices used by Healthcare Providers. Int J Telerehabil 2017 Nov;9(2):39-59 [FREE Full text] [doi: 10.5195/ijt.2017.6231] [Medline: 29238448]

XSL•FO
RenderX

78. 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]
79. Maeder A, Poultney N, Morgan G, Lippiatt R. Patient Compliance in Home-Based Self-Care Telehealth Projects. J Telemed Telecare 2015 Dec;21(8):439-442. [doi: 10.1177/1357633X15612382] [Medline: 26556057]
80. Yokota RTC, Van der Heyden J, Demarest S, Tafforeau J, Nusselder WJ, Deboosere P, et al. Contribution of chronic diseases to the mild and severe disability burden in Belgium. Arch Public Health 2015;73(1):37 [FREE Full text] [doi: 10.1186/s13690-015-0083-y] [Medline: 26240753]
81. Gu J, Chao J, Chen W, Xu H, Wu Z, Chen H, et al. Multimorbidity in the community-dwelling elderly in urban China. Archives of Gerontology and Geriatrics 2017 Jan;68:62-67. [doi: 10.1016/j.archger.2016.09.001]
82. King DE, Xiang J, Pilkerton CS. Multimorbidity Trends in United States Adults, 1988-2014. J Am Board Fam Med 2018;31(4):503-513 [FREE Full text] [doi: 10.3122/jabfm.2018.04.180008] [Medline: 29986975]
83. Salisbury C, Procter S, Stewart K, Bowen L, Purdy S, Ridd M, et al. The content of general practice consultations: cross-sectional study based on video recordings. Br J Gen Pract 2013 Nov 01;63(616):e751-e759. [doi: 10.3399/bjgp13x674431]
84. Muth C, Blom JW, Smith SM. Evidence supporting the best clinical management of patients with multimorbidity and polypharmacy: a systematic guideline review and expert consensus. J Intern Med 2019 Oct;286(4):487 [FREE Full text] [doi: 10.1111/joim.12923] [Medline: 31556213]

Abbreviations

AI: artificial intelligence
COPD: chronic obstructive pulmonary disease
MD: mean difference
OR: odds ratio
PICO: patient, intervention, comparison, and outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
SMD: standardized mean difference

Edited by T Leung; submitted 16.06.22; peer-reviewed by E Ghanbari, M Mehrabi; comments to author 07.09.22; revised version received 22.09.22; accepted 10.10.22; published 02.11.22.

Please cite as:
Lee AYL, Wong AKC, Hung TTM, Yan J, Yang S
Nurse-led Telehealth Intervention for Rehabilitation (Telerehabilitation) Among Community-Dwelling Patients With Chronic Diseases: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(11):e40364
URL: https://www.jmir.org/2022/11/e40364
doi:10.2196/40364
PMID:36322107

©Athena Yin Lam Lee, Arkers Kwan Ching Wong, Tommy Tsz Man Hung, Jing Yan, Shulan Yang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Interactive Remote Patient Monitoring Devices for Managing Chronic Health Conditions: Systematic Review and Meta-analysis

Donato Giuseppe Leo[1,2], BSc, MSc, PhD; Benjamin J R Buckley[1,2], BSc, MSc, PhD; Mahin Chowdhury[3], MRes, MBChB; Stephanie L Harrison[1,2], BSc, MSc, PhD; Masoud Isanejad[4], BSc, MSc, PhD; Gregory Y H Lip[1,2], MD; David J Wright[2], MBChB, MD; Deirdre A Lane[1,2], BSc, PhD; the TAILOR investigators[5]

[1]Department of Cardiovascular and Metabolic Medicine, Institute of Life Course and Medical Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, United Kingdom

[2]Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool, United Kingdom

[3]School of Medicine, University of Liverpool, Liverpool, United Kingdom

[4]Department of Musculoskeletal Ageing, Institute of Life Course and Medical Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, United Kingdom

[5]See acknowledgements

Corresponding Author:
Deirdre A Lane, BSc, PhD
Department of Cardiovascular and Metabolic Medicine
Institute of Life Course and Medical Sciences, Faculty of Health and Life Sciences
University of Liverpool
William Henry Duncan Building
6 West Derby Street
Liverpool, L7 8TX
United Kingdom
Phone: 44 01517949334
Email: Deirdre.Lane@liverpool.ac.uk

Abstract

Background: Telemedicine is an expanding and feasible approach to improve medical care for patients with long-term conditions. However, there is a poor understanding of patients' acceptability of this technology and their rate of uptake.

Objective: The aim of this study was to systematically review the current evidence on telemonitoring in the management of patients with long-term conditions and evaluate the patients' uptake and acceptability of this technology.

Methods: MEDLINE, Scopus, and CENTRAL (the Cochrane Central Register of Controlled Trials) were searched from the date of inception to February 5, 2021, with no language restrictions. Studies were eligible for inclusion if they reported any of the following outcomes: intervention uptake and adherence; study retention; patient acceptability, satisfaction, and experience using the intervention; changes in physiological values; all-cause and cardiovascular-related hospitalization; all-cause and disease-specific mortality; patient-reported outcome measures; and quality of life. In total, 2 reviewers independently assessed the articles for eligibility.

Results: A total of 96 studies were included, and 58 (60%) were pooled for the meta-analyses. Meta-analyses showed a reduction in mortality (risk ratio=0.71, 95% CI 0.56-0.89; P=.003; I^2=0%) and improvements in blood pressure (mean difference [MD]=−3.85 mm Hg, 95% CI −7.03 to −0.68; P=.02; I^2=100%) and glycated hemoglobin (MD=−0.33, 95% CI −0.57 to −0.09; P=.008; I^2=99%) but no significant improvements in quality of life (MD=1.45, 95% CI −0.10 to 3; P=.07; I^2=80%) and an increased risk of hospitalization (risk ratio=1.02, 95% CI 0.85-1.23; P=.81; I^2=79%) with telemonitoring compared with usual care. A total of 12% (12/96) of the studies reported adherence outcomes, and 9% (9/96) reported on satisfaction and acceptance outcomes; however, heterogeneity in the assessment methods meant that a meta-analysis could not be performed.

Conclusions: Telemonitoring is a valid alternative to usual care, reducing mortality and improving self-management of the disease, with patients reporting good satisfaction and adherence. Further studies are required to address some potential concerns regarding higher hospitalization rates and a lack of positive impact on patients' quality of life.

Trial Registration: PROSPERO CRD42021236291; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=236291

XSL•FO
RenderX

(J Med Internet Res 2022;24(11):e35508) doi:10.2196/35508

KEYWORDS

chronic condition; telemonitoring; telemedicine; eHealth; self-monitoring; systematic review; meta-analysis

Introduction

Background

In the United Kingdom, 15 million people live with at least one long-term condition [1], with their care accounting for 70% of the National Health Service budget [1]. Those with long-term conditions have significantly reduced quality of life (QoL) as well as an increased risk of morbidity and mortality [2,3]. Cardiovascular disease, diabetes mellitus, and chronic obstructive pulmonary disease (COPD) are the most common chronic conditions worldwide [4]. Lack of care coordination [5,6] and care planning consultation [5,6] are among the common barriers that patients with long-term conditions face. In addition, the restrictions induced by the COVID-19 pandemic have amplified the challenges that people living with chronic diseases experience in terms of managing their health and accessing health care [7].

Advances in technology have the potential to support patients with long-term conditions in managing their health at home, making the provision of remote health care more accessible and efficient [8]. Web-based health care and telemedicine include the remote delivery of care using communication technology (eg, videoconference software, web-based applications, and home-based health measurement) to enable consultations between patients and their care team, providing continuous monitoring of relevant health parameters. This allows health care professionals to promptly respond to changes in patient health status and adapt their clinical management in real time [9].

Objectives

Recent evidence has deemed telemedicine feasible for patients with long-term conditions and effective in terms of improving medical care [10]. As telemedicine is a rapidly expanding and changing field, recent umbrella reviews [10,11] that consider older primary studies have potentially made conclusions based on noncontemporary data. Therefore, the aim of this systematic review was to update and expand the current literature on telemonitoring by better defining the interventions included to encompass the role that interactive, 2-way communication devices play in improving the care of patients with long-term conditions, as well as evaluate patient uptake and acceptability of this technology.

Methods

Overview

This systematic review was registered on PROSPERO (International Prospective Register of Systematic Reviews; CRD42021236291) and conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [12].

This review aimed to address the following research questions: (1) What is the rate of uptake, patient retention, and patient satisfaction when using 2-way (patient-health care provider) remote patient monitoring devices to manage chronic health conditions? (2) What factors are associated with patient retention and satisfaction when using 2-way (patient-health care provider) remote patient monitoring devices to manage chronic health conditions? (3) Does the use of 2-way (patient-health care provider) remote patient monitoring devices for the management of chronic health conditions affect patient outcomes (eg, changes in physiological measurements, QoL, all-cause and cardiovascular-related hospitalizations, and all-cause and disease-specific mortality)?

Criteria for Considering Studies to Include in the Review

Studies carried out in any setting aiming to evaluate telemonitoring interventions for participants with at least one chronic condition among the following—cardiovascular disease, COPD, or diabetes mellitus—were eligible for inclusion. All randomized controlled trials (RCTs) and nonrandomized trials, before-and-after (pre-post) studies, and interrupted time series were considered for inclusion. Cross-sectional studies and case reports were excluded. Qualitative studies were included to assess participant satisfaction. Ongoing studies (if any) were also considered and presented in a dedicated table.

Participants

Adult participants (aged ≥18 years) were eligible for inclusion in this review if they reported one or more of the following chronic health conditions: cardiovascular diseases (eg, coronary artery disease, atrial fibrillation, stroke, heart failure, and hypertension), COPD, or diabetes mellitus.

Intervention

Interventions designed to remotely collect health information from patients using digital technologies and electronically transfer the information to health care professionals for monitoring and assessment were eligible for inclusion. Only interventions where the participant received a digital device for remote patient monitoring and the participant or their caregiver took physiological measurements and either input the information into the device or the device automatically uploaded the data were included. Health devices suitable for inclusion had to transmit data to the participant's health care team, and the participant's health care team had to monitor the information received, assessing it and making appropriate changes to the participant's treatment accordingly. A 2-way exchange of information was required for a study to be included.

Comparator

Studies in which usual care or a different intervention was used as control or comparator were also considered as eligible for inclusion, as were studies that did not have a control group.

XSL•FO
RenderX

Outcomes

The primary outcomes of interest were (1) intervention uptake (number of people willing to participate in the intervention) and adherence (level of commitment of the patient to the prescribed intervention); (2) study retention (number of people who completed the intervention); and (3) patient acceptability (level of acceptance of the intervention by the participants), satisfaction (number of participants pleased with the intervention), and experience using the intervention. Secondary outcomes included (1) changes in physiological measurements (eg, oxygen saturation, blood pressure [BP], and blood glucose level); (2) all-cause and cardiovascular-related hospitalizations; (3) all-cause and disease-specific mortality; (4) patient-reported outcome measures (eg, mental well-being, depression, and anxiety questionnaires); and (5) QoL, quality-adjusted life years, and any other health economic outcomes reported in the studies. All the studies that reported one or more of these outcomes were considered eligible for inclusion.

Search Strategy

The search strategy was developed by the review team, which agreed on the key terms. Medical Subject Headings terms and synonyms for the different terms, such as "telemedicine," "digital monitoring," and "e-health" (Table S1 in Multimedia Appendix 1 [13-163]), were used and combined with Boolean operators, proximity operators, truncations, and wildcards. MEDLINE, Scopus, and CENTRAL (the Cochrane Central Register of Controlled Trials) were searched from the date of inception to February 5, 2021, for relevant studies. There were no language restrictions, but the availability of the full text was a requirement for inclusion. Search results were managed using EndNote (version X9.3.3; Clarivate Analytics).

Study Selection

Two reviewers (MC and DGL) independently screened the titles and abstracts of the studies retrieved from the databases against the search criteria. Additional screening of the preliminary results was independently undertaken by 3 other reviewers (BB, SH, and MI). The full texts of all potentially relevant articles were retrieved and independently assessed by the reviewers in duplicate. Any disagreement was resolved through discussion with the senior author (DL).

Data Extraction

Data extraction was conducted independently by 2 reviewers (DGL and MC). The following information was extracted: (1) authors, year, country, and reference; (2) study aim; (3) study characteristics (study design and sample size); (4) participant characteristics (age, sex, and ethnicity); (5) health condition; (6) intervention (type of telemedicine device, input of the data [manual or automated], delivery of the intervention, staff involved, duration and frequency of the intervention, and follow-up points); (7) comparators (usual care, different intervention, or no intervention); and (8) outcomes (primary and secondary, as reported in the study).

Risk of Bias Assessment

Six authors (DGL, MC, BB, SH, MI, and DL) independently assessed the individual studies for risk of bias in duplicate, and any discrepancies were resolved via discussion or referral to a third reviewer, as required. For RCTs, the Cochrane Risk of Bias version 2 tool [164] was used. For nonrandomized studies, the Risk Of Bias In Non-randomized Studies of Interventions [165] was used.

Data Synthesis

Meta-analyses were conducted on comparable studies. Primary and secondary outcome effect measures with 95% CIs were pooled using the RevMan software (The Cochrane Collaboration) [166]. The results are presented visually using forest plots. Where continuous data were not homogeneous, an estimate of the standardized mean difference (MD) with 95% CIs was calculated. For studies in which quantitative data were too few or too heterogeneous, a narrative synthesis approach was used.

Dichotomous analyses were conducted using the number of events and total sample size as reported in the included studies. The results of the selected studies were combined using the Mantel-Haenszel method. Effect sizes are expressed as relative risk and 95% CIs. Random effect models were applied to all meta-analyses owing to heterogeneity in study characteristics and populations. Heterogeneity was quantitatively assessed using the Higgins index (I^2).

For the analysis of QoL, the postintervention scores, as reported in the included studies, were used. Where the SD was not reported, it was calculated using the calculator function available in RevMan. For analysis of changes in physiological parameters (BP and glycated hemoglobin [HbA_{1c}]) and QoL, the results of the selected studies were combined using the generic inverse variance method. Effect sizes are expressed as the MD and SD.

Findings from the included qualitative studies will be synthesized elsewhere using a meta-aggregative approach to data synthesis.

Results

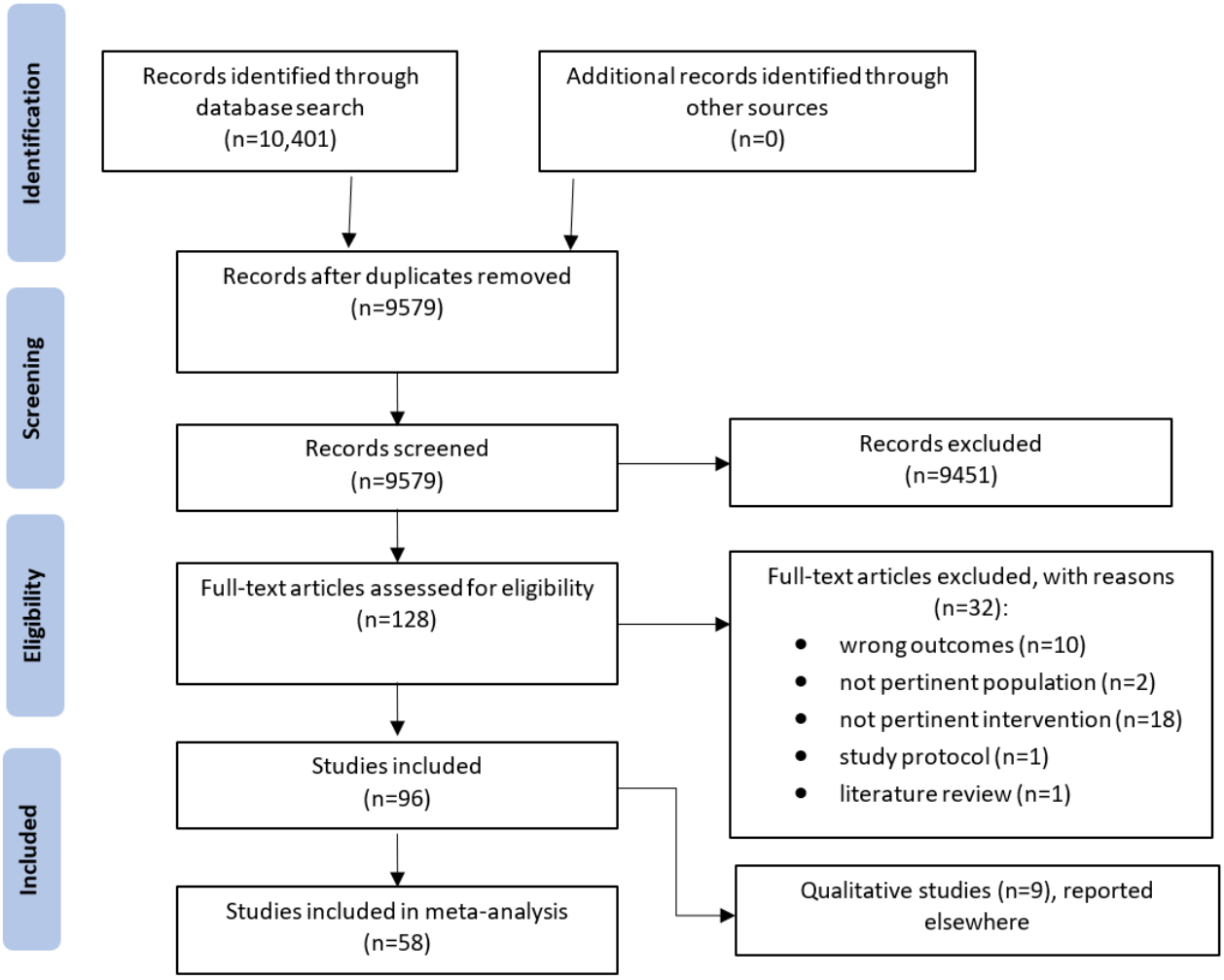
Overview

The database searches identified 10,401 papers. After independent screening of titles and abstracts by 2 study authors, 98.77% (10,273/10,401) of papers were determined to be duplicates or not eligible. After screening against the inclusion and exclusion criteria, of the remaining 128 papers, 96 (75%) were included. No ongoing studies were found (Figure 1). A full list of the excluded studies with reasons for exclusion is provided in Table S2 in Multimedia Appendix 1. Full texts of all 96 included papers [13-109] were retrieved.

No study reporting outcomes related to intervention uptake, study retention, and patient acceptability were identified in our search and, therefore, these outcomes could not be analyzed. The following analyses and results concern only patient adherence and satisfaction as well as clinical and patient-reported outcomes.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram depicting the screening and study selection process.



Characteristics of the Included Studies

The included studies were published between 1998 and 2020, with sample sizes ranging from 20 [36,99] to 3562 [102] participants and a total sample of 26,167 participants. The mean age ranged from 44 [22] to 78 [107] years, and the proportion of men varied from 25% [51] to 76% [91]. Most of the included studies were conducted in the United Kingdom (21/96, 22%) and the United States (29/96, 30%), with additional studies conducted in Belgium (2/96, 2%), Canada (4/96, 4%), Denmark (5/96, 5%), Poland (2/96, 2%), Singapore (2/96, 2%), South Korea (2/96, 2%), Spain (9/96, 9%), Germany (4/96, 4%), and Italy (6/96, 6%; Multimedia Appendix 2 [13-109,136]). In addition, the following countries had 1% (1/96) of the studies each: Australia [37], China [99], Finland [106], Greece [49], Hong Kong [28], Israel [14], Japan [66], Malaysia [67], the Netherlands [25], and Taiwan [29] (Multimedia Appendix 2).

Populations in the included studies comprised patients with diabetes (27/96, 28% of the studies), cardiovascular disease (stroke, atrial fibrillation, hypertension, and heart failure; 52/96, 54% of the studies), COPD (12/96, 12% of the studies), and mixed chronic conditions (diabetes, hypertension, and COPD; 5/96, 5% of the studies; Multimedia Appendix 2).

Types of Interventions

The studies varied in their design, type of telemonitoring system used, and method of delivery (Multimedia Appendix 2). Most (64/96, 67%) were RCTs, with 4% (4/96) being nonrandomized controlled studies, 2% (2/96) being cluster randomized studies, 10% (10/96) being longitudinal studies, 4% (4/96) being retrospective analyses, 3% (3/96) being pre-post analyses, and 9% (9/96) having a mixed methods or qualitative design. Most studies (88/96, 92%) used telemonitoring systems that collected patient information via computers, tablets, or dedicated devices (eg, modem) and transferred these data to a web-based server. Some studies collected patient data via SMS text message (3/96, 3%) or by telephone (4/96, 4%). A total of 4% (4/96) of the studies provided educational videos to increase the patients' knowledge of the disease. The length of the intervention was highly variable, with 5% (5/96) of the studies assessing it over a short period (7-45 days), 21% (20/96) assessing it over a 2- to 4-month period, and most interventions (76/96, 79%) lasting 6 to 12 months. The follow-up periods were inconsistent among the studies and, where present, ranged from 3 to 18 months.

Types of Comparators

Most studies (79/96, 82%) compared the intervention with usual care, which consisted of routine visits (outpatient clinics) and in-person consultations with general practitioners or the hospital

XSL•FO
RenderX

care team (Multimedia Appendix 2). A total of 10% (10/96) of the studies did not have a control group. A total of 1% (1/96) of the studies asked the control group to manually record their data in a diary. In total, 2% (2/96) of the studies used educational videos in the control group to improve patients' knowledge of the disease, another 2% (2/96) compared the intervention with another telemonitoring device, and 1% (1/96) compared the intervention (telemonitoring device) with telephone communication. A total of 1% (1/96) of the studies used a similar intervention as the control group comparing patients with and without heart failure.

Types of Outcomes

In total, 12 studies reported adherence to the intervention, including 9 (75%) in patients with cardiovascular disease, 2 (17%) in patients with diabetes, and 1 (8%) in patients with COPD (Multimedia Appendix 2). Patient satisfaction with the intervention was assessed in 9% (9/96) of the studies (2/9, 22% in patients with cardiovascular disease; 3/9, 33% in patients with diabetes; 2/9, 22% in patients with COPD; and 2/9, 22% in a mixed population; Multimedia Appendix 2).

Most studies (31/96, 32%) reported changes in physiological parameters, which varied depending on the population observed, with 39% (12/31) of these studies reporting BP values for patients with cardiovascular disease, 55% (17/31) reporting HbA_{1c} values for patients with diabetes, and 6% (2/31) reporting multiple physiological values in mixed populations (Multimedia Appendix 2).

Hospital admission during the intervention was recorded in 29% (28/96) of the studies (21/28, 75% in patients with cardiovascular disease; 4/28, 14% in patients with COPD; and 3/28, 11% in a mixed sample), and death was noted in 18% (17/96) of the studies (14/17, 82% in patients with cardiovascular disease; 2/17, 12% in patients with COPD; and 1/17, 6% in a mixed population; Multimedia Appendix 2).

QoL before and after the intervention was recorded in 22% (21/96) of the studies (11/21, 52% in patients with cardiovascular disease; 2/21, 10% in patients with diabetes; 6/21, 29% in patients with COPD; and 2/21, 10% in a mixed population; Multimedia Appendix 2).

Excluded Studies

A total of 25% (32/128) of the studies assessed for eligibility [110-141] were excluded. A summary of these studies can be found in Table S2 in Multimedia Appendix 1. Most (18/32, 56%) were excluded as they were not related to a telemonitoring intervention, 6% (2/32) included disease populations not covered in this review, 31% (10/32) reported outcomes outside the scope of this review, 3% (1/32) were literature reviews, and 3% (1/32) were study protocols.

Risk of Bias Assessment

A summary of the risk of bias assessment of the included studies can be found in Tables S3-S5 in Multimedia Appendix 1. Overall, most RCTs (48/66, 73%) and non-RCTs (17/20, 85%) included in this review showed either some concerns or a high risk of bias. Most RCT studies (45/66, 68%) showed either some concerns or a high risk of bias in the randomization process as well as in the selection of the reported results. Some RCTs (18/66, 27%) showed either some concerns or a high risk of bias in missing outcome data. Few RCTs (17/66, 26%) showed either some concerns or a high risk of bias in the measurement of the outcomes.

Most of the non-RCTs (18/20, 90%) showed either some concerns or a high risk of bias in the *bias due to confounding* category. A total of 50% (10/20) of the studies showed either some concerns or a high risk of bias in the *bias in measurement of outcomes* category. Few of the non-RCTs (9/20, 45%) showed either some concerns or a high risk of bias in the *bias due to missing data* category as well as in the *bias due to deviations from the intended intervention* category.

The studies included in the meta-analyses were assessed for publication bias. Funnel plots and Egger tests were performed only where ≥10 studies were available [167].

Funnel plots for the outcomes of systolic BP (SBP), HbA_{1c}, and mortality can be found in Figures S1-S6 in Multimedia Appendix 1. The Egger test results revealed no evidence of publication bias for SBP, HbA_{1c}, or mortality.

Ongoing Studies

The database search did not return any protocols for ongoing studies. Searches on ClinicalTrials.gov (updated to February 5, 2021) identified 22 ongoing studies [142-163] (n=14, 64% on patients with cardiovascular disease; n=4, 18% on patients with diabetes; and n=4, 18% on patients with COPD), which are reported in detail in Table S6 in Multimedia Appendix 1.

Primary Outcomes

Adherence

Adherence was assessed in 12 studies at different time points: 1 month (n=3, 25%) [51,66,84], 6 weeks (n=2, 17%) [58,103], 2 months (n=1, 8%) [13], 3 months (n=1, 8%) [30], 6 months (n=4, 33%) [42,48,59,92], and 12 months (n=1, 8%) [36]. Of the 12 studies, 7 (58%) [13,36,42,48,58,59,92] demonstrated a benefit of telemonitoring on patient adherence when compared with a comparator, whereas 4 (33%) [30,51,66,84] showed no difference when compared with a comparator. A total of 8% (1/12) of the studies [103] compared 2 telemonitoring systems and showed that educational support combined with telemonitoring positively influenced adherence compared with telemonitoring alone. Owing to variations in how adherence was defined in the studies, a meta-analysis was not performed. A summary of these studies is presented in Table 1.

XSL•FO
RenderX

Table 1. Studies examining the impact of telemonitoring interventions versus comparator on adherence (N=12).

Study type and authors, year, and country	Study population, N	Condition	Intervention type, number of participants, age (years), men (n [%])	Comparator, number of participants, age (years), mean (n [%])	Outcomes	Follow-up	Impact of telemonitoring
Randomized controlled trials							
Ong et al [84], 2016, United States	1437	CHF[a]	Automated upload of data on dedicated device or software, 715, mean 73 (SD not reported), men: 382 (53.4); women: 333 (46.6)	Usual care, 722, mean 73 (SD not reported), men: 382 (53.4); women: 333 (46.6)	Adherence electronically recorded; 82.7%	1 month	=[b]
Gallagher et al [51], 2017, United States	40	HF[c]	Manual upload of data on dedicated device or software, 20, median 68 (IQR 49-79), men: 15 (75); women: 5 (25)	Usual care, 20, median 62 (IQR 52-75), men: 15 (75); women: 5 (25)	Adherence recorded electronically; 81% in both groups	1 month	=
Kotooka et al [66], 2018, Japan	183	CHF	Automated upload of data on dedicated device or software, 93, mean 67.1 (SD 12.8), men: 51 (56); women: 39 (44)	Usual care, 91, mean 65.4 (SD 15.6), men: 56 (61); women: 35 (39)	Adherence recorded electronically; 90% at 12 months	12 months	=
Varon et al [103], 2015, United Kingdom	534	HF	Docobo system (telemonitoring only), 135, mean 69.1 (SD 12.6), not reported	Motiva system (telemonitoring+ educational videos), 399, mean 69.1 (SD 12.6), not reported	Adherence assessed by the amount of missing data during the telemonitoring period	6 weeks	–[d]
Kardas et al [58], 2016, Poland	60	Type 2 diabetes	Automated upload of data on dedicated device or software, 30, mean 59.9 (SD 5.31), men: 17 (57); women: 13 (43)	Usual care, 30, mean 59 (SD 8.9), men: 19 (63); women: 11 (47)	Adherence expressed as medication taken vs medication prescribed; 92.9%	6 weeks	+[e]
Cho et al [30], 2009, South Korea	69	Type 2 diabetes	Mobile app, 35, mean 51.1 (SD 13.1), 26 men; 74 women[f]	Web-based telemonitoring system, 34, mean 51.1 (SD 13.1), 26 men; 74 women[f]	Adherence, self-reported; >70% in both groups	3 months	=
Seto et al [92], 2012, Canada	100	CHF	Automated upload of data on dedicated device or software, 50, mean 55.1 (SD 13.7), men: 41 (82); women: 9 (18)	Usual care, 50, mean 52.3 (SD 13.7), men: 38 (76); women: 12 (24)	Adherence registered electronically; 80%	6 months	+
Evans et al [48], 2016, United States	441	HF and healthy	Disease group: automated upload of data on dedicated device or software, 421, mean 71.8 (SD 8.8), 46 men; 54 women[f]	Healthy group: automated upload of data on dedicated device or software, 20, mean 72.2 (SD 4.3), 50 men; 50 women[f]	Adherence checking the amount of data against the participants' time spent in the study; between 71% and 81%	6 months	+
Nonrandomized studies							
Agboola et al [13], 2013, United States	30	Hypertension	Web-based device, 15, mean 61.9 (SD not reported), 20 men; 80 women[f]	Mobile blood pressure device, 15, mean 61.6 (SD not reported), 20 men; 80 women[f]	Adherence recorded electronically based on frequency of data transmission	2 months	+
Domingo et al [42], 2012, Spain	97	HF	Automated upload of data on dedicated device or software, 46, mean 66.5 (SD 11.5), men: 14 (30); women: 32 (70)	Usual care, 51, mean 66.5 (SD 11.5), men: 15 (30); women: 36 (70)	Adherence based on the number of educational videos watched; between 67% and 85%	6 months	+

XSL•FO
RenderX

Study type and authors, year, and country	Study population, N	Condition	Intervention type, number of participants, age (years), men (n [%])	Comparator, number of participants, age (years), mean (n [%])	Outcomes	Follow-up	Impact of telemonitoring
Karg et al [59], 2012, Germany	36	COPD[g]	Automated upload of data on dedicated device or software, 36, mean 67.9 (SD 6.9), men: 27 (75); women: 9 (25)	N/A[h]	Adherence: use of the device for at least two-thirds of working days; full compliance	6 months	+
De Lusignan et al [36], 2001, United Kingdom	20	CHF	Manual upload of data on dedicated device or software, 10, mean 75.2 (SD not reported), not reported	Usual care, 10, mean 75.2 (SD not reported), not reported	Adherence based on the frequency of the uploaded data; 90%	12 months	+

[a]CHF: congestive heart failure.

[b]No differences between telemonitoring and usual care.

[c]HF: heart failure.

[d]Negative impact of telemonitoring over comparator.

[e]Positive impact of telemonitoring over comparator.

[f]Absolute value not reported in the paper.

[g]COPD: chronic obstructive pulmonary disease.

[h]N/A: not applicable.

Satisfaction

Patient satisfaction with the intervention was assessed in 9 studies (n=2, 22% in patients with cardiovascular disease; n=3, 33% in patients with diabetes; n=2, 22% in patients with COPD; and n=2, 22% in a mixed population; Table 2). A total of 56% (5/9) of the studies [22,28,42,78,91] demonstrated a benefit of telemonitoring on patient satisfaction when compared with a comparator, whereas 44% (4/9) [30,43,44,95] showed no difference when compared with a comparator. Owing to variations in how satisfaction was defined in the studies, a meta-analysis was not performed. A summary of these studies is provided in Table 2.

XSL•FO
RenderX

Table 2. Studies examining the impact of telemonitoring interventions versus comparator on satisfaction (N=9).

Study type and authors, year, and country	Study population, N	Condition	Intervention type, number of participants, age (years), mean (n [%])	Comparator, number of participants, age (years), mean (n [%])	Outcomes	Follow-up	Impact of telemonitoring
Randomized controlled trials							
Bergenstal et al [22], 2005, United States	47	Type 2 diabetes	Automated data transmitted via modem, 24, mean 44 (SD 17), 37 men; 63 women[a]	Data transmitted via telephone, 23, mean 45 (SD 13), 39 men; 61 women[a]	Satisfaction: 5-point questionnaire; 4.30 in the phone group and 4.52 in the modem group	4 weeks	=[b]
Chau et al [28], 2012, Hong Kong	40	COPD[c]	Manual upload of data on dedicated device or software, 22, mean 73.5 (SD 6), men: 21 (95); women: 1 (5)	Usual care, 18, mean 72.2 (SD 6), men: 18 (100); women: 0 (0)	Satisfaction: 10-item questionnaire based on a 5-point system; 91%	2 months	+[d]
Edmonds et al [44], 1998, Canada	35	Type 2 diabetes	Mobile phone data transmission, 16, not reported, not reported	Usual care, 19, not reported, not reported	Satisfaction: patient questionnaire	3 months	Further studies required
Cho et al [30], 2009, South Korea	69	Type 2 diabetes	Mobile app, 35, mean 51.1 (SD 13.1), 26 men; 74 women[a]	Web-based telemonitoring system, 34, mean 51.1 (SD 13.1), 26 men; 74 women[a]	Satisfaction: questionnaire, internet vs phone; 81% vs 79%, respectively	3 months	=
Sicotte et al [95], 2011, Canada	46	COPD	Manual upload of data on dedicated device or software, 23, mean 73.7 (SD 9.6), men: 13 (56); women: 10 (44)	Usual care, 23, mean 75.4 (SD 9.7), men: 13 (56); women: 10 (44)	Satisfaction: 5-point questionnaire; 4.50 score	3 months	=
Domingo et al [42], 2012, Spain	97	HF[e]	Automated upload of data on dedicated device or software, 46, mean 66.5 (SD 11.5), men: 14 (30); women: 32 (70)	Usual care, 51, mean 66.5 (SD 11.5), men: 15 (30); women: 36 (70)	Satisfaction: 10-point questionnaire; 8.4 score	6 months	+
Nonrandomized studies							
Schoenfeld et al [91], 2004, United States	59	CHF[f]	Manual upload of data on dedicated device or software, 59, mean 64 (SD 14), men: 45 (76); women: 14 (24)	N/A[g]	Satisfaction: 3-point questionnaire; 98.1% indicating ease of use of the device	7 days	+
Donate-Martinez et al [43], 2016, Spain	74	Chronic conditions (COPD, type 2 diabetes, and HF)	Manual upload of data on dedicated device or software, 74, mean 67.95 (SD 11.14), men: 49 (66); women: 25 (44)	N/A	Satisfaction: 11-item questionnaire with 10-point score; 8.63 score overall	12 months	=
Mira-Solves et al [77], 2014, Spain	410	Chronic conditions (type 2 diabetes, hypertension, CHF, and COPD)	Automated upload of data on dedicated device or software, 410, not reported, 64 men; 36 women[a]	N/A	Satisfaction: questionnaire, 89.4% were satisfied with the ease of use.	24 months	+

[a]Absolute value not reported in the paper.

[b]No differences between telemonitoring and usual care.

[c]COPD: chronic obstructive pulmonary disease.

[d]Positive impact of telemonitoring over comparator.

[e]HF: heart failure.

[f]CHF: congestive HF.

[g]N/A: not applicable.

XSL•FO
RenderX

Secondary Outcomes

QoL Measurement

Studies included in the meta-analyses were pooled by comparable scales (eg, the Short Form 36 Health Survey Questionnaire) and end points (eg, 6 or 12 months), with 8% (8/96) of the studies [16,31,33,35,47,96,101,104] included in the meta-analyses.

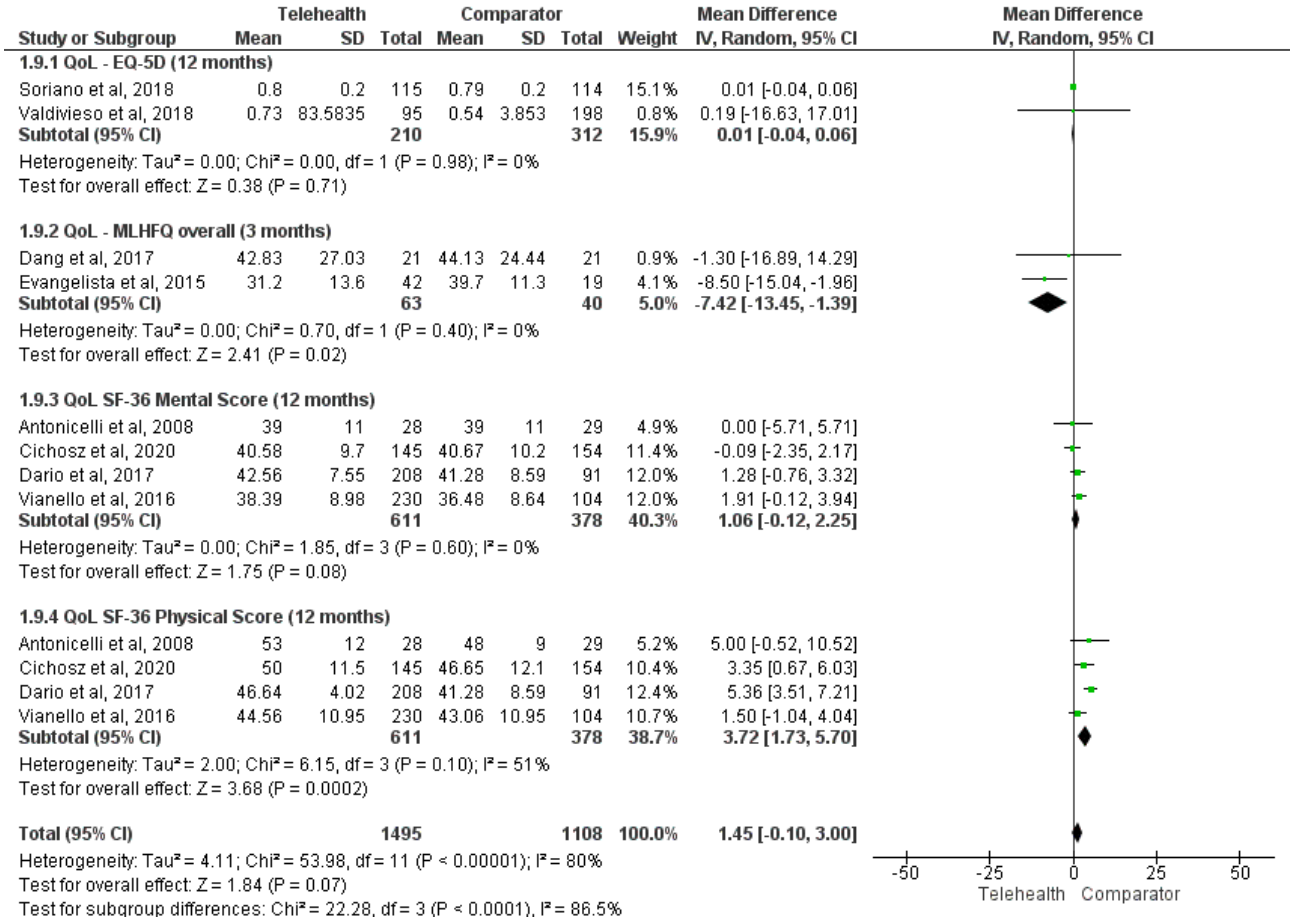
A total of 50% (4/8) of these studies [16,31,35,104] reported the Short Form 36 Health Survey Questionnaire scores (mental and physical) at comparable end points (12 months) and were included in the meta-analyses (Figure 2 [15,31,35,47,96,101,104,136], subgroups 1.9.3 and 1.9.4). From the meta-analysis, telemonitoring showed greater improvements compared with usual care on physical component scores (weighted MD=3.72, 95% CI 1.73-5.70; P<.001; I^2=51%; Figure 2) compared with the comparator but no difference in mental component scores (weighted MD=1.06, 95% CI −0.12 to 2.25; P=.08; I^2=0%; Figure 3 [15,39,40,50,60,64,84,96,101,105,107]).

In total, 25% (2/8) of the studies [96,101] reported EQ-5D scores at comparable end points (12 months) and were included in the meta-analysis (Figure 2, subgroup 1.9.1). There was no difference in QoL between the groups (weighted MD=0.01, 95% CI −0.04 to 0.06; P=.71; I^2=0%)

A total of 25% (2/8) of the studies [33,47] using the Minnesota Living with Heart Failure Questionnaire overall scores at 3 months were included in the meta-analysis (Figure 2, subgroup 1.9.2), demonstrating that the telemonitoring group showed greater improvements in QoL (weighted MD=−7.42, 95% CI −13.45 to −1.39; P=.02; I^2=0%) compared with the comparator.

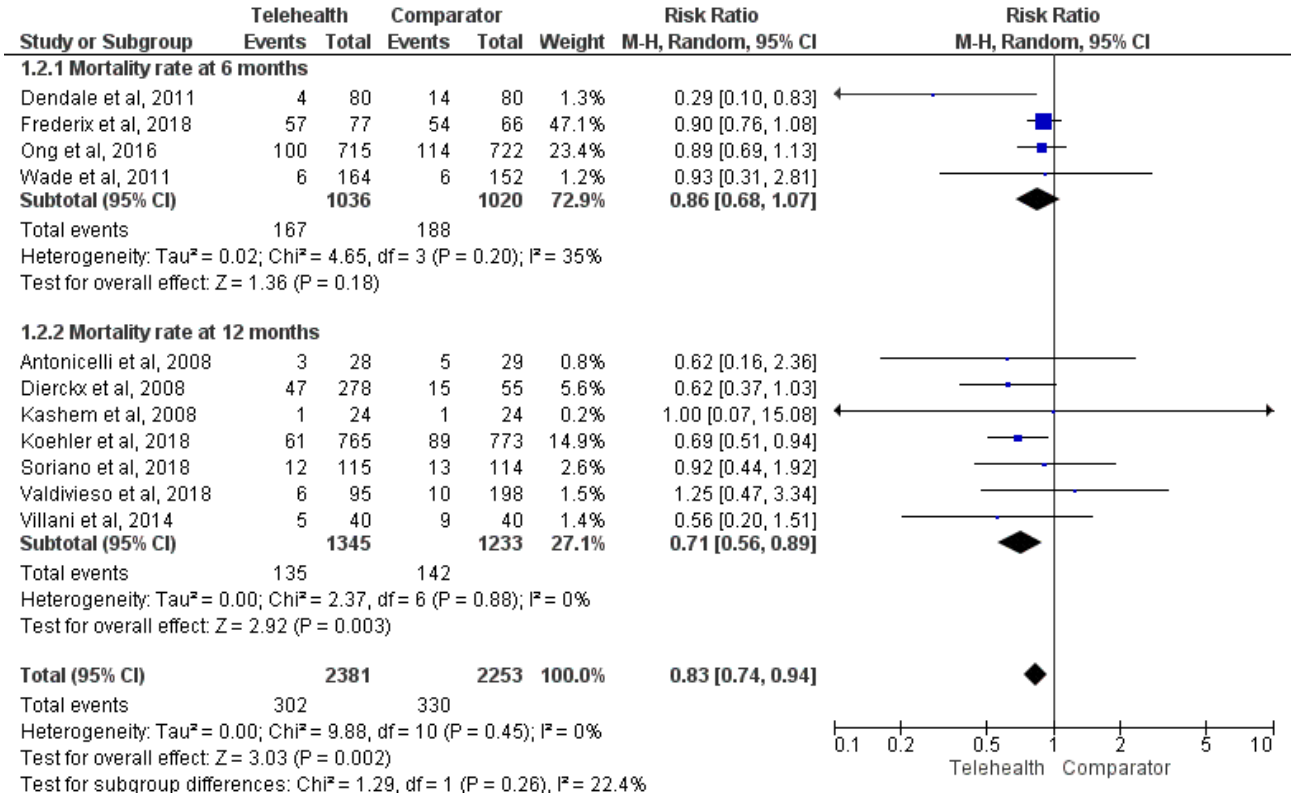
A total of 14% (13/96) of the studies [20,23,36,43,58,62,65,70,92,100,103,107,108] could not be included in the meta-analysis because they reported different time points and used different questionnaires to assess QoL. Of these 13 studies, 4 (31%) reported a significant improvement in QoL in the telemonitoring group compared with usual care at 6 weeks [58], 6 months [92,100], and 12 months [43] measured using a variety of questionnaires (Minnesota Living with Heart Failure Questionnaire [92], EQ-5D [43,58], and 15D [100]), whereas 9 (69%) reported no difference in QoL between telemonitoring and usual care at 4 weeks [70], 6 weeks [65,103], 7 weeks [70], 3 months [36], 6 months [23,62,107], 9 months [108], and 12 months [36]. A total of 8% (1/13) of the studies [20] reported significant improvement in QoL in the usual care group compared with telemonitoring at 2 and 6 months using the St George's Respiratory Questionnaire.

Figure 2. Impact of telemonitoring versus comparator on quality of life (QoL). 1.9.1: EQ-5D; 1.9.2: Minnesota Living with Heart Failure Questionnaire (MLHFQ); 1.9.3: SF-36 mental score; and 1.9.4: SF-36 physical component [15,31,35,47,96,101,104,136].



XSL•FO
RenderX

Figure 3. Impact of telemonitoring versus comparator on the mortality rate at 6 and 12 months. The study by Mortara et al [80] was not included in the mortality meta-analyses because of the use of a composite outcome of mortality and hospitalization where absolute mortality results were not available. The study by Seto et al [92] was not included in the mortality meta-analyses because of 0 events in the control group [15,39,40,50,60,64,84,96,101,105,107].



Mortality

Meta-analyses for mortality were conducted at the 6- and 12-month follow-up (Figure 3). Sensitivity analyses were conducted at the 6- and 12-month follow-up excluding studies at high risk of bias and at 12 months excluding non-RCTs (Figure S1 in Multimedia Appendix 1). A sensitivity analysis with the exclusion of non-RCTs at 6 months was not conducted as all the studies included were RCTs.

A total of 11 studies contributed to the all-cause mortality meta-analysis: 4 (36%) [39,50,84,107] (N=2056) provided data at 6 months, and 7 (64%) [16,40,61,64,96,101,105] (N=2578) provided data at 12 months. There was no significant difference in all-cause mortality between telemonitoring and the comparator at 6 months (risk ratio [RR]=0.86, 95% CI 0.68-1.07; P=.18; I^2=35%; Figure 3). This finding was consistent when studies evaluated as having a high risk of bias were excluded (Figure S1 in Multimedia Appendix 1). There was a significantly lower risk of all-cause mortality with telemonitoring than with the comparator at 12 months (RR=0.71, 95% CI 0.56-0.89; P=.003; I^2=0%; Figure 3). This finding was consistent following the exclusion of non-RCTs and studies evaluated as having a high risk of bias (Figure S1 in Multimedia Appendix 1).

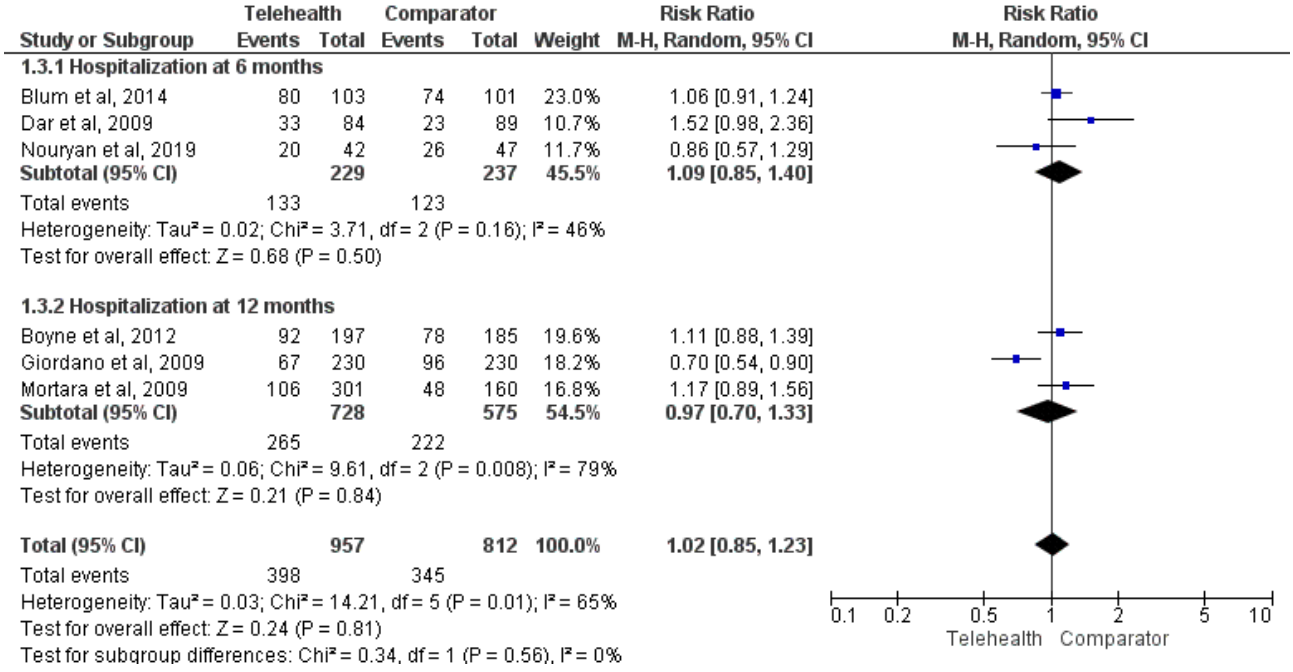
Hospitalization

Meta-analyses for hospitalization at the 6- and 12-month follow-up were conducted (Figure 4 [23,25,34,52,80,83]), with sensitivity analyses excluding studies classified as having a high risk of bias (Figure S2 in Multimedia Appendix 1) and a subgroup analysis including only studies on patients with heart failure (12/96, 12%). Subgroup analyses for studies on patients with COPD and multiple chronic conditions were not possible because of a lack of absolute values or comparator [29,85].

A total of 8 studies contributed to the all-cause hospitalization meta-analyses: 3 (38%) [23,34,83] (n=466) provided data at 6 months, and 5 (62%) [25,52,80,96,101] (n=1825) provided data at 12 months. There was no significant difference in the risk of all-cause hospitalization between the groups at 6 months (RR=1.09, 95% CI 0.85-1.40; P=.50; I^2=46%) or 12 months (RR=0.97, 95% CI 0.70-1.33; P=.84; I^2=79%; Figure 4). This result was also consistent after the exclusion of studies evaluated as having a high risk of bias (Figure S2 in Multimedia Appendix 1). The meta-analysis that included only patients with heart failure showed no difference in the risk of hospitalization between the telemonitoring and comparator groups (RR=0.99, 95% CI 0.81-1.22; P=.94; I^2=69%; Figure S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 4. Impact of telemonitoring versus comparator on hospitalization at 6 and 12 months [23,25,34,52,80,83].



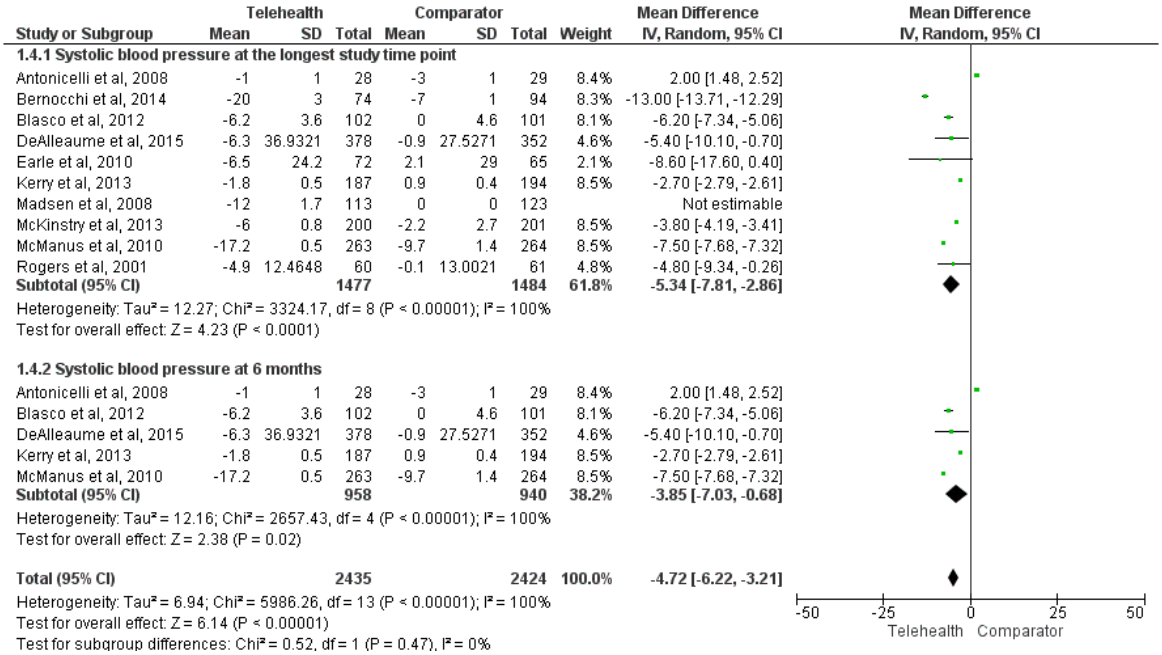
Changes in BP

A total of 10% (10/96) of the studies [16,17,24,38,45,62,72,75,77] reporting on the change in SBP and 8% (8/96) of the studies [15,17,24,45,62,72,75,77,90] reporting on the change in diastolic BP (DBP) between a telemonitoring intervention and usual care were included in the meta-analyses. Further details on the analyses of BP are provided in Multimedia Appendix 1.

Changes in SBP

SBP was significantly reduced in the telemonitoring group (n=1477) compared with that in the usual care group (n=1484; weighted MD=−5.34 mm Hg, 95% CI −7.81 to −2.86; P<.001; I^2=100%; Figure 5 [15,17,24,38,45,62,72,75,77,90]). In the subgroup analysis according to study time points, similar results were observed for SBP at 6 months (weighted MD=−3.85 mm Hg, 95% CI −7.03 to −0.68; P=.02; I^2=100%; Figure 5) and 12 months (weighted MD=−3.85 mm Hg, 95% CI −7.03 to −0.68; P=.02; I^2=100%; Figure S3 in Multimedia Appendix 1) in favor of telemonitoring.

Figure 5. Impact of telemonitoring versus usual care on changes in systolic blood pressure (mean difference) at the longest study time point and at 6 months [15,17,24,38,45,62,72,75,77,90].



XSL•FO
RenderX

The sensitivity analysis, excluding studies where the SD was not reported directly [38,45,90], did not materially change the results (weighted MD=−5.19 mm Hg, 95% CI −8.01 to −2.37; P<.001; I^2=100%; Figure S3 in Multimedia Appendix 1). The sensitivity analysis was also performed excluding studies with a high risk of bias (Figure S3 in Multimedia Appendix 1); the results remained in favor of telemonitoring (weighted MD=–2.84 mm Hg, 95% CI −4.22 to −1.46; P<.001; I^2=98%).

Changes in DBP

A meta-analysis including the longest time point demonstrated a significant reduction in DBP in favor of telemonitoring (n=1218) compared with the comparator (n=1255; weighted MD=−2.83 mm Hg, 95% CI −3.98 to −1.68; P<.001; I^2=99%; Figure S4 in Multimedia Appendix 1). In the subgroup analysis, a similar result was observed for DBP reduction at 6 months (weighted MD=−5.44 mm Hg, 95% CI −9.00 to −1.87; P=.003; I^2=100%; Figure S4 in Multimedia Appendix 1) in favor of telemonitoring but not for DBP at 12 months (weighted MD=−1.09 mm Hg, 95% CI −4.76 to 2.57; P=.56; I^2=97%; Figure S4 in Multimedia Appendix 1). Sensitivity analyses at the longest time point excluding studies with high risk of bias (Figure S4 in Multimedia Appendix 1) showed no significant reduction in DBP in the telemonitoring group (weighted MD=−1.07 mm Hg, 95% CI −2.58 to 0.44; P=.16; I^2=98%) compared with usual care.

Changes in HbA_{1c}

A total of 19% (18/96) of the studies reported on HbA_{1c}, and all the studies (18/18, 100%) compared telemonitoring with usual care, with 61% (11/18; n=3277) included in the meta-analysis [27,30,35,46,49,58,63,87,89,94,109]. Further details on the excluded studies for the meta-analysis are provided in Multimedia Appendix 1.

The duration of the interval before and after varied, with 18% (2/11) of these studies reporting a 6-week assessment [58,87], 45% (5/11) [27,30,46,49,63] reporting 3-month assessments, 9% (1/11) reporting 9-month assessments [109], and 27% (3/11) [35,89] reporting 12-month assessments. A sensitivity analysis was performed excluding studies with a high risk of bias [58,94].

The overall mean change in HbA_{1c} is shown in Figure S5 in Multimedia Appendix 1. The pooled estimate showed a reduction in the mean change in HbA_{1c} in the telemonitoring group (n=1703; weighted MD=−0.33, 95% CI −0.57 to −0.09; P=.008; I^2=99%; Figure S5 in Multimedia Appendix 1). The results did not materially change after the sensitivity analysis excluding studies at high risk of bias [58,87] (Figure S5 in Multimedia Appendix 1). Subgroup analyses according to study time points showed no significant difference in the change in HbA_{1c} values between telemonitoring and the comparator (Figure S5 in Multimedia Appendix 1).

Discussion

Principal Findings

Our results suggest that telemonitoring interventions are associated with good patient adherence and satisfaction. Although this review did not demonstrate improvements in QoL with telemonitoring, there was evidence to suggest reductions in all-cause mortality and improvements in BP and blood glucose control. Conversely, there was evidence to suggest that telemonitoring interventions may be associated with a higher rate of hospitalizations, which could be interpreted as a positive role of telemonitoring in detecting patients' health issues more than usual care.

Comparison With Prior Work

Our review showed improvements in physiological parameters (BP and blood glucose) in patients receiving telemonitoring interventions. These findings demonstrate the positive role of telemonitoring in improving patients' self-management of their conditions. This is in line with other reviews that have shown similar improvements in hypertension [168] and type 2 diabetes self-management [169] after telemonitoring interventions.

The studies included in this review consistently showed that patients receiving telemonitoring interventions had lower all-cause mortality compared with patients receiving usual care. A recent umbrella review [170] examining the effects of telemonitoring on mortality in several clinical populations (cardiovascular, COPD, and neurological) reported similar findings for the cardiovascular population, where the mortality rate was either reduced in the telemedicine users or remained unchanged compared with usual care. The same review [170] did not find any difference in mortality between telemonitoring and usual care in patients with COPD. The impact on death is an important outcome when considering the administration of remote interventions over in-person visits, and the reduced mortality rate with telemonitoring reported in our review suggests the effectiveness of telemonitoring for patients with chronic conditions.

Surprisingly, the overall results of our review showed a higher risk of hospitalization among patients undergoing telemonitoring interventions. There is inconsistency in the previous literature on the role that telemonitoring plays in reducing the risk of rehospitalization, with some studies reporting no differences compared with usual care [171] and others concluding that telemonitoring is an effective tool to reduce all-cause hospitalization in adults with heart failure [172]. Thurmond et al [173] noted the importance that the type of telemonitoring intervention has on its acceptability by patients and, consequently, their adherence to it, which, when poor, may influence the rate of rehospitalization. This would suggest the need to identify common characteristics of effective telemonitoring interventions (or "active ingredients") that facilitate patient acceptability. It may also be possible that increased hospitalizations with telemonitoring is a positive finding (ie, reasons for hospitalization may be identified earlier by telemonitoring, and hospitalization may be initiated earlier than with usual care, averting serious outcomes and death). Hypothetically, this could have contributed to the reduced

XSL•FO
RenderX

mortality at 12 months; however, future research is needed to substantiate this.

The results of this review are in line with those of previous systematic reviews assessing patient satisfaction with telemonitoring interventions [174,175]. From qualitative reports, the convenience of decreased travel time and costs and the reassurance of being monitored are the most likely reasons for patients preferring telemonitoring over usual care [176]. It is important to note that patient satisfaction may differ with the type of telemonitoring device used; indeed, available evidence suggests that higher patient satisfaction is reported for videoconferences and devices that allow for automated data transmission [174].

The included studies did not report significant improvements in the QoL of patients receiving a telemonitoring intervention compared with usual care. Our findings confirm previous reviews [177,178] while expanding the results to populations outside care homes [178] and including study designs other than RCTs [177]. Although telemonitoring does not seem to improve QoL compared with usual care, previous findings [178] have shown important benefits of telemonitoring in improving patients' confidence in accessing health care services.

Strengths and Limitations

This review used a strict definition of telemonitoring, only including studies that used a device to collect health measures and facilitated 2-way communication or action between the patient and health care team. Despite the inclusion of studies with low methodological quality, sensitivity analyses were conducted where appropriate, reducing the potential for bias to affect the results of this review. The studies included in this review presented a wide range of telemonitoring interventions that differed in the personnel involved, administration of the intervention, and technology used and that were examined in a variety of populations with different long-term conditions, thus making the results highly generalizable. A robust methodology was used, with independent screening and data extraction by 2 reviewers and risk of bias assessment in duplicate.

Several limitations are noteworthy. First, despite our initial plans to investigate uptake, patient retention and satisfaction, and associated factors when using 2-way (patient-health care provider) remote patient monitoring devices to manage chronic health conditions, no studies reported uptake and retention outcomes and, therefore, these outcomes could not be reported in this review. Most of the included studies assessed similar outcomes but used different measurement tools, thus making comparison difficult, particularly in studies investigating patient adherence [13,30,36,42,48,51,58,59,66,84,92,103] and satisfaction [22,28,30,42-44,78,91,95] with the intervention. Second, despite our efforts to define the best search strategy to identify all relevant articles for our review, the possible omission of papers because of the heterogeneity in the key terms used by the authors cannot be ruled out. We did not conduct any searches for gray literature. Third, most outcomes analyzed in this review have been infrequently investigated in the literature (eg, mortality was reported only in 17/96, 18% of the included studies; adherence was reported in only 12/96, 12% of the studies; and satisfaction was reported in only 9/96, 9% of the studies), and further research is required to properly assess the effects of telemonitoring on these outcomes. Moreover, some conditions (eg, COPD) were underrepresented as few studies investigating the effects of telemonitoring interventions on these populations were available; thus, we could not conduct a separate meta-analysis for each condition. The type and quality of usual care also varied throughout the included studies, which may have influenced the results in favor of or against telemonitoring.

Conclusions

Telemonitoring is a promising tool to manage long-term conditions, with the potential to reduce the associated costs and alleviate patient difficulties in accessing primary health care. Patient satisfaction and adherence to telemonitoring appear, overall, to be promising. Although telemonitoring resulted in improvement in physiological parameters and reduced all-cause mortality compared with usual care, there was no improvement in QoL and an increased risk of hospitalization with telemonitoring. Although the latter may be a positive finding indicating earlier detection of health issues and action (resulting in hospitalization), this result warrants further investigation. Telemonitoring is expanding rapidly, more so since the COVID-19 pandemic, and has been shown to be a viable alternative to usual care for the management of patients with long-term health conditions.

Acknowledgments

The authors would like to thank all the Telehealth and Artificial Intelligence for Older People (TAILOR) investigators: Dr Asan Akpan, Dr Girvan Burnside (University of Liverpool), Mr Robert Halhead, Mr Stephen Hope, Mr Peter Levene, Mr Geoff Hayllar (Docobo Ltd, Leatherhead, United Kingdom), Mr Peter Almond (Mersey Care National Health Service Trust), Ms Sarah Dyas (Clinical Research Network, North West Coast), and Ms Lindsay Sharples (Innovation Agency). The authors would also like to thank Dr Marie Held (University of Liverpool) for her help in translating some of the included papers from German to English. This project has received funding from the Liverpool Clinical Commissioning Group, Research Capability Funding (LCCG_RCF20-21_01).

Conflicts of Interest

BJRB has received research funding from the Bristol Myers Squibb (BMS)-Pfizer Alliance. SLH has received an investigator-initiated grant from BMS. GYHL has been a consultant and speaker for the BMS-Pfizer Alliance, Boehringer Ingelheim, and Daiichi-Sankyo. No fees were received personally. DJW has been a consultant and speaker for Medtronic and

XSL•FO
RenderX

Boston Scientific. DAL has received investigator-initiated educational grants from BMS; been a speaker for Boehringer Ingelheim, Bayer, and the BMS-Pfizer Alliance; and consulted for Boehringer Ingelheim, Bayer, and the BMS-Pfizer Alliance, all outside the submitted work.

Multimedia Appendix 1
Supplementary figures and tables that were not included in the main manuscript.
[DOC File , 1432 KB - jmir_v24i11e35508_app1.doc]

Multimedia Appendix 2
Summary of the included studies (N=96).
[DOCX File , 57 KB - jmir_v24i11e35508_app2.docx]

References

1. Goodwin N, Sonola L, Thiel V, Kodner D. Co-ordinated care for people with complex chronic conditions: key lessons and markers for success. The King's Fund. 2013. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/co-ordinated-care-for-people-with-complex-chronic-conditions-kingsfund-oct13.pdf [accessed 2021-10-01]
2. Masnoon N, Kalisch Ellett L, Shakib S, Caughey GE. Predictors of mortality in the older population: the role of polypharmacy and other medication and chronic disease-related factors. Drugs Aging 2020 Oct;37(10):767-776. [doi: 10.1007/s40266-020-00794-7] [Medline: 32885396]
3. Tyack Z, Frakes KA, Barnett A, Cornwell P, Kuys S, McPhail S. Predictors of health-related quality of life in people with a complex chronic disease including multimorbidity: a longitudinal cohort study. Qual Life Res 2016 Oct;25(10):2579-2592. [doi: 10.1007/s11136-016-1282-x] [Medline: 27048497]
4. Noncommunicable diseases. World Health Organization. 2021 Apr 13. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2022-04-01]
5. Coulter A, Roberts S, Dixon A. Delivering better services for people with long-term conditions: building the house of care. The King's Fund. 2013 Oct. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/delivering-better-services-for-people-with-long-term-conditions.pdf [accessed 2021-10-01]
6. How to deliver high-quality, patient-centred, cost-effective care: consensus solutions from the voluntary sector. The King's Fund. 2010. URL: https://www.kingsfund.org.uk/sites/default/files/how-to-deliver-high-quality-patient-centred-cost-effective-care-16-september-2010-kings-fund.pdf [accessed 2021-10-01]
7. Topriceanu CC, Wong A, Moon JC, Hughes AD, Bann D, Chaturvedi N, et al. Evaluating access to health and care services during lockdown by the COVID-19 survey in five UK national longitudinal studies. BMJ Open 2021 Mar 18;11(3):e045813 [FREE Full text] [doi: 10.1136/bmjopen-2020-045813] [Medline: 33737441]
8. Salisbury C, Thomas C, O'Cathain A, Rogers A, Pope C, Yardley L, et al. TElehealth in CHronic disease: mixed-methods study to develop the TECH conceptual model for intervention design and evaluation. BMJ Open 2015 Feb 06;5(2):e006448 [FREE Full text] [doi: 10.1136/bmjopen-2014-006448] [Medline: 25659890]
9. Alvarez P, Sianis A, Brown J, Ali A, Briasoulis A. Chronic disease management in heart failure: focus on telemedicine and remote monitoring. Rev Cardiovasc Med 2021 Jun 30;22(2):403-413 [FREE Full text] [doi: 10.31083/j.rcm2202046] [Medline: 34258907]
10. Eze ND, Mateus C, Cravo Oliveira Hashiguchi T. Telemedicine in the OECD: an umbrella review of clinical and cost-effectiveness, patient experience and implementation. PLoS One 2020 Aug 13;15(8):e0237585 [FREE Full text] [doi: 10.1371/journal.pone.0237585] [Medline: 32790752]
11. Snoswell CL, Chelberg G, De Guzman KR, Haydon HH, Thomas EE, Caffery LJ, et al. The clinical effectiveness of telehealth: a systematic review of meta-analyses from 2010 to 2019. J Telemed Telecare (forthcoming) 2021 Jun 29:1357633X211022907. [doi: 10.1177/1357633X211022907] [Medline: 34184580]
12. 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]
13. Agboola S, Havasy R, Myint-U K, Kvedar J, Jethwani K. The impact of using mobile-enabled devices on patient engagement in remote monitoring programs. J Diabetes Sci Technol 2013 May 01;7(3):623-629 [FREE Full text] [doi: 10.1177/193229681300700306] [Medline: 23759394]
14. Amir O, Ben-Gal T, Weinstein JM, Schliamser J, Burkhoff D, Abbo A, et al. Evaluation of remote dielectric sensing (ReDS) technology-guided therapy for decreasing heart failure re-hospitalizations. Int J Cardiol 2017 Aug 01;240:279-284 [FREE Full text] [doi: 10.1016/j.ijcard.2017.02.120] [Medline: 28341372]
15. Antonicelli R, Testarmata P, Spazzafumo L, Gagliardi C, Bilo G, Valentini M, et al. Impact of telemonitoring at home on the management of elderly patients with congestive heart failure. J Telemed Telecare 2008;14(6):300-305. [doi: 10.1258/jtt.2008.071213] [Medline: 18776075]

XSL•FO
RenderX

16. Antonicelli R, Mazzanti I, Abbatecola AM, Parati G. Impact of home patient telemonitoring on use of β-blockers in congestive heart failure. Drugs Aging 2010 Oct 01;27(10):801-805. [doi: 10.2165/11538210-000000000-00000] [Medline: 20883060]
17. Bernocchi P, Scalvini S, Bertacchini F, Rivadossi F, Muiesan ML. Home based telemedicine intervention for patients with uncontrolled hypertension--a real life non-randomized study. BMC Med Inform Decis Mak 2014 Jun 12;14:52 [FREE Full text] [doi: 10.1186/1472-6947-14-52] [Medline: 24920046]
18. Baron JS, Hirani S, Newman SP. A randomised, controlled trial of the effects of a mobile telehealth intervention on clinical and patient-reported outcomes in people with poorly controlled diabetes. J Telemed Telecare 2017 Feb;23(2):207-216. [doi: 10.1177/1357633X16631628] [Medline: 26880694]
19. Baron JS, Hirani SP, Newman SP. Investigating the behavioural effects of a mobile-phone based home telehealth intervention in people with insulin-requiring diabetes: results of a randomized controlled trial with patient interviews. J Telemed Telecare 2017 Jun;23(5):503-512. [doi: 10.1177/1357633X16655911] [Medline: 27377790]
20. Bentley CL, Mountain GA, Thompson J, Fitzsimmons DA, Lowrie K, Parker SG, et al. A pilot randomised controlled trial of a Telehealth intervention in patients with chronic obstructive pulmonary disease: challenges of clinician-led data collection. Trials 2014 Aug 06;15:313 [FREE Full text] [doi: 10.1186/1745-6215-15-313] [Medline: 25100550]
21. Beran M, Asche SE, Bergdall AR, Crabtree B, Green BB, Groen SE, et al. Key components of success in a randomized trial of blood pressure telemonitoring with medication therapy management pharmacists. J Am Pharm Assoc (2003) 2018;58(6):614-621 [FREE Full text] [doi: 10.1016/j.japh.2018.07.001] [Medline: 30077564]
22. Bergenstal RM, Anderson RL, Bina DM, Johnson ML, Davidson JL, Solarz-Johnson B, et al. Impact of modem-transferred blood glucose data on clinician work efficiency and patient glycemic control. Diabetes Technol Ther 2005 Apr;7(2):241-247. [doi: 10.1089/dia.2005.7.241] [Medline: 15857225]
23. Blum K, Gottlieb SS. The effect of a randomized trial of home telemonitoring on medical costs, 30-day readmissions, mortality, and health-related quality of life in a cohort of community-dwelling heart failure patients. J Card Fail 2014 Jul;20(7):513-521. [doi: 10.1016/j.cardfail.2014.04.016] [Medline: 24769270]
24. Blasco A, Carmona M, Fernández-Lozano I, Salvador CH, Pascual M, Sagredo PG, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. J Cardiopulm Rehabil Prev 2012;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
25. Boyne JJ, Vrijhoef HJ, Crijns HJ, De Weerd G, Kragten J, Gorgels AP, TEHAF investigators. Tailored telemonitoring in patients with heart failure: results of a multicentre randomized controlled trial. Eur J Heart Fail 2012 Jul;14(7):791-801 [FREE Full text] [doi: 10.1093/eurjhf/hfs058] [Medline: 22588319]
26. Buis LR, Roberson DN, Kadri R, Rockey NG, Plegue MA, Danak SU, et al. Understanding the feasibility, acceptability, and efficacy of a clinical pharmacist-led mobile approach (BPTrack) to hypertension management: mixed methods pilot study. J Med Internet Res 2020 Aug 11;22(8):e19882 [FREE Full text] [doi: 10.2196/19882] [Medline: 32780026]
27. Bujnowska-Fedak MM, Puchała E, Steciwko A. The impact of telehome care on health status and quality of life among patients with diabetes in a primary care setting in Poland. Telemed J E Health 2011 Apr;17(3):153-163. [doi: 10.1089/tmj.2010.0113] [Medline: 21375410]
28. Chau JP, Lee DT, Yu DS, Chow AY, Yu WC, Chair SY, et al. A feasibility study to investigate the acceptability and potential effectiveness of a telecare service for older people with chronic obstructive pulmonary disease. Int J Med Inform 2012 Oct;81(10):674-682. [doi: 10.1016/j.ijmedinf.2012.06.003] [Medline: 22789911]
29. Chen YH, Lin YH, Hung CS, Huang CC, Yeih DF, Chuang PY, et al. Clinical outcome and cost-effectiveness of a synchronous telehealth service for seniors and nonseniors with cardiovascular diseases: quasi-experimental study. J Med Internet Res 2013 Apr 24;15(4):e87 [FREE Full text] [doi: 10.2196/jmir.2091] [Medline: 23615318]
30. Cho JH, Lee HC, Lim DJ, Kwon HS, Yoon KH. Mobile communication using a mobile phone with a glucometer for glucose control in type 2 patients with diabetes: as effective as an Internet-based glucose monitoring system. J Telemed Telecare 2009;15(2):77-82. [doi: 10.1258/jtt.2008.080412] [Medline: 19246607]
31. Cichosz SL, Udsen FW, Hejlesen O. The impact of telehealth care on health-related quality of life of patients with heart failure: results from the Danish TeleCare North heart failure trial. J Telemed Telecare 2020;26(7-8):452-461. [doi: 10.1177/1357633X19832713] [Medline: 30975047]
32. Cleland JG, Louis AA, Rigby AS, Janssens U, Balk AH, TEN-HMS Investigators. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. J Am Coll Cardiol 2005 May 17;45(10):1654-1664 [FREE Full text] [doi: 10.1016/j.jacc.2005.01.050] [Medline: 15893183]
33. Dang S, Dimmick S, Kelkar G. Evaluating the evidence base for the use of home telehealth remote monitoring in elderly with heart failure. Telemed J E Health 2009 Oct;15(8):783-796. [doi: 10.1089/tmj.2009.0028] [Medline: 19831704]
34. Dar O, Riley J, Chapman C, Dubrey SW, Morris S, Rosen SD, et al. A randomized trial of home telemonitoring in a typical elderly heart failure population in North West London: results of the Home-HF study. Eur J Heart Fail 2009 Mar;11(3):319-325 [FREE Full text] [doi: 10.1093/eurjhf/hfn050] [Medline: 19174529]
35. Dario C, Toffanin R, Calcaterra F, Saccavini C, Stafylas P, Mancin S, et al. Telemonitoring of type 2 diabetes mellitus in Italy. Telemed J E Health 2017 Feb;23(2):143-152. [doi: 10.1089/tmj.2015.0224] [Medline: 27379995]

XSL•FO
RenderX

36. de Lusignan S, Wells S, Johnson P, Meredith K, Leatham E. Compliance and effectiveness of 1 year's home telemonitoring. The report of a pilot study of patients with chronic heart failure. Eur J Heart Fail 2001 Dec;3(6):723-730 [FREE Full text] [doi: 10.1016/s1388-9842(01)00190-8] [Medline: 11738225]
37. De San Miguel K, Smith J, Lewin G. Telehealth remote monitoring for community-dwelling older adults with chronic obstructive pulmonary disease. Telemed J E Health 2013 Sep;19(9):652-657. [doi: 10.1089/tmj.2012.0244] [Medline: 23808885]
38. DeAlleaume L, Parnes B, Zittleman L, Sutter C, Chavez R, Bernstein J, et al. Success in the achieving cardiovascular excellence in Colorado (A CARE) home blood pressure monitoring program: a report from the shared networks of Colorado Ambulatory Practices and Partners (SNOCAP). J Am Board Fam Med 2015;28(5):548-555 [FREE Full text] [doi: 10.3122/jabfm.2015.05.150024] [Medline: 26355126]
39. Dendale P, De Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar;14(3):333-340 [FREE Full text] [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]
40. Dierckx R, Cleland JG, Pellicori P, Zhang J, Goode K, Putzu P, et al. If home telemonitoring reduces mortality in heart failure, is this just due to better guideline-based treatment? J Telemed Telecare 2015 Sep;21(6):331-339. [doi: 10.1177/1357633X15574947] [Medline: 25766855]
41. Dinesen B, Haesum LK, Soerensen N, Nielsen C, Grann O, Hejlesen O, et al. Using preventive home monitoring to reduce hospital admission rates and reduce costs: a case study of telehealth among chronic obstructive pulmonary disease patients. J Telemed Telecare 2012 Jun;18(4):221-225. [doi: 10.1258/jtt.2012.110704] [Medline: 22653618]
42. Domingo M, Lupón J, González B, Crespo E, López R, Ramos A, et al. Evaluation of a telemedicine system for heart failure patients: feasibility, acceptance rate, satisfaction and changes in patient behavior: results from the CARME (CAtalan Remote Management Evaluation) study. Eur J Cardiovasc Nurs 2012 Dec;11(4):410-418. [doi: 10.1016/j.ejcnurse.2011.02.003] [Medline: 21402493]
43. Doñate-Martínez A, Ródenas F, Garcés J. Impact of a primary-based telemonitoring programme in HRQOL, satisfaction and usefulness in a sample of older adults with chronic diseases in Valencia (Spain). Arch Gerontol Geriatr 2016;62:169-175. [doi: 10.1016/j.archger.2015.09.008] [Medline: 26446784]
44. Edmonds M, Bauer M, Osborn S, Lutfiyya H, Mahon J, Doig G, et al. Using the Vista 350 telephone to communicate the results of home monitoring of diabetes mellitus to a central database and to provide feedback. Int J Med Inform 1998;51(2-3):117-125. [doi: 10.1016/s1386-5056(98)00109-9] [Medline: 9794328]
45. Earle KA, Istepanian RS, Zitouni K, Sungoor A, Tang B. Mobile telemonitoring for achieving tighter targets of blood pressure control in patients with complicated diabetes: a pilot study. Diabetes Technol Ther 2010 Jul;12(7):575-579. [doi: 10.1089/dia.2009.0090] [Medline: 20597833]
46. Egede LE, Williams JS, Voronca DC, Knapp RG, Fernandes JK. Randomized controlled trial of technology-assisted case management in low income adults with type 2 diabetes. Diabetes Technol Ther 2017 Aug;19(8):476-482. [doi: 10.1089/dia.2017.0006] [Medline: 28581821]
47. Evangelista LS, Lee JA, Moore AA, Motie M, Ghasemzadeh H, Sarrafzadeh M, et al. Examining the effects of remote monitoring systems on activation, self-care, and quality of life in older patients with chronic heart failure. J Cardiovasc Nurs 2015;30(1):51-57 [FREE Full text] [doi: 10.1097/JCN.0000000000000110] [Medline: 24365871]
48. Evans J, Papadopoulos A, Silvers CT, Charness N, Boot WR, Schlachta-Fairchild L, et al. Remote health monitoring for older adults and those with heart failure: adherence and system usability. Telemed J E Health 2016 Jun;22(6):480-488 [FREE Full text] [doi: 10.1089/tmj.2015.0140] [Medline: 26540369]
49. Fountoulakis S, Papanastasiou L, Gryparis A, Markou A, Piaditis G. Impact and duration effect of telemonitoring on HbA1c, BMI and cost in insulin-treated diabetes mellitus patients with inadequate glycemic control: a randomized controlled study. Hormones (Athens) 2015;14(4):632-643 [FREE Full text] [doi: 10.14310/horm.2002.1603] [Medline: 26188234]
50. Frederix I, Vanderlinden L, Verboven AS, Welten M, Wouters D, De Keulenaer G, et al. Long-term impact of a six-month telemedical care programme on mortality, heart failure readmissions and healthcare costs in patients with chronic heart failure. J Telemed Telecare 2019 Jun;25(5):286-293. [doi: 10.1177/1357633X18774632] [Medline: 29742959]
51. Gallagher BD, Moise N, Haerizadeh M, Ye S, Medina V, Kronish IM. Telemonitoring adherence to medications in heart failure patients (TEAM-HF): a pilot randomized clinical trial. J Card Fail 2017 Apr;23(4):345-349 [FREE Full text] [doi: 10.1016/j.cardfail.2016.11.001] [Medline: 27818309]
52. Giordano A, Scalvini S, Zanelli E, Corrà U, Longobardi GL, Ricci VA, et al. Multicenter randomised trial on home-based telemanagement to prevent hospital readmission of patients with chronic heart failure. Int J Cardiol 2009 Jan 09;131(2):192-199. [doi: 10.1016/j.ijcard.2007.10.027] [Medline: 18222552]
53. Grady M, Cameron H, Levy BL, Katz LB. Remote health consultations supported by a diabetes management Web application with a new glucose meter demonstrates improved glycemic control. J Diabetes Sci Technol 2016 May;10(3):737-743 [FREE Full text] [doi: 10.1177/1932296815622646] [Medline: 26685995]

XSL•FO
RenderX

54. Grant S, Hodgkinson J, Schwartz C, Bradburn P, Franssen M, Hobbs FR, et al. Using mHealth for the management of hypertension in UK primary care: an embedded qualitative study of the TASMINH4 randomised controlled trial. Br J Gen Pract 2019 Sep;69(686):e612-e620 [FREE Full text] [doi: 10.3399/bjgp19X704585] [Medline: 31262847]
55. Greenwood DA, Blozis SA, Young HM, Nesbitt TS, Quinn CC. Overcoming clinical inertia: a randomized clinical trial of a telehealth remote monitoring intervention using paired glucose testing in adults with type 2 diabetes. J Med Internet Res 2015 Jul 21;17(7):e178 [FREE Full text] [doi: 10.2196/jmir.4112] [Medline: 26199142]
56. Hanley J, Fairbrother P, McCloughan L, Pagliari C, Paterson M, Pinnock H, et al. Qualitative study of telemonitoring of blood glucose and blood pressure in type 2 diabetes. BMJ Open 2015 Dec 23;5(12):e008896 [FREE Full text] [doi: 10.1136/bmjopen-2015-008896] [Medline: 26700275]
57. Istepanian RS, Zitouni K, Harry D, Moutosammy N, Sungoor A, Tang B, et al. Evaluation of a mobile phone telemonitoring system for glycaemic control in patients with diabetes. J Telemed Telecare 2009;15(3):125-128. [doi: 10.1258/jtt.2009.003006] [Medline: 19364893]
58. Kardas P, Lewandowski K, Bromuri S. Type 2 diabetes patients benefit from the COMODITY12 mHealth system: results of a randomised trial. J Med Syst 2016 Dec;40(12):259. [doi: 10.1007/s10916-016-0619-x] [Medline: 27722974]
59. Karg O, Weber M, Bubulj C, Esche B, Weber N, Geiseler J, et al. Akzeptanz einer telemedizinischen Intervention bei Patienten mit chronisch-obstruktiver Lungenerkrankung. Dtsch Med Wochenschr 2012 Mar;137(12):574-579. [doi: 10.1055/s-0031-1299033] [Medline: 22415618]
60. Kashem A, Droogan MT, Santamore WP, Wald JW, Bove AA. Managing heart failure care using an internet-based telemedicine system. J Card Fail 2008 Mar;14(2):121-126. [doi: 10.1016/j.cardfail.2007.10.014] [Medline: 18325458]
61. Kashem A, Droogan MT, Santamore WP, Wald JW, Marble JF, Cross RC, et al. Web-based Internet telemedicine management of patients with heart failure. Telemed J E Health 2006 Aug;12(4):439-447. [doi: 10.1089/tmj.2006.12.439] [Medline: 16942416]
62. Kerry SM, Markus HS, Khong TK, Cloud GC, Tulloch J, Coster D, et al. Home blood pressure monitoring with nurse-led telephone support among patients with hypertension and a history of stroke: a community-based randomized controlled trial. CMAJ 2013 Jan 08;185(1):23-31 [FREE Full text] [doi: 10.1503/cmaj.120832] [Medline: 23128283]
63. Kim SI, Kim HS. Effectiveness of mobile and internet intervention in patients with obese type 2 diabetes. Int J Med Inform 2008 Jun;77(6):399-404. [doi: 10.1016/j.ijmedinf.2007.07.006] [Medline: 17881285]
64. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan BA, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
65. Konstam V, Gregory D, Chen J, Weintraub A, Patel A, Levine D, et al. Health-related quality of life in a multicenter randomized controlled comparison of telephonic disease management and automated home monitoring in patients recently hospitalized with heart failure: SPAN-CHF II trial. J Card Fail 2011 Feb;17(2):151-157. [doi: 10.1016/j.cardfail.2010.08.012] [Medline: 21300305]
66. Kotooka N, Kitakaze M, Nagashima K, Asaka M, Kinugasa Y, Nochioka K, HOMES-HF study investigators. The first multicenter, randomized, controlled trial of home telemonitoring for Japanese patients with heart failure: home telemonitoring study for patients with heart failure (HOMES-HF). Heart Vessels 2018 Aug;33(8):866-876. [doi: 10.1007/s00380-018-1133-5] [Medline: 29450689]
67. Lee JY, Chan CK, Chua SS, Paraidathathu T, Lee KK, Tan CS, et al. Using telemedicine to support care for people with type 2 diabetes mellitus: a qualitative analysis of patients' perspectives. BMJ Open 2019 Oct 22;9(10):e026575 [FREE Full text] [doi: 10.1136/bmjopen-2018-026575] [Medline: 31640990]
68. Lee PA, Greenfield G, Pappas Y. Patients' perception of using telehealth for type 2 diabetes management: a phenomenological study. BMC Health Serv Res 2018 Jul 13;18(1):549 [FREE Full text] [doi: 10.1186/s12913-018-3353-x] [Medline: 30005696]
69. Leng Chow W, Aung CY, Tong SC, Goh GS, Lee S, MacDonald MR, et al. Effectiveness of telemonitoring-enhanced support over structured telephone support in reducing heart failure-related healthcare utilization in a multi-ethnic Asian setting. J Telemed Telecare 2020 Jul;26(6):332-340. [doi: 10.1177/1357633X18825164] [Medline: 30782070]
70. Lewis KE, Annandale JA, Warm DL, Hurlin C, Lewis MJ, Lewis L. Home telemonitoring and quality of life in stable, optimised chronic obstructive pulmonary disease. J Telemed Telecare 2010;16(5):253-259. [doi: 10.1258/jtt.2009.090907] [Medline: 20483881]
71. Madigan E, Schmotzer BJ, Struk CJ, DiCarlo CM, Kikano G, Piña IL, et al. Home health care with telemonitoring improves health status for older adults with heart failure. Home Health Care Serv Q 2013;32(1):57-74 [FREE Full text] [doi: 10.1080/01621424.2012.755144] [Medline: 23438509]
72. Madsen LB, Kirkegaard P, Pedersen EB. Blood pressure control during telemonitoring of home blood pressure. A randomized controlled trial during 6 months. Blood Press 2008;17(2):78-86. [doi: 10.1080/08037050801915468] [Medline: 18568696]
73. Margolis KL, Asche SE, Bergdall AR, Dehmer SP, Groen SE, Kadrmas HM, et al. Effect of home blood pressure telemonitoring and pharmacist management on blood pressure control: a cluster randomized clinical trial. JAMA 2013 Jul 03;310(1):46-56 [FREE Full text] [doi: 10.1001/jama.2013.6549] [Medline: 23821088]
74. Margolis KL, Asche SE, Dehmer SP, Bergdall AR, Green BB, Sperl-Hillen JM, et al. Long-term outcomes of the effects of home blood pressure telemonitoring and pharmacist management on blood pressure among adults with uncontrolled

XSL•FO
RenderX

hypertension: follow-up of a cluster randomized clinical trial. JAMA Netw Open 2018 Sep 07;1(5):e181617 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1617] [Medline: 30646139]
75. McKinstry B, Hanley J, Wild S, Pagliari C, Paterson M, Lewis S, et al. Telemonitoring based service redesign for the management of uncontrolled hypertension: multicentre randomised controlled trial. BMJ 2013 May 24;346:f3030 [FREE Full text] [doi: 10.1136/bmj.f3030] [Medline: 23709583]
76. Michaud TL, Siahpush M, Schwab RJ, Eiland LA, DeVany M, Hansen G, et al. Remote patient monitoring and clinical outcomes for postdischarge patients with type 2 diabetes. Popul Health Manag 2018 Oct;21(5):387-394. [doi: 10.1089/pop.2017.0175] [Medline: 29583057]
77. McManus RJ, Mant J, Bray EP, Holder R, Jones MI, Greenfield S, et al. Telemonitoring and self-management in the control of hypertension (TASMINH2): a randomised controlled trial. Lancet 2010 Jul 17;376(9736):163-172. [doi: 10.1016/S0140-6736(10)60964-6] [Medline: 20619448]
78. Mira-Solves JJ, Orozco-Beltrán D, Sánchez-Molla M, Sánchez García JJ, en nombre de los investigadores del programa ValCrònic. Evaluación de la satisfacción de los pacientes crónicos con los dispositivos de telemedicina y con el resultado de la atención recibida. Programa ValCrònic. Aten Primaria 2014 Jun;46 Suppl 3:16-23 [FREE Full text] [doi: 10.1016/S0212-6567(14)70061-7] [Medline: 25262307]
79. Moon EW, Tan NC, Allen JC, Jafar TH. The use of wireless, smartphone app-assisted home blood pressure monitoring among hypertensive patients in Singapore: pilot randomized controlled trial. JMIR Mhealth Uhealth 2019 May 28;7(5):e13153 [FREE Full text] [doi: 10.2196/13153] [Medline: 30905872]
80. Mortara A, Pinna GD, Johnson P, Maestri R, Capomolla S, La Rovere MT, HHH Investigators. Home telemonitoring in heart failure patients: the HHH study (Home or Hospital in Heart Failure). Eur J Heart Fail 2009 Mar;11(3):312-318 [FREE Full text] [doi: 10.1093/eurjhf/hfp022] [Medline: 19228800]
81. Neumann CL, Menne J, Rieken EM, Fischer N, Weber MH, Haller H, et al. Blood pressure telemonitoring is useful to achieve blood pressure control in inadequately treated patients with arterial hypertension. J Hum Hypertens 2011 Dec;25(12):732-738. [doi: 10.1038/jhh.2010.119] [Medline: 21228822]
82. Nissen L, Lindhardt T. A qualitative study of COPD-patients' experience of a telemedicine intervention. Int J Med Inform 2017 Nov;107:11-17. [doi: 10.1016/j.ijmedinf.2017.08.004] [Medline: 29029687]
83. Nouryan CN, Morahan S, Pecinka K, Akerman M, Lesser M, Chaikin D, et al. Home telemonitoring of community-dwelling heart failure patients after home care discharge. Telemed J E Health 2019 Jun;25(6):447-454. [doi: 10.1089/tmj.2018.0099] [Medline: 30036166]
84. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the better effectiveness after transition -- heart failure (BEAT-HF) randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
85. Orozco-Beltran D, Sánchez-Molla M, Sanchez JJ, Mira JJ, ValCrònic Research Group. Telemedicine in primary care for patients with chronic conditions: the ValCrònic quasi-experimental study. J Med Internet Res 2017 Dec 15;19(12):e400 [FREE Full text] [doi: 10.2196/jmir.7677] [Medline: 29246881]
86. Pekmezaris R, Williams MS, Pascarelli B, Finuf KD, Harris YT, Myers AK, et al. Adapting a home telemonitoring intervention for underserved Hispanic/Latino patients with type 2 diabetes: an acceptability and feasibility study. BMC Med Inform Decis Mak 2020 Dec 07;20(1):324 [FREE Full text] [doi: 10.1186/s12911-020-01346-0] [Medline: 33287815]
87. Pressman AR, Kinoshita L, Kirk S, Barbosa GM, Chou C, Minkoff J. A novel telemonitoring device for improving diabetes control: protocol and results from a randomized clinical trial. Telemed J E Health 2014 Feb;20(2):109-114. [doi: 10.1089/tmj.2013.0157] [Medline: 24404816]
88. Ralston JD, Cook AJ, Anderson ML, Catz SL, Fishman PA, Carlson J, et al. Home blood pressure monitoring, secure electronic messaging and medication intensification for improving hypertension control: a mediation analysis. Appl Clin Inform 2014 Mar 12;5(1):232-248 [FREE Full text] [doi: 10.4338/ACI-2013-10-RA-0079] [Medline: 24734136]
89. Rodríguez-Idígoras MI, Sepúlveda-Muñoz J, Sánchez-Garrido-Escudero R, Martínez-González JL, Escolar-Castelló JL, Paniagua-Gómez IM, et al. Telemedicine influence on the follow-up of type 2 diabetes patients. Diabetes Technol Ther 2009 Jul;11(7):431-437. [doi: 10.1089/dia.2008.0114] [Medline: 19580356]
90. Rogers MA, Small D, Buchan DA, Butch CA, Stewart CM, Krenzer BE, et al. Home monitoring service improves mean arterial pressure in patients with essential hypertension. A randomized, controlled trial. Ann Intern Med 2001 Jun 05;134(11):1024-1032. [doi: 10.7326/0003-4819-134-11-200106050-00008] [Medline: 11388815]
91. Schoenfeld MH, Compton SJ, Mead RH, Weiss DN, Sherfesee L, Englund J, et al. Remote monitoring of implantable cardioverter defibrillators: a prospective analysis. Pacing Clin Electrophysiol 2004 Jun;27(6 Pt 1):757-763. [doi: 10.1111/j.1540-8159.2004.00524.x] [Medline: 15189530]
92. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]

XSL•FO
RenderX

93. Shea S, Weinstock RS, Starren J, Teresi J, Palmas W, Field L, et al. A randomized trial comparing telemedicine case management with usual care in older, ethnically diverse, medically underserved patients with diabetes mellitus. J Am Med Inform Assoc 2006;13(1):40-51 [FREE Full text] [doi: 10.1197/jamia.M1917] [Medline: 16221935]
94. Shea S, Weinstock RS, Teresi JA, Palmas W, Starren J, Cimino JJ, IDEATel Consortium. A randomized trial comparing telemedicine case management with usual care in older, ethnically diverse, medically underserved patients with diabetes mellitus: 5 year results of the IDEATel study. J Am Med Inform Assoc 2009;16(4):446-456 [FREE Full text] [doi: 10.1197/jamia.M3157] [Medline: 19390093]
95. Sicotte C, Paré G, Morin S, Potvin J, Moreault MP. Effects of home telemonitoring to support improved care for chronic obstructive pulmonary diseases. Telemed J E Health 2011 Mar;17(2):95-103. [doi: 10.1089/tmj.2010.0142] [Medline: 21214399]
96. Soriano JB, García-Río F, Vázquez-Espinosa E, Conforto JI, Hernando-Sanz A, López-Yepes L, et al. A multicentre, randomized controlled trial of telehealth for the management of COPD. Respir Med 2018 Nov;144:74-81 [FREE Full text] [doi: 10.1016/j.rmed.2018.10.008] [Medline: 30366588]
97. Stuckey M, Fulkerson R, Read E, Russell-Minda E, Munoz C, Kleinstiver P, et al. Remote monitoring technologies for the prevention of metabolic syndrome: the Diabetes and Technology for Increased Activity (DaTA) study. J Diabetes Sci Technol 2011 Jul 01;5(4):936-944 [FREE Full text] [doi: 10.1177/193229681100500417] [Medline: 21880237]
98. Trudel M, Cafazzo JA, Hamill M, Igharas W, Tallevi K, Picton P, et al. A mobile phone based remote patient monitoring system for chronic disease management. Stud Health Technol Inform 2007;129(Pt 1):167-171. [Medline: 17911700]
99. Tsang MW, Mok M, Kam G, Jung M, Tang A, Chan U, et al. Improvement in diabetes control with a monitoring system based on a hand-held, touch-screen electronic diary. J Telemed Telecare 2001;7(1):47-50. [doi: 10.1258/1357633011936138] [Medline: 11265938]
100. Tupper OD, Gregersen TL, Ringbaek T, Brøndum E, Frausing E, Green A, et al. Effect of tele-health care on quality of life in patients with severe COPD: a randomized clinical trial. Int J Chron Obstruct Pulmon Dis 2018 Aug 29;13:2657-2662 [FREE Full text] [doi: 10.2147/COPD.S164121] [Medline: 30214183]
101. Valdivieso B, García-Sempere A, Sanfélix-Gimeno G, Faubel R, Librero J, Soriano E, GeChronic Group. The effect of telehealth, telephone support or usual care on quality of life, mortality and healthcare utilization in elderly high-risk patients with multiple chronic conditions. A prospective study. Med Clin (Barc) 2018 Oct 23;151(8):308-314. [doi: 10.1016/j.medcli.2018.03.013] [Medline: 29705155]
102. van Berkel C, Almond P, Hughes C, Smith M, Horsfield D, Duckworth H. Retrospective observational study of the impact on emergency admission of telehealth at scale delivered in community care in Liverpool, UK. BMJ Open 2019 Jul 31;9(7):e028981 [FREE Full text] [doi: 10.1136/bmjopen-2019-028981] [Medline: 31371293]
103. Varon C, Alao M, Minter J, Stapleton M, Thomson S, Jaecques S, et al. Telehealth on heart failure: results of the Recap project. J Telemed Telecare 2015 Sep;21(6):340-347. [doi: 10.1177/1357633X15577310] [Medline: 25962654]
104. Vianello A, Fusello M, Gubian L, Rinaldo C, Dario C, Concas A, et al. Home telemonitoring for patients with acute exacerbation of chronic obstructive pulmonary disease: a randomized controlled trial. BMC Pulm Med 2016 Nov 22;16(1):157 [FREE Full text] [doi: 10.1186/s12890-016-0321-2] [Medline: 27876029]
105. Villani A, Malfatto G, Compare A, Della Rosa F, Bellardita L, Branzi G, et al. Clinical and psychological telemonitoring and telecare of high risk heart failure patients. J Telemed Telecare 2014 Dec;20(8):468-475. [doi: 10.1177/1357633X14555644] [Medline: 25339632]
106. Vuorinen AL, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van Gils M, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014 Dec 11;16(12):e282 [FREE Full text] [doi: 10.2196/jmir.3651] [Medline: 25498992]
107. Wade MJ, Desai AS, Spettell CM, Snyder AD, McGowan-Stackewicz V, Kummer PJ, et al. Telemonitoring with case management for seniors with heart failure. Am J Manag Care 2011 Mar 01;17(3):e71-e79 [FREE Full text] [Medline: 21504262]
108. Walker PP, Pompilio PP, Zanaboni P, Bergmo TS, Prikk K, Malinovschi A, et al. Telemonitoring in chronic obstructive pulmonary disease (CHROMED). A randomized clinical trial. Am J Respir Crit Care Med 2018 Sep 01;198(5):620-628. [doi: 10.1164/rccm.201712-2404OC] [Medline: 29557669]
109. Wild SH, Hanley J, Lewis SC, McKnight JA, McCloughan LB, Padfield PL, et al. Supported telemonitoring and glycemic control in people with type 2 diabetes: the telescot diabetes pragmatic multicenter randomized controlled trial. PLoS Med 2016 Jul;13(7):e1002098 [FREE Full text] [doi: 10.1371/journal.pmed.1002098] [Medline: 27458809]
110. 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]
111. Abraham WT. Disease management: remote monitoring in heart failure patients with implantable defibrillators, resynchronization devices, and haemodynamic monitors. Europace 2013 Jun;15 Suppl 1:i40-i46. [doi: 10.1093/europace/eut105] [Medline: 23737229]

XSL•FO
RenderX

112. Abraham WT, Adamson PB, Bourge RC, Aaron MF, Costanzo MR, Stevenson LW, CHAMPION Trial Study Group. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. Lancet 2011 Feb 19;377(9766):658-666. [doi: 10.1016/S0140-6736(11)60101-3] [Medline: 21315441]
113. Abraham WT, Perl L. Implantable hemodynamic monitoring for heart failure patients. J Am Coll Cardiol 2017 Jul 18;70(3):389-398 [FREE Full text] [doi: 10.1016/j.jacc.2017.05.052] [Medline: 28705321]
114. Adamson PB, Abraham WT, Bourge RC, Costanzo MR, Hasan A, Yadav C, et al. Wireless pulmonary artery pressure monitoring guides management to reduce decompensation in heart failure with preserved ejection fraction. Circ Heart Fail 2014 Nov;7(6):935-944. [doi: 10.1161/CIRCHEARTFAILURE.113.001229] [Medline: 25286913]
115. Adamson PB, Abraham WT, Stevenson LW, Desai AS, Lindenfeld J, Bourge RC, et al. Pulmonary artery pressure-guided heart failure management reduces 30-day readmissions. Circ Heart Fail 2016 Jun;9(6):e002600. [doi: 10.1161/CIRCHEARTFAILURE.115.002600] [Medline: 27220593]
116. Ades PA, Pashkow FJ, Fletcher G, Pina IL, Zohman LR, Nestor JR. A controlled trial of cardiac rehabilitation in the home setting using electrocardiographic and voice transtelephonic monitoring. Am Heart J 2000 Mar;139(3):543-548. [doi: 10.1016/s0002-8703(00)90100-5] [Medline: 10689271]
117. Ahring KK, Ahring JP, Joyce C, Farid NR. Telephone modem access improves diabetes control in those with insulin-requiring diabetes. Diabetes Care 1992 Aug;15(8):971-975. [doi: 10.2337/diacare.15.8.971] [Medline: 1505329]
118. Ajay VS, Jindal D, Roy A, Venugopal V, Sharma R, Pawar A, et al. Development of a smartphone-enabled hypertension and diabetes mellitus management package to facilitate evidence-based care delivery in primary healthcare facilities in India: the mPower heart project. J Am Heart Assoc 2016 Dec 21;5(12):e004343 [FREE Full text] [doi: 10.1161/JAHA.116.004343] [Medline: 28003248]
119. Al-Khatib SM, Piccini JP, Knight D, Stewart M, Clapp-Channing N, Sanders GD. Remote monitoring of implantable cardioverter defibrillators versus quarterly device interrogations in clinic: results from a randomized pilot clinical trial. J Cardiovasc Electrophysiol 2010 May;21(5):545-550. [doi: 10.1111/j.1540-8167.2009.01659.x] [Medline: 20021522]
120. Amara W, Montagnier C, Cheggour S, Boursier M, Barnay C, Georger F, et al. Early detection and treatment of supraventricular arrhythmia by remote monitoring prevents its progression in pacemaker patients: the randomized, multicenter SETAM trial. Europace 2015 Mar;17(5):iii8. [doi: 10.1093/europace/euv151]
121. Ando K, Koyama J, Abe Y, Sato T, Shoda M, Soga Y, et al. Feasibility evaluation of a remote monitoring system for implantable cardiac devices in Japan. Int Heart J 2011;52(1):39-43 [FREE Full text] [doi: 10.1536/ihj.52.39] [Medline: 21321467]
122. Bastyr 3rd EJ, Zhang S, Mou J, Hackett AP, Raymond SA, Chang AM. Performance of an electronic diary system for intensive insulin management in global diabetes clinical trials. Diabetes Technol Ther 2015 Aug;17(8):571-579 [FREE Full text] [doi: 10.1089/dia.2014.0407] [Medline: 25826466]
123. Batalik L, Dosbaba F, Hartman M, Batalikova K, Spinar J. Benefits and effectiveness of using a wrist heart rate monitor as a telerehabilitation device in cardiac patients: a randomized controlled trial. Medicine (Baltimore) 2020 Mar;99(11):e19556 [FREE Full text] [doi: 10.1097/MD.0000000000019556] [Medline: 32176113]
124. Bekelman DB, Plomondon ME, Carey EP, Sullivan MD, Nelson KM, Hattler B, et al. Primary results of the patient-centered disease management (PCDM) for heart failure study: a randomized clinical trial. JAMA Intern Med 2015 May;175(5):725-732. [doi: 10.1001/jamainternmed.2015.0315] [Medline: 25822284]
125. Benhamou PY, Melki V, Boizel R, Perreal F, Quesada JL, Bessieres-Lacombe S, et al. One-year efficacy and safety of Web-based follow-up using cellular phone in type 1 diabetic patients under insulin pump therapy: the PumpNet study. Diabetes Metab 2007 Jun;33(3):220-226. [doi: 10.1016/j.diabet.2007.01.002] [Medline: 17395516]
126. Bernocchi P, Vitacca M, La Rovere MT, Volterrani M, Galli T, Baratti D, et al. Home-based telerehabilitation in older patients with chronic obstructive pulmonary disease and heart failure: a randomised controlled trial. Age Ageing 2018 Jan 01;47(1):82-88. [doi: 10.1093/ageing/afx146] [Medline: 28985325]
127. Bhavnani SP, Sola S, Adams D, Venkateshvaran A, Dash PK, Sengupta PP, ASEF-VALUES Investigators. A randomized trial of pocket-echocardiography integrated mobile health device assessments in modern structural heart disease clinics. JACC Cardiovasc Imaging 2018 Apr;11(4):546-557 [FREE Full text] [doi: 10.1016/j.jcmg.2017.06.019] [Medline: 28917688]
128. Biermann E, Dietrich W, Standl E. Telecare of diabetic patients with intensified insulin therapy. A randomized clinical trial. Stud Health Technol Inform 2000;77:327-332. [Medline: 11187566]
129. Billiard A, Rohmer V, Roques MA, Joseph MG, Suraniti S, Giraud P, et al. Telematic transmission of computerized blood glucose profiles for IDDM patients. Diabetes Care 1991 Feb;14(2):130-134. [doi: 10.2337/diacare.14.2.130] [Medline: 2060415]
130. Böhm M, Drexler H, Oswald H, Rybak K, Bosch R, Butter C, OptiLink HF Study Investigators. Fluid status telemedicine alerts for heart failure: a randomized controlled trial. Eur Heart J 2016 Nov 01;37(41):3154-3163. [doi: 10.1093/eurheartj/ehw099] [Medline: 26984864]
131. Bosworth HB, Powers BJ, Olsen MK, McCant F, Grubber J, Smith V, et al. Home blood pressure management and improved blood pressure control: results from a randomized controlled trial. Arch Intern Med 2011 Jul 11;171(13):1173-1180. [doi: 10.1001/archinternmed.2011.276] [Medline: 21747013]

XSL•FO
RenderX

132. Boyne JJ, Vrijhoef HJ, Spreeuwenberg M, De Weerd G, Kragten J, Gorgels AP, TEHAF investigators. Effects of tailored telemonitoring on heart failure patients' knowledge, self-care, self-efficacy and adherence: a randomized controlled trial. Eur J Cardiovasc Nurs 2014 Jun;13(3):243-252. [doi: 10.1177/1474515113487464] [Medline: 23630403]
133. Bray EP, Jones MI, Banting M, Greenfield S, Hobbs FD, Little P, et al. Performance and persistence of a blood pressure self-management intervention: telemonitoring and self-management in hypertension (TASMINH2) trial. J Hum Hypertens 2015 Jul;29(7):436-441. [doi: 10.1038/jhh.2014.108] [Medline: 25566874]
134. Chan WM, Woo J, Hui E, Lau WW, Lai JC, Lee D. A community model for care of elderly people with diabetes via telemedicine. Appl Nurs Res 2005 May;18(2):77-81. [doi: 10.1016/j.apnr.2004.11.002] [Medline: 15991104]
135. Chase HP, Pearson JA, Wightman C, Roberts MD, Oderberg AD, Garg SK. Modem transmission of glucose values reduces the costs and need for clinic visits. Diabetes Care 2003 May;26(5):1475-1479. [doi: 10.2337/diacare.26.5.1475] [Medline: 12716807]
136. Dang S, Karanam C, Gómez-Marín O. Outcomes of a mobile phone intervention for heart failure in a minority county hospital population. Telemed J E Health 2017 Jun;23(6):473-484. [doi: 10.1089/tmj.2016.0211] [Medline: 28051357]
137. Desai AS. Home monitoring heart failure care does not improve patient outcomes: looking beyond telephone-based disease management. Circulation 2012 Feb 14;125(6):828-836. [doi: 10.1161/CIRCULATIONAHA.111.031179] [Medline: 22331920]
138. Ding H, Jayasena R, Maiorana A, Dowling A, Chen SH, Karunanithi M, et al. Innovative Telemonitoring Enhanced Care Programme for Chronic Heart Failure (ITEC-CHF) to improve guideline compliance and collaborative care: protocol of a multicentre randomised controlled trial. BMJ Open 2017 Oct 08;7(10):e017550 [FREE Full text] [doi: 10.1136/bmjopen-2017-017550] [Medline: 28993389]
139. Schwarz KA, Mion LC, Hudock D, Litman G. Telemonitoring of heart failure patients and their caregivers: a pilot randomized controlled trial. Prog Cardiovasc Nurs 2008;23(1):18-26. [doi: 10.1111/j.1751-7117.2008.06611.x] [Medline: 18326990]
140. Trief PM, Teresi JA, Eimicke JP, Shea S, Weinstock RS. Improvement in diabetes self-efficacy and glycaemic control using telemedicine in a sample of older, ethnically diverse individuals who have diabetes: the IDEATel project. Age Ageing 2009 Mar;38(2):219-225. [doi: 10.1093/ageing/afn299] [Medline: 19171951]
141. Zakeri R, Morgan JM, Phillips P, Kitt S, Ng GA, McComb JM, REM-HF Investigators. Impact of remote monitoring on clinical outcomes for patients with heart failure and atrial fibrillation: results from the REM-HF trial. Eur J Heart Fail 2020 Mar;22(3):543-553 [FREE Full text] [doi: 10.1002/ejhf.1709] [Medline: 31908129]
142. Ancker J. Comparative Effectiveness of Telemedicine in Primary Care. Clinical Trials. 2020 Dec 28. URL: https://ClinicalTrials.gov/show/NCT04684836; [accessed 2021-10-01]
143. Bessonov IS. Telemedicine Follow-up for Post-ACS Patients. Clinical Trials. 2020 Jul 24. URL: https://ClinicalTrials.gov/show/NCT04485754; [accessed 2021-10-01]
144. Colet JC. Heart Failure Events Reduction With Remote Monitoring and eHealth Support Investigator Initiated Trial (HERMeS). Clinical Trials. 2018 Sep 10. URL: https://ClinicalTrials.gov/show/NCT03663907; [accessed 2021-10-01]
145. Dominguez H. Brazilian Heart Insufficiency With Telemedicine (BRAHIT). Clinical Trials. 2020 Jul 10. URL: https://ClinicalTrials.gov/show/NCT04466852; [accessed 2021-10-01]
146. Esteban C. Impact of the Artificial Intelligence in a Telemonitoring Programme of COPD Patients With Multiple Hospitalizations. Clinical Trials. 2021 Jul 27. URL: https://ClinicalTrials.gov/show/NCT04978922; [accessed 2021-10-01]
147. Hoffman J. Influence of Telemonitoring on the Management of LVAD-patients. Clinical Trials. 2020 Nov 3. URL: https://ClinicalTrials.gov/show/NCT04613401; [accessed 2021-10-01]
148. Iversen MM. Telemedicine Follow-up in Primary Health Care for Diabetes-related Foot Ulcers (DiaFOTo). Clinical Trials. 2012 Oct 19. URL: https://ClinicalTrials.gov/show/NCT01710774; [accessed 2021-10-01]
149. Krzowski B. Mobile App and Digital System for Patients After Myocardial Infarction (afterAMI). Clinical Trials. 2021 Mar 11. URL: https://ClinicalTrials.gov/show/NCT04793425; [accessed 2021-10-01]
150. Mahler SA. Enhancing Rural Health Via Cardiovascular Telehealth for Rural Patients Implementation (E-VICTORS). Clinical Trials. 2020 Nov 5. URL: https://ClinicalTrials.gov/show/NCT04617834; [accessed 2021-10-01]
151. Marinello R. Telemonitoring of Patients Admitted in Hospital at Home With Acute Decompensated Heart Failure - Pilot Study (MONTEROSA). Clinical Trials. 2020 May 27. URL: https://ClinicalTrials.gov/show/NCT04403659; [accessed 2021-10-01]
152. Martin S. TeLIPro Health Program - Active With Diabetes (TeLIPro). Clinical Trials. 2018 Sep 18. URL: https://ClinicalTrials.gov/show/NCT03675919; [accessed 2021-10-01]
153. Naqvi IA. Telehealth After Stroke Care: Integrated Multidisciplinary Access to Post-stroke Care (TASC). Clinical Trials. 2020 Nov 23. URL: https://ClinicalTrials.gov/show/NCT04640519; [accessed 2021-10-01]
154. Noruzbaeva A. Effects of Remote Monitoring of Patients With Heart Failure Based on Smartphone Application (ERICA-HF). Clinical Trials. 2020 Oct 19. URL: https://ClinicalTrials.gov/show/NCT04591964; [accessed 2021-10-01]
155. Nouira S. The Impact of Telemonitoring in the Management of Hypertension (HOROSCOPE). Clinical Trials. 2020 Oct 29. URL: https://ClinicalTrials.gov/show/NCT04607239; [accessed 2021-10-01]
156. Omboni S. Telemonitoring of Blood Pressure in Local Pharmacies (TEMPLAR). Clinical Trials. 2018 Dec 19. URL: https://ClinicalTrials.gov/show/NCT03781401; [accessed 2021-10-01]

XSL•FO
RenderX

157. Pekmezaris R. Diabetes Management Program for Hispanic/Latino. Clinical Trials. 2019 May 23. URL: https://ClinicalTrials.gov/show/NCT03960424; [accessed 2021-10-01]
158. Prigent A. Study Evaluating Telemonitoring and Experimentation in Telemedicine for the Improvement of Healthcare Pathways (ETAPES Program) Compared to Standard of Care in Patients With Chronic Respiratory Failure Receiving Non-invasive Home Ventilation (e-VENT). Clinical Trials. 2020 Nov 4. URL: https://ClinicalTrials.gov/show/NCT04615078; [accessed 2021-10-01]
159. Schwartz J, Ssinabulya I. mHealth for Self-care of Heart Failure in Uganda. Clinical Trials. 2020 Jun 11. URL: https://ClinicalTrials.gov/show/NCT04426630; [accessed 2021-10-01]
160. Siriwardena AN. Use of MonitorMe in COPD. Clinical Trials. 2019 Sep 27. URL: https://ClinicalTrials.gov/show/NCT04108143; [accessed 2021-10-01]
161. Skanes AC. Virtual for Care Atrial Fibrillation Patients Using VIRTUES. Clinical Trials. 2020 Oct 22. URL: https://ClinicalTrials.gov/show/NCT04599114; [accessed 2021-10-01]
162. Sun L, Zhu Y, Sun X, Jiang J. Integrative Management of Patients With Atrial Fibrillation Via Hospital-Community-Family-Based Telemedicine (HCFT-AF) Program. Clinical Trials. 2019 Oct 16. URL: https://ClinicalTrials.gov/show/NCT04127799; [accessed 2021-10-01]
163. Yin W. COVID-19 Lockdown Related Telemedicine for Type 2 Diabetes. Clinical Trials. 2021 Jan 25. URL: https://ClinicalTrials.gov/show/NCT04723550; [accessed 2021-10-01]
164. 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. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
165. 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]
166. Review Manager (RevMan). 5.3. Copenhagen, Denmark: The Cochrane Collaboration; 2014. URL: https://training.cochrane.org/online-learning/core-software/revman/revman-5-download [accessed 2021-10-01]
167. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.3. London, UK: The Cochrane Collaboration; 2022.
168. Shahaj O, Denneny D, Schwappach A, Pearce G, Epiphaniou E, Parke H, et al. Supporting self-management for people with hypertension: a meta-review of quantitative and qualitative systematic reviews. J Hypertens 2019 Feb;37(2):264-279. [doi: 10.1097/HJH.0000000000001867] [Medline: 30020240]
169. 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]
170. Snoswell CL, Stringer H, Taylor ML, Caffery LJ, Smith AC. An overview of the effect of telehealth on mortality: a systematic review of meta-analyses. J Telemed Telecare (forthcoming) 2021 Jun 29:1357633X211023700. [doi: 10.1177/1357633X211023700] [Medline: 34184578]
171. Kalankesh LR, Pourasghar F, Nicholson L, Ahmadi S, Hosseini M. Effect of telehealth interventions on hospitalization indicators: a systematic review. Perspect Health Inf Manag 2016 Oct 1;13(Fall):1h [FREE Full text] [Medline: 27843425]
172. Zhu Y, Gu X, Xu C. Effectiveness of telemedicine systems for adults with heart failure: a meta-analysis of randomized controlled trials. Heart Fail Rev 2020 Mar;25(2):231-243 [FREE Full text] [doi: 10.1007/s10741-019-09801-5] [Medline: 31197564]
173. Thurmond VA, Boyle DK. An integrative review of patients' perceptions regarding telehealth used in their health care. Online J Knowl Synth Nurs 2002 Apr 16;9:2. [Medline: 12089636]
174. 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]
175. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. J Med Internet Res 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: 10.2196/20786] [Medline: 32810841]
176. Nguyen M, Waller M, Pandya A, Portnoy J. A review of patient and provider satisfaction with telemedicine. Curr Allergy Asthma Rep 2020 Sep 22;20(11):72 [FREE Full text] [doi: 10.1007/s11882-020-00969-7] [Medline: 32959158]
177. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
178. McFarland S, Coufopolous A, Lycett D. The effect of telehealth versus usual care for home-care patients with long-term conditions: a systematic review, meta-analysis and qualitative synthesis. J Telemed Telecare 2021 Feb;27(2):69-87. [doi: 10.1177/1357633X19862956] [Medline: 31394973]

Abbreviations

BP: blood pressure

XSL•FO
RenderX

COPD: chronic obstructive pulmonary disease
DBP: diastolic blood pressure
HbA_{1c}: glycated hemoglobin
MD: mean difference
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
RR: risk ratio
SBP: systolic blood pressure

Edited by T Leung; submitted 09.12.21; peer-reviewed by C Guinemer, B Smith, K Matthias; comments to author 18.02.22; revised version received 07.04.22; accepted 29.04.22; published 03.11.22.

Please cite as:
Leo DG, Buckley BJR, Chowdhury M, Harrison SL, Isanejad M, Lip GYH, Wright DJ, Lane DA, the TAILOR investigators
Interactive Remote Patient Monitoring Devices for Managing Chronic Health Conditions: Systematic Review and Meta-analysis
J Med Internet Res 2022;24(11):e35508
URL: https://www.jmir.org/2022/11/e35508
doi:10.2196/35508
PMID:36326818

©Donato Giuseppe Leo, Benjamin J R Buckley, Mahin Chowdhury, Stephanie L Harrison, Masoud Isanejad, Gregory Y H Lip, David J Wright, Deirdre A Lane, the TAILOR investigators. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

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:

This is a corrected version. See correction statement: https://www.jmir.org/2022/12/e45081

Abstract

Background: Ambient assisted living (AAL) is a common name for various artificial intelligence (AI)—infused applications and platforms that support their users in need in multiple activities, from health to daily living. These systems use different approaches to learn about their users and make automated decisions, known as AI models, for personalizing their services and increasing outcomes. Given the numerous systems developed and deployed for people with different needs, health conditions, and dispositions toward the technology, it is critical to obtain clear and comprehensive insights concerning AI models used, along with their domains, technology, and concerns, to identify promising directions for future work.

Objective: This study aimed to provide a scoping review of the literature on AI models in AAL. In particular, we analyzed specific AI models used in A A L systems, the target domains of the models, the technology using the models, and the major concerns from the end-user perspective. Our goal was to consolidate research on this topic and inform end users, health care professionals and providers, researchers, and practitioners in developing, deploying, and evaluating future intelligent AAL systems.

Methods: This study was conducted as a scoping review to identify, analyze, and extract the relevant literature. It used a natural language processing toolkit to retrieve the article corpus for an efficient and comprehensive automated literature search. Relevant articles were then extracted from the corpus and analyzed manually. This review included 5 digital libraries: IEEE, PubMed, Springer, Elsevier, and MDPI.

Results: We included a total of 108 articles. The annual distribution of relevant articles showed a growing trend for all categories from January 2010 to July 2022. The AI models mainly used unsupervised and semisupervised approaches. The leading models are deep learning, natural language processing, instance-based learning, and clustering. Activity assistance and recognition were the most common target domains of the models. Ambient sensing, mobile technology, and robotic devices mainly implemented the models. Older adults were the primary beneficiaries, followed by patients and frail persons of various ages. Availability was a top beneficiary concern.

Conclusions: This study presents the analytical evidence of AI models in AAL and their domains, technologies, beneficiaries, and concerns. Future research on intelligent AAL should involve health care professionals and caregivers as designers and users,

XSL•FO
RenderX

comply with health-related regulations, improve transparency and privacy, integrate with health care technological infrastructure, explain their decisions to the users, and establish evaluation metrics and design guidelines.

Trial Registration: PROSPERO (International Prospective Register of Systematic Reviews) CRD42022347590; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022347590

(J Med Internet Res 2022;24(11):e36553) doi:10.2196/36553

KEYWORDS

ambient assisted living; AAL; assisted living; active living; digital health; digital well-being; automated learning approach; artificial intelligence algorithms; human-centered AI; review; implications; artificial intelligence; mobile phone

Introduction

Background

Ambient assisted living (AAL) is an umbrella term describing a general approach to technology design to construct safe environments around assisted users and help them maintain independent living [1]. Over time, it has focused mainly on older adults (people aged >65 years) as target users.

Developing technology for this group is an increasingly important design challenge because of specific deficits in later life [2]. Beyond usability, there is an increased emphasis on designing technology for older adults that will enable them not merely to satisfy their needs but also to transform their mental and physical health and well-being [3,4]. This challenge is particularly significant because older people are becoming the largest demographic group. In 2020, more than one-fifth (20.6%) of the population in the European Union was aged ≥65 years [5], and by 2030, an estimated 16.6% of the world's population will be aged ≥60 years [6].

Over the last decade, many technological devices have been developed to support an active lifestyle as people age, concerning health promotion [4,7,8]. Health promotion refers to "the process of empowering people to increase control over their health and its determinants through health literacy efforts and multisectoral action to increase healthy behaviors" [9]. Concerning technology design, the objective is to find cost-effective solutions to help independent living and provide health care and well-being [10]. A comprehensive analysis of information and communication technology research and development revealed the common goals of these technologies as the provision of health, accessibility, and safety [11,12]. Many technologies seek to assist older adults with everyday activities [4,11,13-15].

To support disability-free and independent living and the well-being of older users, AAL systems use automated decision-making mechanisms that integrate, analyze, and interpret complex multimodal and multidevice information [7]. These systems have focused on 2 general scenarios of automated decision-making involving older users—*health monitoring* and *activity recognition* [8,16,17].

Different *monitoring* contexts were targeted by a variety of technological systems, ranging from monitoring systems for fall prevention using wearable and ambient sensing technology [18] and social robots for the well-being of people with dementia and mild cognitive impairments [13] to games for leisure and user engagement during therapy and rehabilitation [14].

Robotic technologies have been widely exploited as tools to support health monitoring and mobility capacities, such as strength, balance, and range of motion [15], or as companions [19] to assist older adults in daily and social activities at home. The former may be nonsocial robots, whereas the latter are social robots with the primary goal of offering companionship.

Remote telepresence robots have been successfully used to support the autonomy of older adults in doing daily activities at home. Giraff is a telepresence robot that uses a video interface to allow caregivers and relatives to visit older people in their homes remotely [20]. It runs implicit data collection (blood pressure, body temperature, movement, and fall) and then analyzes the data to alert the caregivers for emergencies. Similarly, Matilda is a social robot with human attributes (such as baby-face–like appearance, human voices, gestures, and body movements) that can recognize voices and faces and perform activities such as playing music, dancing, and playing card games [21].

Although biological aging cannot be stopped, regular exercise can minimize its physiological effects, increase life satisfaction, and prolong the decline in functional abilities in older adults [22]. Studies on the favorite activities of older adults show the prominence of physical activities such as walking, jogging, and outdoor maintenance [23]. Specific technologies, such as exergames [10] or web-based exercises and activities [24,25], have motivated, sustained, and monitored physical and social activities at older adults' homes. Coupling with the features from theories of human behavior, such as goal setting, self-monitoring, achievements, and personalized feedback and progression, has been associated with the higher effectiveness of these applications for older adults (ie, increased engagement in physical activities and associated health outcomes) [26].

In the context of AAL, *activity recognition* concerns tracking the daily behavior of older and frail people. It can detect falls and recognize activities of daily living (ADL), which are crucial for identifying complex patterns associated with the development of specific diseases. Zdravevski et al [27] suggested an automated approach for analyzing multivariate time series originating from various sensors and facilitating the robust classification of daily activities.

Wearable [28] and *mobile* technologies [29,30] have been used for implicit data collection and analysis to recognize older

XSL•FO
RenderX

adults' activities for tracking their health and detecting emergencies.

From a technical perspective, the energy efficiency of wearable technologies appears to be the primary constraint for continuous measurement and activity recognition [28]. It further affects the provision of timely and informative feedback and recommendations for the users. The major user-related concerns are privacy and acceptance [28] due to unclear use cases and difficulties in device pairing with a smartphone for older adults. A more stable commitment to wearables requires use cases with apparent benefits and reduced effort of use for older adults.

Mobile technologies represent a versatile source for older adults' health and activity data collection [29,30]. They facilitate home care and self-management of the health and well-being of older adults. These applications implement various services based on target activity and health recognition features to support health care and independent living (ie, reminders, companionship, or recommendation of favorite activities or treatments). However, significant obstacles to using mobile technologies in practice include privacy [30] and technological literacy and usability of touch screen interaction styles [31].

AAL technologies use a variety of artificial intelligence (AI) models in learning about their users' habits and health conditions to provide adequate services with automated decision-making. Table 1 shows the common AI classification, whereas Table 2 summarizes existing AI models concerning their learning and decision-making techniques and the problems they address (with corresponding algorithms) [32-34]. We separated the classification and models because multiple models can belong to the same class. Conversely, some models can implement different classes (ie, clustering can be done in both supervised and unsupervised manners).

Table 1. The artificial intelligence classification as common learning approaches [32-34].

Name	Description	Problem or algorithm
Supervised learning	Input (training) data or examples are labeled with known output values. The model uses the data in a training process to make predictions, and the predictions are corrected when they are false. The process runs until the model achieves a required level of the prediction accuracy.	Classification and regression
Unsupervised learning	Input data are not labeled, and output values are unknown. Instead, the model is trained by removing structures from the input data to extract general rules, reduce redundancy, or organize data by similarity.	Clustering, dimensionality reduction, and association rule learning
Semisupervised learning	Input data contains labeled and unlabeled examples. The model learns the structures to organize the data to create predictions. It models the unlabeled data.	Classification and regression
Reinforcement learning	The model rewards desired behaviors and eliminates undesired ones. It is represented by a learning agent (process) that perceives and interprets its environment, takes actions, and learns through trial and error.	Markov Decision Process, Q learning, and Monte Carlo methods

XSL•FO
RenderX

Table 2. The summary of artificial intelligence models [32-34].

Model	Learning technique	Problem or algorithm
Regression learning	Models a relationship between input and output data (or variables). The relation is iteratively refined by measuring errors in the model's predictions.	Variations such as linear and logistic regression
Instance-based learning	Models a decision based on instances of input data that are considered relevant or necessary. Creates a database of reference examples used to compare with new data to find optimal matches using similarity metrics to make a decision.	K-nearest neighbor and support vector machines
Regularization learning	The extension or modification of another model (eg, regression learning) in a way that reduces the complexity of the model by converting it into a simpler form.	Ridge regression and elastic net regression
Decision tree learning	Models a decision based on the values of the input data attributes. It follows a tree structure in making a decision for given input data.	Classification and regression trees and conditional decision tree
Bayesian learning	The models use Bayes' theorem to solve problems of classification and regression.	Naïve Bayes and Gaussian naïve Bayes
Clustering learning	The model organizes the input data into groups (or clusters) where group membership or commonality criteria are taken or derived from the data (eg, centroid based or hierarchical).	K-means, K-medians, and hierarchical clustering
Association rule learning	The model discovers associations in input data to make a decision. It extracts rules that describe relationships between observed variables in input data.	A priori algorithm and Eclat algorithm
Artificial neural network	The model is driven by the structure and function of the human neural networks. Represents a class of pattern matching models and their commonly used variations for regression and classification problems.	Perceptron, multilayer perceptrons, and back propagation
Deep learning	Special category of large and complex neural networks for handling vast amounts of labeled input data, including text, images, audio, and video.	Convolutional neural network, recurrent neural networks, and long short-term memory networks
Dimensionality reduction learning	The model analyzes the input structure in the data to represent and describe the data with less information. The simplified data can be visualized and used by other learning methods.	Principal component analysis, principal component regression, and linear discriminant analysis
Ensemble learning	Multiple models that are independently trained, where individual predictions are combined to make the final prediction. The models are combined owing to their weaknesses in making the desired prediction.	Boosting, random forest, AdaBoost, and weighted average (blending)
Natural language processing	Specific for conversational artificial intelligence and includes natural language understanding, dialog management, and natural language generation.	Rule-based algorithms, statistics, neural networks, and deep learning

Goal of the Study

This study investigated the AI models of existing AAL technologies to support independent living. The quality of the models' decision-making can benefit positive behavior change to maintain an active and healthy lifestyle for older adults and other user groups in need of assistance. This is critical for preventing functional decline and supporting health treatments. Our work aimed to identify the positive aspects and gaps in research and practice to provide implications for future AAL systems.

This scoped analysis focused on the following research questions (RQs):

1. RQ1: What AI models are implemented in AAL systems?
   First, we identified, described, and systematized AI classifications and models in the current landscape of AAL systems. For this purpose, we extracted common terminology to describe current AI models and AAL.
2. RQ2: What are the domains of the models?
   Second, we described existing target domains with their concrete activities to propose suitable application strategies that reinforce positive aspects and highlight critical parts in which further research is necessary.
3. RQ3: What technologies are using the models?
   Third, we investigated different technologies using AI models to consolidate and provide design and development guidelines for intelligent AAL systems.
4. RQ4: What are the significant concerns regarding the models from an end-user perspective?
   Finally, we examined end-user groups and their perceptions of AAL system use to indicate specific requirements that the systems should meet or improve.

This study reviews AI models in AAL concerning their domains, technologies, and concerns published in the literature covering 2010 to 2022. The findings are intended for health and care professionals, researchers, technology providers, and end users to consult when developing, deploying, and evaluating intelligent AAL technologies.

XSL•FO
RenderX

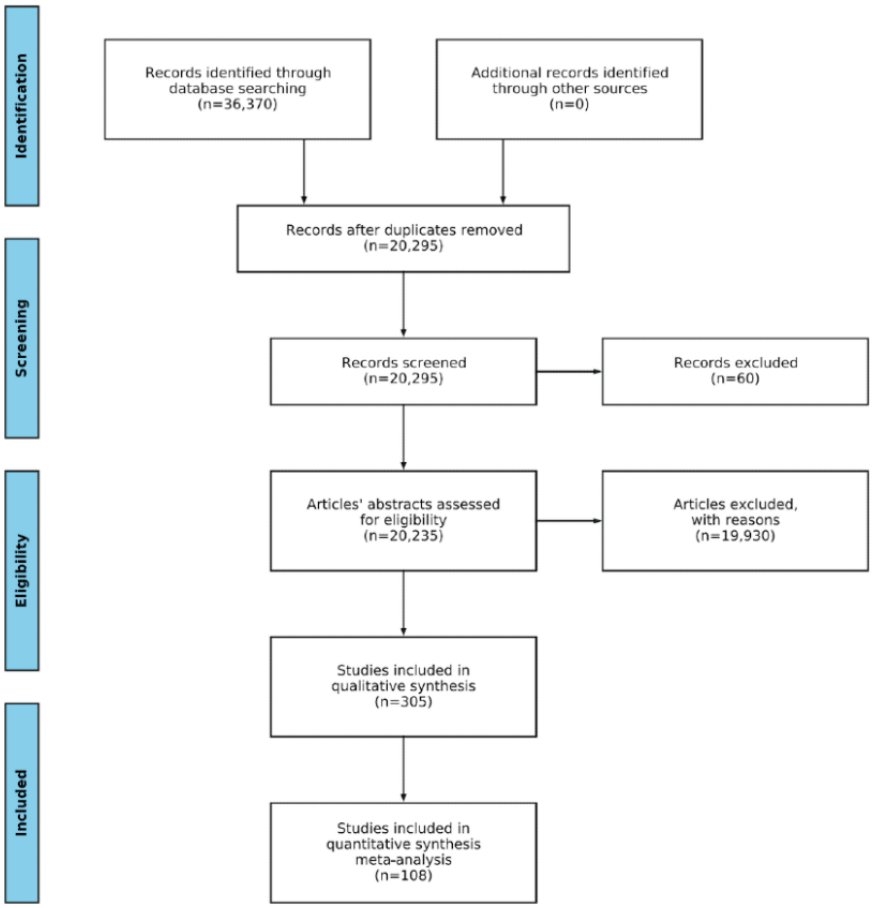
The paper continues as follows: the Methods section includes the methodology of the scoped literature review; the Results section describes the results of the analysis of the 108 selected articles; and the Discussion section contains the discussion of the review's findings concerning the RQs and outlines conclusions, limitations, and implications for future work.

Methods

Study Type

This paper has been organized as a scoping review, involving the synthesis and analysis of the existing literature to provide a conceptual framework that systematizes and clarifies the specific phenomena—AI models in AAL systems. We identified the articles to be reviewed by conducting a systematic literature search within the IEEE, PubMed, Springer, Elsevier, and MDPI research article databases. The study implemented the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) workflow for systematic reviews [35], as illustrated in Figure 1.

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow of the review process illustrates identification, screening, eligibility, and inclusion of relevant articles.



Identification

During the search, the titles, abstracts, and keywords of the library articles were queried with the search terms structured as in Table 3. The search terms of AAL and AI classification and model categories were mandatory for all articles, and the remaining categories were optional. We ran the search based on categories as properties and used keywords. For this purpose, we used a natural language processing (NLP) toolkit we had developed for automated literature search, screening, and analysis [36]. The toolkit accepts a collection of keywords as input to retrieve potentially relevant articles, combined with the set of properties (or categories) and property groups (as subcategories) to be satisfied by the articles. The input can be expanded with keyword and property synonyms to fine-tune the search and screening process. The details of the toolkit can be found in the article by Zdravevski et al [36].

The search was conducted in July 2022 and included research articles written in English and published between 2010 and 2022. Given the rapid advancements in AI that also influenced the significant growth of technology-supported AAL, we wanted to cover a sufficient research landscape concerning the time frame.

The search process sometimes identified the same article by multiple keywords and phrases from Table 3. For example, the article could describe the use of multiple AI models or classifications. In these situations, we counted the articles multiple times, per each found keyword, and have presented it in Figures 2-13 in the Results section.

XSL•FO
RenderX

Table 3. Key terminology for the scoping review's natural language processing search toolkit.

Category	Criteria	Keywords
Ambient assisted living	Mandatory	*Ambient assisted living, ambient-assisted living, assisted living, active and assisted living,* and *active-assisted living*
Artificial intelligence class	Mandatory	*Supervised learning, unsupervised learning, semi-supervised learning,* and *reinforcement learning*
Artificial intelligence model	Mandatory	*Classification, regression, clustering, dimensionality reduction, association rule learning, instance-based learning, regularization learning, decision tree learning, Bayesian learning, ANN, DL, ensemble learning,* and *natural language processing*
Domain	Optional	*Activity recognition, health monitoring, activity assistance, rehabilitation, therapy, interaction, communication,* and *entertainment*
Technology	Optional	*Mobile technology, mobile device, smartphone, tablet, touch-screen, wearable technology, wearable device, robot, robotic device, ambient sensing, ambient sensors, game, gamification, conversational agent, chatbot, virtual assistant,* and *virtual companion*
Beneficiaries	Optional	*Older adults, frail persons, patients, healthcare staff, caregivers,* and *family*
Concerns	Optional	*Acceptance, adoption, availability, accessibility, privacy, usability, reliability, safety,* and *security*

Screening

In the screening phase, we evaluated the retrieved articles to assess their relevance to the review based on the independent inclusion and exclusion criteria presented in Textbox 1.

The first 3 authors (MJ, GM, and EZ) manually screened the content of each article independently and coded it to indicate its relevance concerning the inclusion criteria. The inclusions were cross-checked, resolved, and confirmed during regular discussions among the authors.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Artificial intelligence (AI) classes and models of ambient assisted living (AAL) applications and platforms, where specific classes and models are explicitly considered and not mentioned without description, analysis, or evaluation.
- Articles contributing to the AI models' domains to support or assist in specific health-related or daily activities, in line with research question 2.
- Articles demonstrating different AAL technologies that use the models and deliver the automated decisions of AAL systems to end users, as per research question 3.
- Articles describing end users' concerns regarding the model's automated decision-making outcomes, according to research question 4.
- The primary end users are older adults, but end users also include other user groups.

Exclusion criteria

- Articles containing the search terms, but AAL, AI classes and models, domains, technology, and end users' concerns were not scrutinized. Thus, they were not relevant to research questions 1-4.
- Literature reviews and surveys on the related topics.

Extraction

In this phase, we analyzed each included article in detail. We identified and extracted AI classes and models in AAL systems, the models' target domains and technologies, and the end user's categories and concerns, where available per article. The extracted information from the articles was kept in a shared spreadsheet to facilitate coding and discussion among the authors. The extracted information included publication venue and date, a summary of the article, AI model or models used including corresponding AI algorithms and tools, the models' target domains if available, the technology using the models, if any, and information on the end users and their concerns regarding the models if available.

Analysis

We conducted a manual, thematic analysis of the extracted information during this phase. Our goal was to categorize the AI classes, models, domains, technologies, and concerns for AAL systems. Coded data were the basis to address the review's RQs. In particular, we grouped articles based on their primary outcomes to guide the analysis as follows: articles that describe AI classes and models of the AAL systems, articles dealing with the models' domains, articles that present the technologies using the models, and articles with the models' beneficiaries and use concerns.

We describe the general approach to analyzing the particular article groups.

The analysis of AI classes and models in AAL systems concerned identifying and describing the systems' automated learning and decision-making functionalities, including the particular AI algorithm or tool. *Analysis of AI models' domains* considered specific application scenarios with supported activities. *Analysis of the AI models' technologies* through which

XSL•FO
RenderX

the automated decisions were generated and communicated to the end users. *Analysis of the models' concerns* included various end users' perceptions and dispositions toward the models' functions and outcomes.

Results

Screening Process and Number of Articles

The NLP search toolkit initially identified 36,370 potentially relevant studies (Figure 1). Duplicates were then eliminated, reducing the number to 20,295. The automated screening process further removed 60 articles published before 2010 or for which the title or abstract could not be analyzed owing to parsing errors, unavailability, or other reasons. The NLP toolkit's advanced functions assessed the eligibility of the remaining 20,235 articles and kept 305 articles. After automated processing, the articles were analyzed in detail according to the inclusion and exclusion criteria. Finally, 108 articles were deemed eligible for the in-depth manual investigation to identify and articulate research results, trends, and implications. The articles are reported in Multimedia Appendix 1 [27,37-142].

We describe the results by responding to the RQs that guided our review.

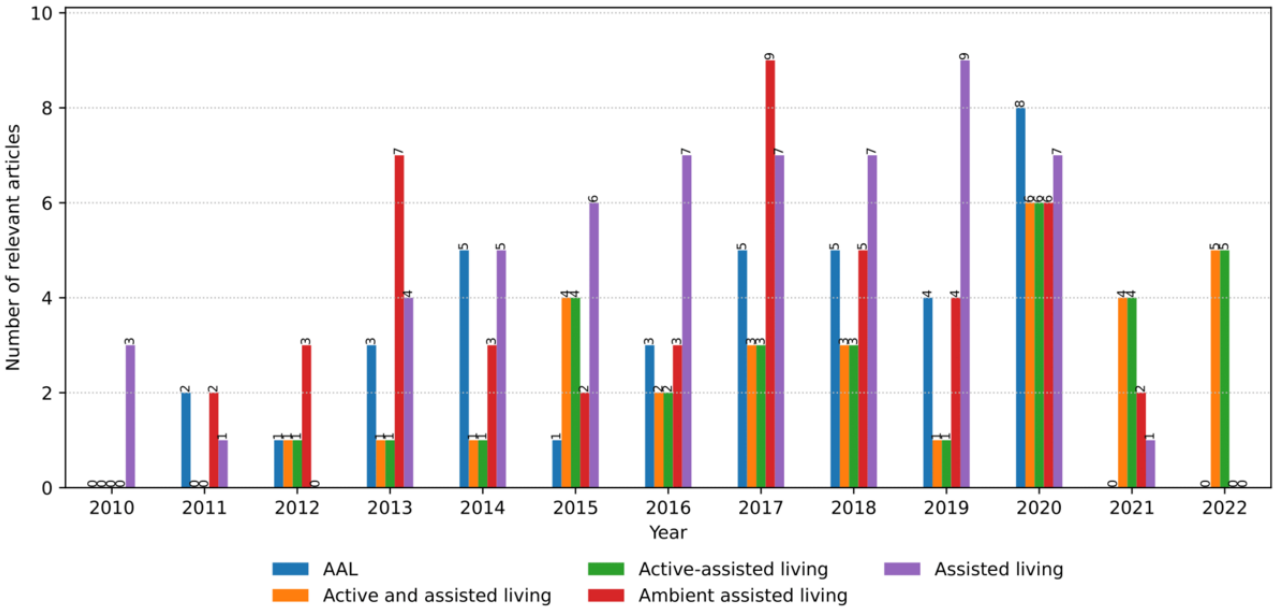
Distribution of Relevant Articles and Categories

Figure 2 illustrates the annual occurrences of the relevant articles containing different AI classes and models. The term "assisted living" has been commonly used in the literature to describe the systems with similar context and purpose of use as per the definition of AAL [1,4]. It outperformed the number of articles in some years (eg, 2019) and was comparable with the AAL in 2018 and 2020. In minor cases, the abbreviation was used solely. In general, there has been a growing trend throughout the search time frame, occasionally decreasing in specific years. The decreases were owing to our search conditions and inclusion criteria. Many articles dealt with AAL without explicit mentions of AI models concerning their application and outcomes.

The combined information on the digital library and publication year of the relevant articles demonstrates that IEEE is a leader, with an increasing trend, reaching a peak in 2020 (Figure 3). This is expected, as the publisher is oriented toward technology, with many venues relevant to AI models and AAL. PubMed follows, dealing more with the end-user aspects of the topics, such as different types of user evaluations. We could notice a growing trend until 2020 and an oscillatory period afterward. The Springer library combines technical articles with user-oriented articles. A smaller number of the relevant articles with an irregular annual trend was found in the Elsevier library, while MDPI published relevant articles from 2020.

As the total number of relevant articles increased within the review time frame, the number of articles pertinent to the associated categories changed accordingly (Figure 4). As for the 3 mandatory categories (*AAL*, *AI classification*, and *AI model*), there is a general growing trend up to 2020, with occasional drops in the previous year and a decrease in 2021. As an optional category, the *domain* follows the leading trend but with fewer articles, indicating that sometimes it was not considered (ie, AI models used or tested in a domain-independent way). The *beneficiaries* follow the leading trend but are smaller than the domain, showing that the AI models are sometimes studied without relating to a particular user group or groups. The *beneficiaries* are comparable with the technology in total amount but with annual oscillations due to different types of AI model verifications across relevant articles (ie, deployments and evaluations with or without users). The *concerns* appear in the smallest amount that grows in time and oscillates in some years, showing that relevant articles focused on various aspects of AI models in AAL, beyond and different from users' concerns (ie, algorithmic accuracy and performance).

Figure 2. The number of relevant articles concerning ambient assisted living (AAL) with artificial intelligence classes and models per year from January 2010 to July 2022.



XSL•FO
RenderX

Figure 3. The number of relevant articles per year from January 2010 to July 2022, grouped by the respective digital library.

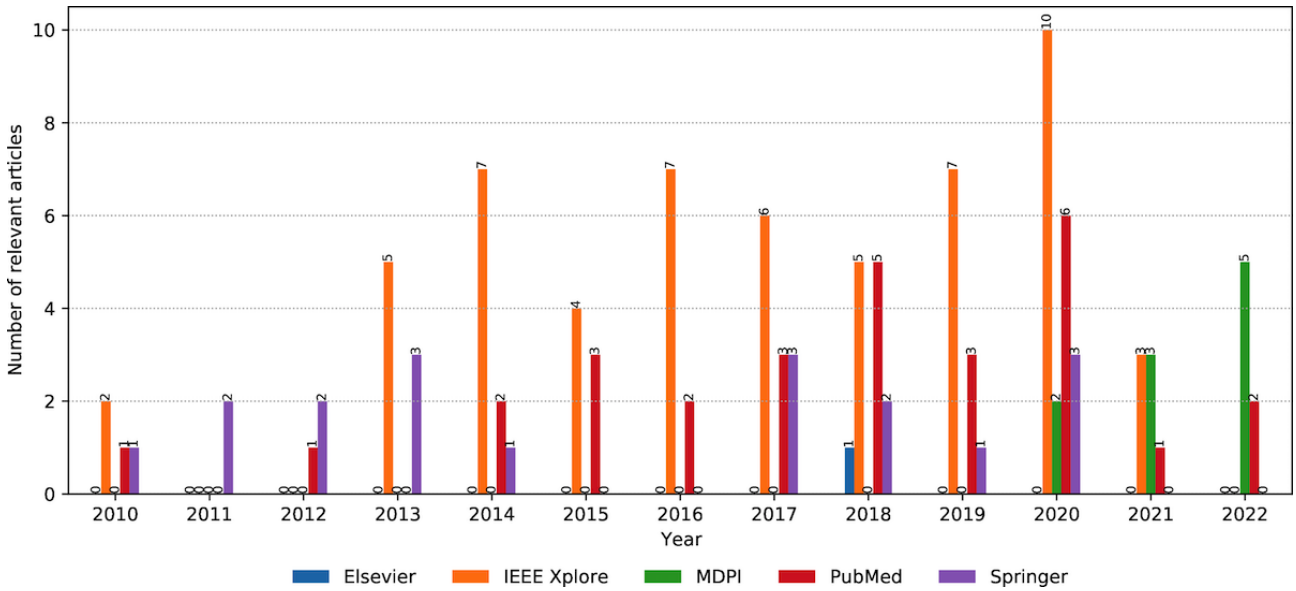
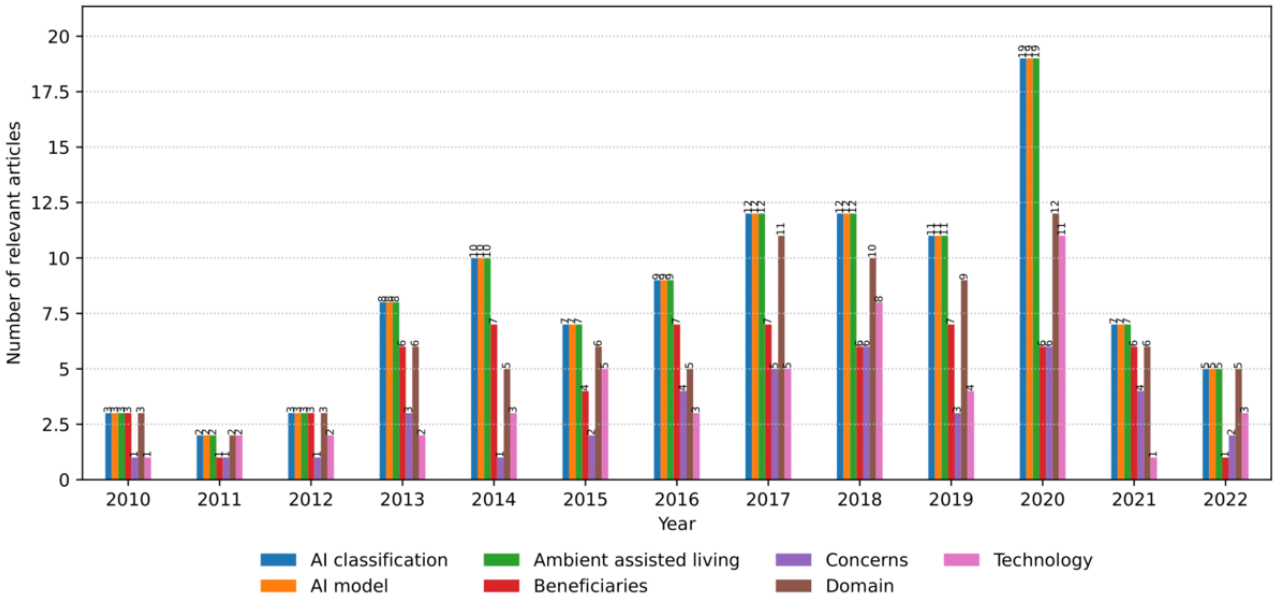


Figure 4. The number of relevant articles for each category per year from January 2010 to July 2022. AI: artificial intelligence.



Connections Between and Within Categories

Our analysis revealed overlap between searched categories. We aimed to represent all categories equally while highlighting particular connections as informative (eg, notably higher co-occurrences of instances from distinct or within the categories).

Figure 5 shows associations of AI models and the classes they use. Semisupervised learning is a dominant approach for deep learning (DL) and NLP models (51 occurrences). Unsupervised learning appears mainly in clustering (14 occurrences), instance-based learning (12 occurrences), and DL (11 occurrences). Supervised learning prevails for instance-based learning and DL (9 occurrences per model). Finally, reinforcement learning was the occasional approach for DL and NLP (7 occurrences per model).

The study reveals specific synergies within categories. Regarding the classes, 20 articles combined supervised and unsupervised learning. Reinforcement learning was used together with the previous 17 times per class. The studies combined the classes in a sequence or for mutual comparison in solving concrete problems. Concerning the models, we noticed that NLP tasks have been mainly tackled with DL algorithms and tools (51 occurrences).

Figure 6 presents combinations of AI models, domains, and beneficiaries. Activity assistance (33 occurrences), activity recognition (25 occurrences), and interaction (14 occurrences) mainly used DL models. Similarly, and to a smaller extent, NLP models helped with activity assistance (26 occurrences), activity recognition (19 occurrences), interaction (15 occurrences), and communication (10 occurrences).

Combinations of AI models and beneficiaries highlight older adults as the leading users of DL (27 occurrences) and NLP (26

XSL•FO
RenderX

occurrences). Patients and frail persons coexisted with DL models 11 times each.

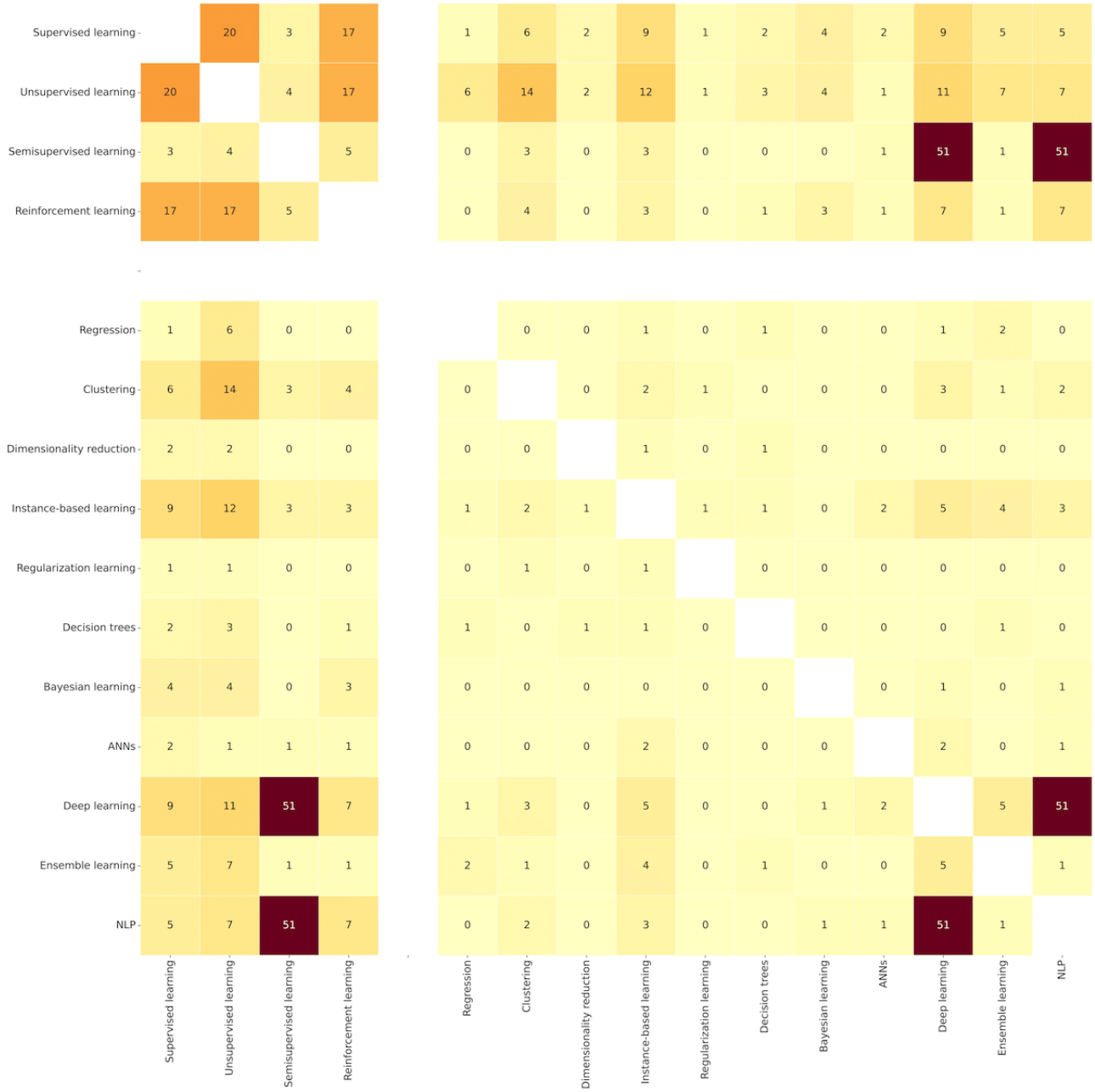
The coappearance of beneficiaries and domains reveals that activity assistance targeted mainly older adults (27 occurrences), followed by activity recognition (17 occurrences) and communication (10 occurrences).

As for connections within categories, activity recognition is a common form of assistance (38 occurrences), followed by communication (12 occurrences), interaction (12 occurrences), and health monitoring (10 occurrences). Patients and frail persons co-occurred 11 times. Older adults were referred to as frail persons and patients 9 times each, indicating that AI models mainly serve healthy older adult users. Family, caregivers, and health care staff rarely appeared together in these articles.

Figure 7 shows instances and connections between the technology, beneficiaries, and concerns. Relationships between the nodes from different categories reveal that older adults commonly used ambient sensing technology (9 occurrences), mobile devices (7 occurrences), and robots (6 occurrences). At the same time, their primary concerns were availability (7 occurrences), usability (5 occurrences), and safety and accessibility (4 occurrences per category). Availability is a concern for patients (4 occurrences). Moreover, availability is the primary concern in wearable technology (5 occurrences), along with ambient sensing and mobile technology (4 per category).

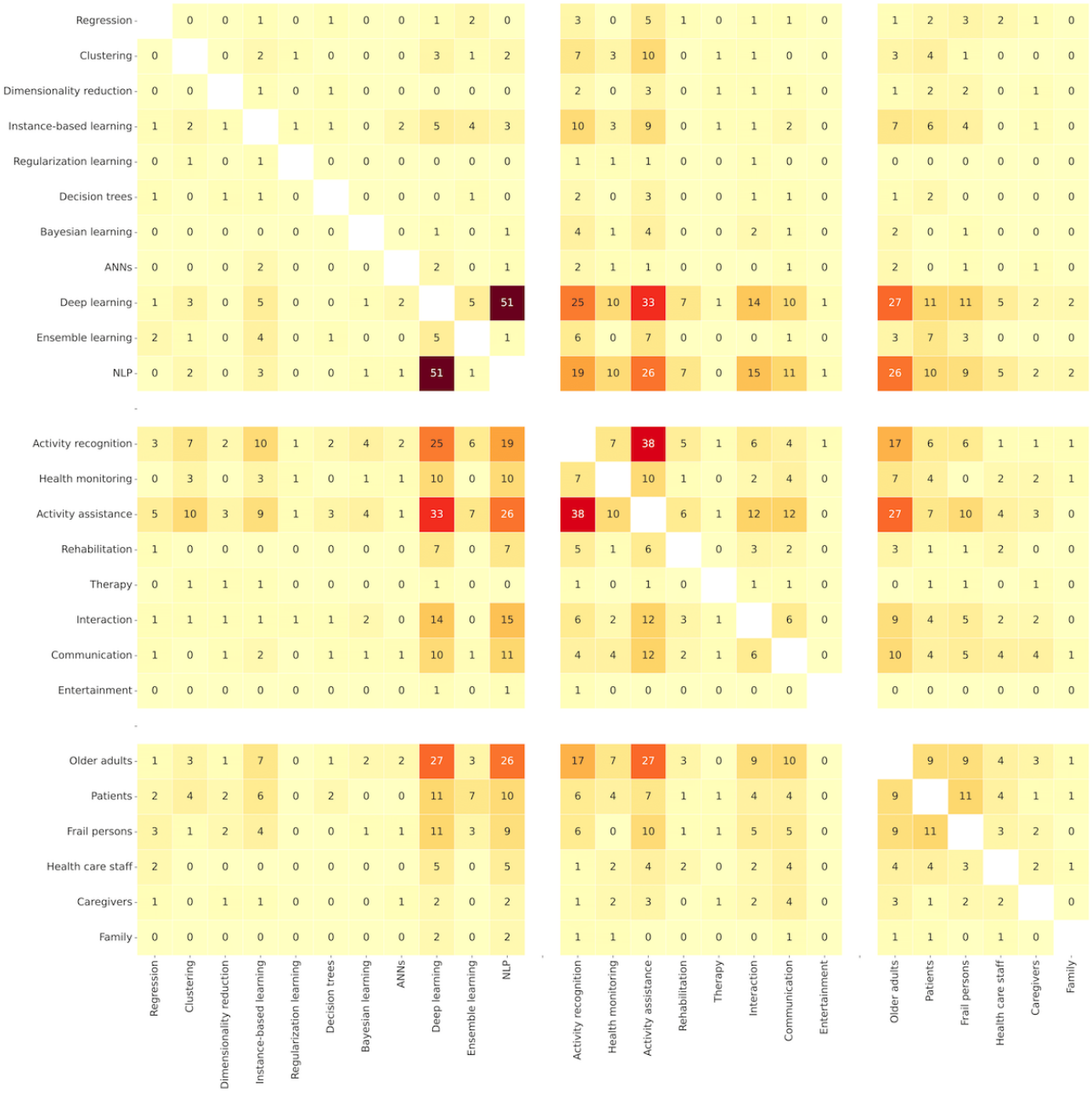
Links between the instances within a category indicate occasional use of ambient sensing and wearable technology with mobile devices 4 and 3 times, respectively.

Figure 5. The heat map describing co-occurrences of artificial intelligence classes and models in relevant articles. ANN: artificial neural network; NLP: natural language processing.



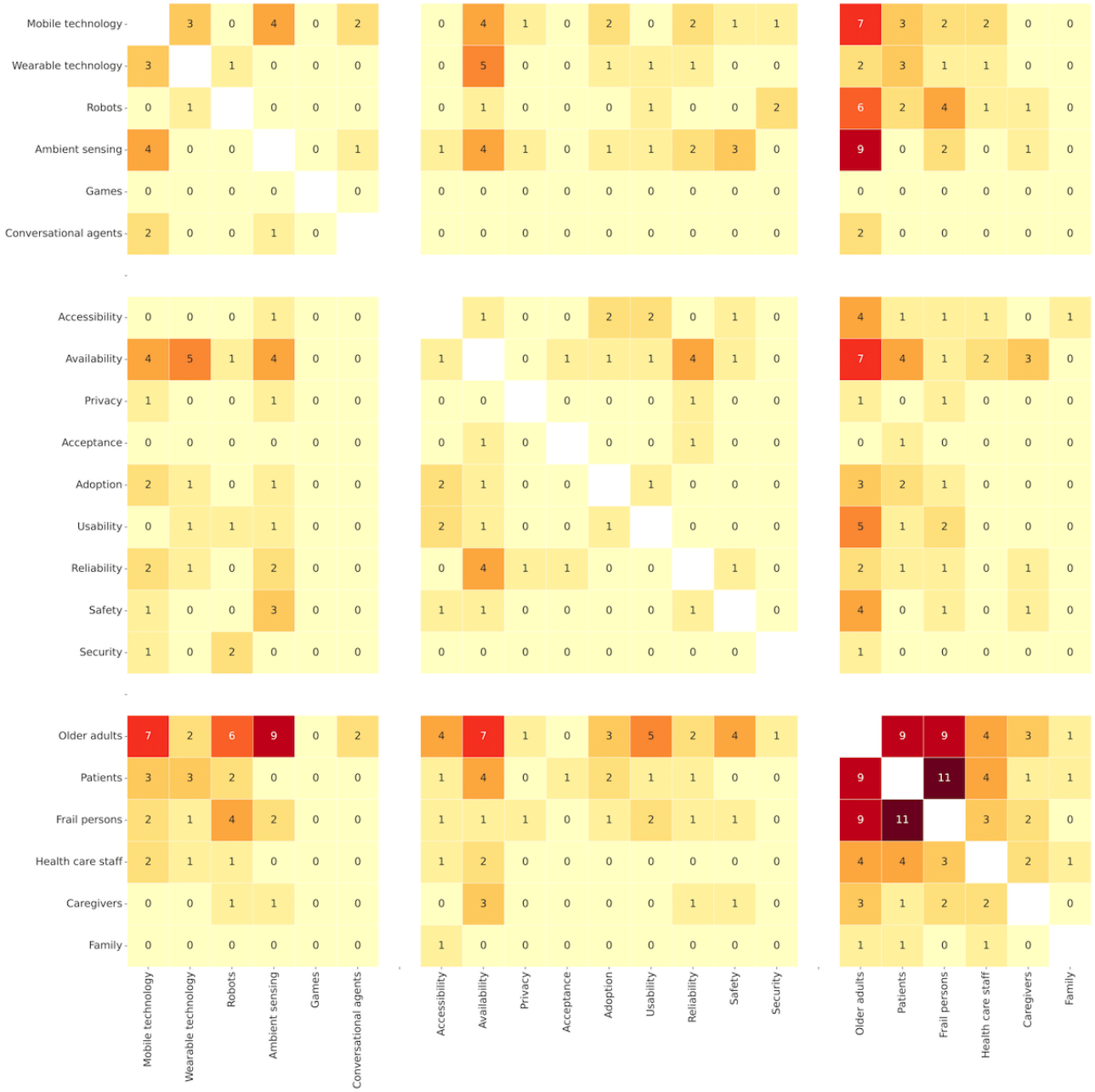
XSL•FO
RenderX

Figure 6. The heat map describing co-occurrences of artificial intelligence models, domains, and beneficiaries in relevant articles. ANN: artificial neural network; NLP: natural language processing.



XSL•FO
RenderX

Figure 7. The heat map describing co-occurrences of technology, beneficiaries, and concerns in relevant articles.



AI Classes and Models in AAL

Concerning the classes, the analysis of relevant articles (Figure 8) showed the highest presence of semisupervised learning (52 occurrences), followed by unsupervised learning (50 occurrences), supervised learning (29 occurrences), and reinforcement learning (20 occurrences).

The distribution of AI classes shows *that semisupervised learning* models had prevailed since the 2010s, with an irregular growing trend until 2017, when they compensated for the lack of a sufficient amount of labeled data for particular inputs. Concrete examples include clustering for physical activity recognition [37], finding relevant input features for improving activity recognition [27], and detecting user-object interactions from sequences of images [143].

The *unsupervised learning* model trend follows the previous category, with a slightly smaller number of appearances. Their use was motivated by a general lack of annotated (or labeled) training data for various activities that early AAL solutions aimed to support [4,8]. Such problems were tackled mainly by either grouping according to shared properties or simplifying input data. The growing trend that followed was caused by the emergence of new health-related domains and activities that solutions were targeting and for which the labeled data did not exist. The examples include the recognition and measurement of everyday activities from unlabeled data [38], clustering to create an ontology of human activities [39], or classification for predicting user movements indoors [40].

Supervised learning models showed general growth until 2020. They complemented other approaches (eg, unsupervised and reinforcement learning; Figure 5) for particular user activities for which labeled data existed. They were used in various

XSL•FO
RenderX

classification tasks within the AAL, such as user reidentification with red green blue depth (RGBD) cameras [41] and ADL recognition using wearable sensors [42].

Reinforcement learning models were used from 2010, increasing use until 2015 and reducing use after 2017. They have served as alternatives to data-driven approaches (ie, clustering and regression) by promoting desirable behaviors and eliminating undesirable user behaviors. Hidden Markov models are the most common algorithms in applications, including user activity recognition from appliance consumption data [43] or with multiple Kinect devices [44].

Regarding the models, the study revealed the prevalence of DL (63 occurrences), followed by NLP (54 occurrences); instance-based learning (20 occurrences); clustering (17 occurrences); ensemble learning (12 occurrences); regression (7 occurrences); Bayesian learning, decision tree learning, and dimensionality reduction (4 each); artificial neural network (ANN; 3 occurrences); and regularization learning (2 occurrences).

In the following sections, we describe each model according to its prevalence (Figure 9).

DL models gained momentum in 2017, expanding their use cases up to date. Their application assumes labeled input data of different structures and semantics are generated at scale. Convolutional neural networks are the most common algorithms used independently or in combination within this model. Examples are activity recognition by transforming data from smartphone sensors into image-based representations [45] or detecting human postures using RGBD cameras [46].

The use of *NLP* models can be divided into 2 stages—earlier applications (up to 2015), which focused on speech recognition and natural language understanding, and later applications, which could also perform dialog management and natural language generation. We can explain this trend with the critical advancements in conversational AI facilitated by the DL algorithms that overlap with our search time frame [34]. For example, the detection of acoustic events (eg, knock, cough, and clap) for older adults in ADL [47] versus conversation with a companion robot [144].

Instance-based learning models were used throughout the search period, with an irregular trend and a recent drop from 2020. They used mainly k-nearest neighbor and support vector machine algorithms. The use cases include recognizing physical activity patterns at home with a multiview infrared motion sensing system [48] or detecting ADL from human joint trajectories captured with a depth camera [49].

Clustering learning models were used in specific years of the search time frame, mainly unsupervised, as an alternative approach in the absence of labeled data concerning particular use cases. The use cases include predicting a sequence of connected users' actions in a robotic device [50] or detecting dining-related postures from motion sensor data [51].

Ensemble learning models were used starting from 2014. Boosting and random forest are the main algorithms in this model, including physical activity classification from wearable sensors [52] and seizure and fall detection from a smartphone's accelerometer data [53], respectively.

The remaining models were used to a smaller extent during the search time frame.

Regression learning models were mainly linear regression, such as real-time energy expenditure estimation when walking with loads and on inclines assisted by an ankle exoskeleton [54] or ADL recognition from hand grasps using electroencephalography [55]. *Bayesian learning* models were applied to classification problems such as ADL recognition (ie, detection and classification) using data collected from wearable motion sensors [42]. *Decision tree* and *dimensionality reduction* models were used for classification tasks. The respective examples classify physical activities based on step counts [56] or Wi-Fi and wearables' data [57]. ANNs were applied to classification problems as a predecessor of DL models, such as activity recognition in safety-critical environments (eg, fall detection) [58]. Finally, *regularization learning* models were applied to regression tasks, such as selecting predictive input features for person identification with RGBD cameras [41].

Although they were mentioned in some articles in the context of previous, relevant, or future work, our analysis did not reveal the examples of *association rule learning* models' algorithms in AAL.

XSL•FO
RenderX

Figure 8. The number and annual distribution of the relevant articles concerning artificial intelligence classes from January 2010 to July 2022.

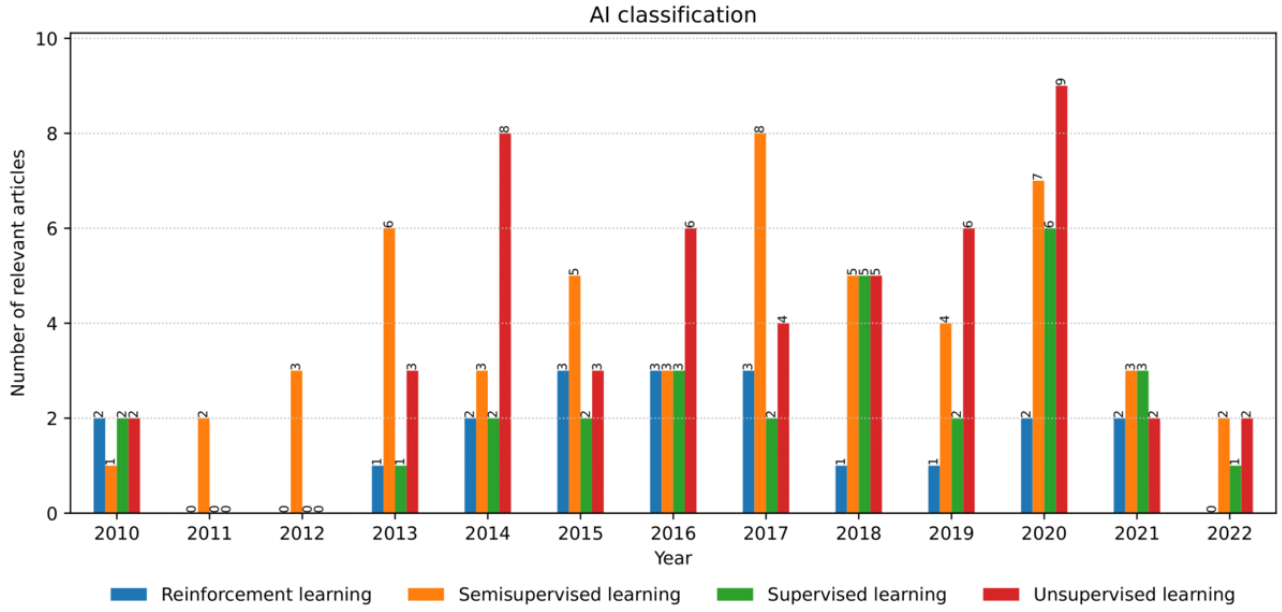
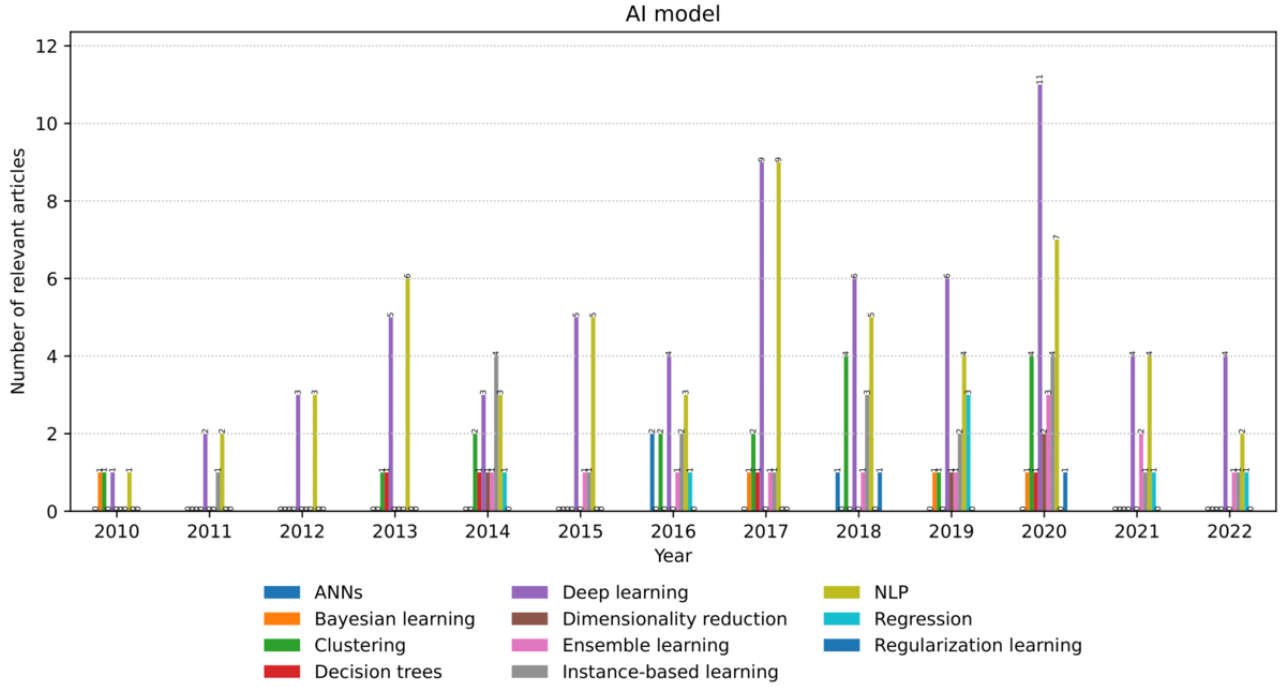


Figure 9. The number and annual distribution of the relevant articles concerning artificial intelligence models from January 2010 to July 2022. ANN: artificial neural network; NLP: natural language processing.



Domains of AI Models

AI models were applied in multiple domains, per domain, or in combination (Figure 10). The most popular domains found were *activity assistance* (61 occurrences) and *activity recognition* (45 occurrences). The domains showed a growing trend until 2020, with periodic oscillations during the time frame. As indicated earlier, they were mainly interconnected in previous studies (38 occurrences; Figure 6). Activity assistance has been a significant target in AAL and assistive technologies in general. Mobility is a common assisted activity, such as a robotic walker for mobility in older adults [59] or smart glasses helping visually impaired users navigate physical spaces [60]. Human activity recognition (HAR) is a commonly used term to describe the recognition of various physical activities. These activities are usually classified into ADL (health focused) and instrumental ADL (IADL; well-being focused), indicating that intelligent AAL systems support health and quality of life. An essential challenge in activity recognition is predicting long-term behavior [61]. Similarly, some studies dealt with the problem of multisensor data fusion in a robotic walker for indoor assistance [62].

Interaction (21 occurrences) referred to the use of different AAL systems, whereas *communication* (17 occurrences) was mainly considered from a technical perspective (eg, communicating between sensors, servers, and cloud-based

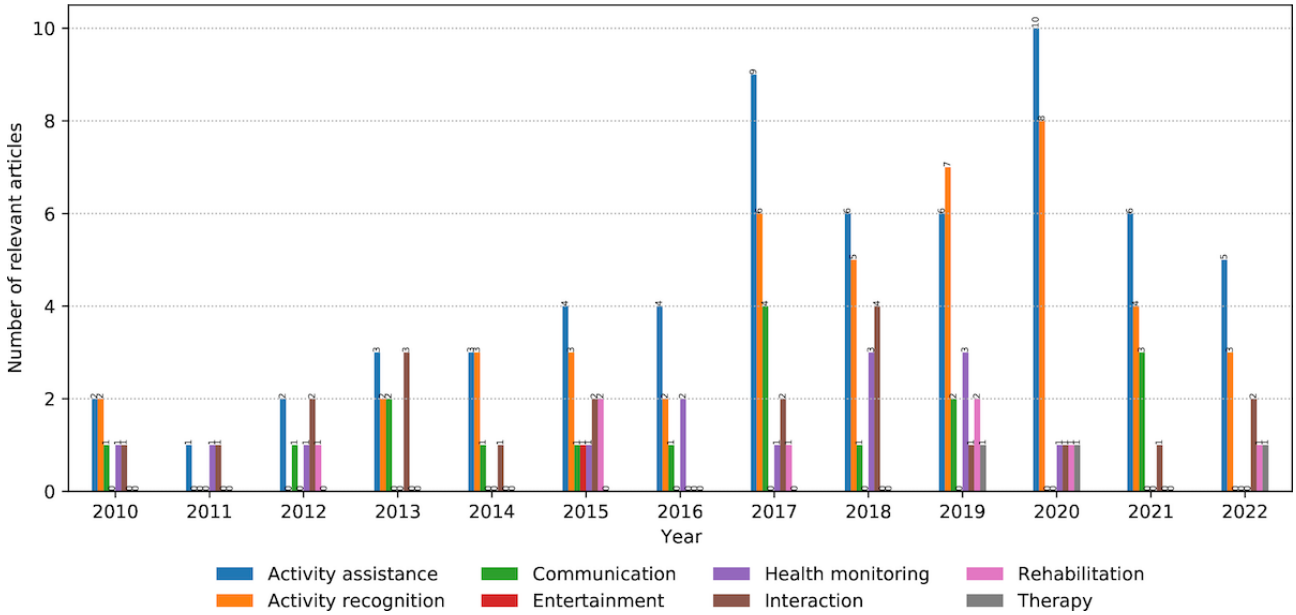
XSL•FO
RenderX

systems). The former examples include interacting with an innovative home platform for facial emotion recognition [46] or a medication delivery application [63]. The latter is a distributed multimedia system for patient data capture [64] or digital footprint applications for the activity prediction of assisted users [65].

Health monitoring (14 occurrences) was commonly referred to as observing users' vital signs to detect changes in health conditions and emergencies, such as health-related data collection for users at their homes [64] or in-home gate analysis using radar sensors [61].

Rehabilitation (8 occurrences), therapy (3 occurrences), and entertainment (1 occurrences) received less attention from the research community in the search time frame. A rehabilitation example is a home system suggesting medications and exercises during fall recovery [66].

Figure 10. The number and annual distribution of relevant articles concerning the artificial intelligence models' domains from January 2010 to July 2022.



Technologies Using AI Models

Ambient sensing and *mobile technology* (15 occurrences per category) prevail in AAL (Figure 11). It is an umbrella term that connotes various sensors that measure the parameters of the observed environment (or ambient environment) to detect and analyze user behavior. In this respect, studies used a particular sensor or combined multiple sensors. In the former case, vision [49] and radar [67] sensors were used for recognizing activities and measuring vital signs, respectively. In the latter case, studies merged signals from various sensors for energy efficiency and improved accuracy and performance (known as sensor fusion). For example, activity recognition combined depth image sequences and audio data [68].

Mobile technologies exposed 2 typical roles. A *passive role* in using their embedded sensors and providing a user interface for measuring the conditions in the users' environments or the state of their behaviors [53]. An *active role* in promoting healthy habits and behaviors among users for a positive lifestyle change by suggesting activities [65].

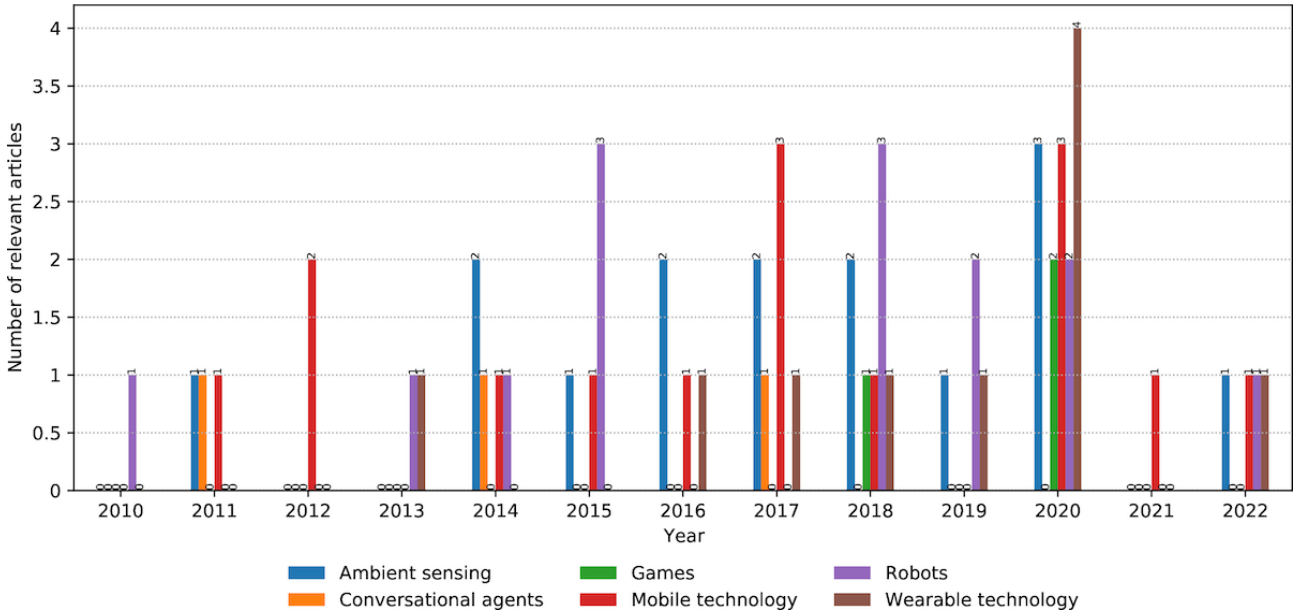
Robotic technology (14 occurrences) was used during the time frame, with an irregular trend. Robots fit well with the AAL paradigm, as they replicate human abilities and characteristics, but the cost of development and deployment may influence their use. In line with related work, we noticed their *assistive* and *companionship* purposes. The former concerns upper-limb gesture recognition to help users with ADL [69]. The latter is demonstrated by interacting with older adults to prevent social isolation and mediating between the older adult, the environment, and the AAL system [144].

Wearable technology was used less frequently than previously (10 occurrences). On the one hand, it can introduce a certain level of intrusiveness compared with ambient sensors when used independently. On the other hand, it is available through mobile devices (eg, smartwatches and bracelets), and the study identified 3 overlaps (Figure 7). The study by Slade et al [54] used wearable sensors attached to users' ankles to estimate energy consumption when walking. Another example is activity recognition, which accounts for measurement uncertainty in wearable sensors [42].

Conversational and *gaming* technologies have 3 occurrences each. The conversation example is a social robot that conducts simplified small-talk dialogs with users [144]. Overall, the dialogs were rare compared with the many occurrences of NLP models used for speech and text recognition. Games were mentioned as the use of gaming technologies (eg, Kinect RGBD camera) that recognized human activity [145].

XSL•FO
RenderX

Figure 11. The number and annual distribution of relevant articles concerning the artificial intelligence models' technology from January 2010 to July 2022.



Beneficiaries of AI Models

Older adults are the primary beneficiaries of intelligent AAL systems (42 occurrences; Figure 12). The shared elements emerging from different AAL systems using AI models for this target group include HAR and measuring vital signs. Accordingly, the study described by Saeed et al [70] used radar sensors' data to infer the activities of community-dwelling older adults, while research by Ejupi et al [71] detected falls by analyzing accelerometer and barometric pressure sensor data.

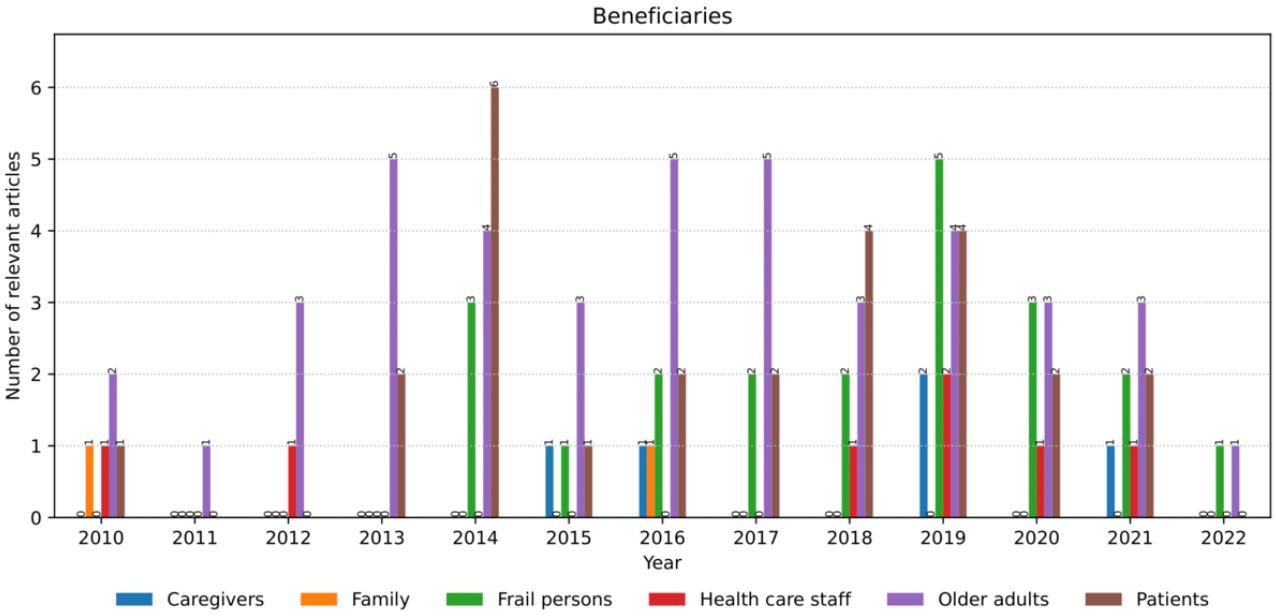
Patients (n=26) were persons with health declines who underwent different medical treatments. Examples include activity prediction for fall prevention of patients at risk [44] and diagnosing clinical abnormalities of patients using multiple vital signs (eg, heart rate, blood pressure, and respiratory rate) [72]. The overlap with older adults (9 times; Figure 6) indicates that for most older people, the purpose of AAL systems was more assistive and aimed toward health promotion rather than therapy.

Frail persons (n=21) were in a specific state of vulnerability with increased risks of falling or disability. The AAL support for these beneficiaries was manifested in diagnosing various health declines. Examples include diagnosing Alzheimer disease from magnetic resonance images [73] and detecting emergencies with users' mobility [54].

Health care staff (n=7), caregivers (n=5), and family (n=2) were considerably less present than previously. Owing to the AAL technology used, they appeared as beneficiaries concerning more efficient and effective caregiving. For example, supporting medical staff in monitoring patients at home [66] and notifying physicians and families if patient conditions decline[64].

XSL•FO
RenderX

Figure 12. The number and annual distribution of relevant articles concerning the artificial intelligence models' beneficiaries from January 2010 to July 2022.



Concerns in AI Models

Availability of intelligent AAL systems was the primary user concern (17 occurrences; Figure 13). It was mentioned mainly concerning a particular technology. For example, beneficiaries preferred off-the-shelf technologies, such as mobile devices because of their availability regarding services they can offer and cost [53]. Conversely, the availability of particular devices, such as exoskeletons [54] or multiple Kinect devices [145], was highlighted as a potential barrier to their use.

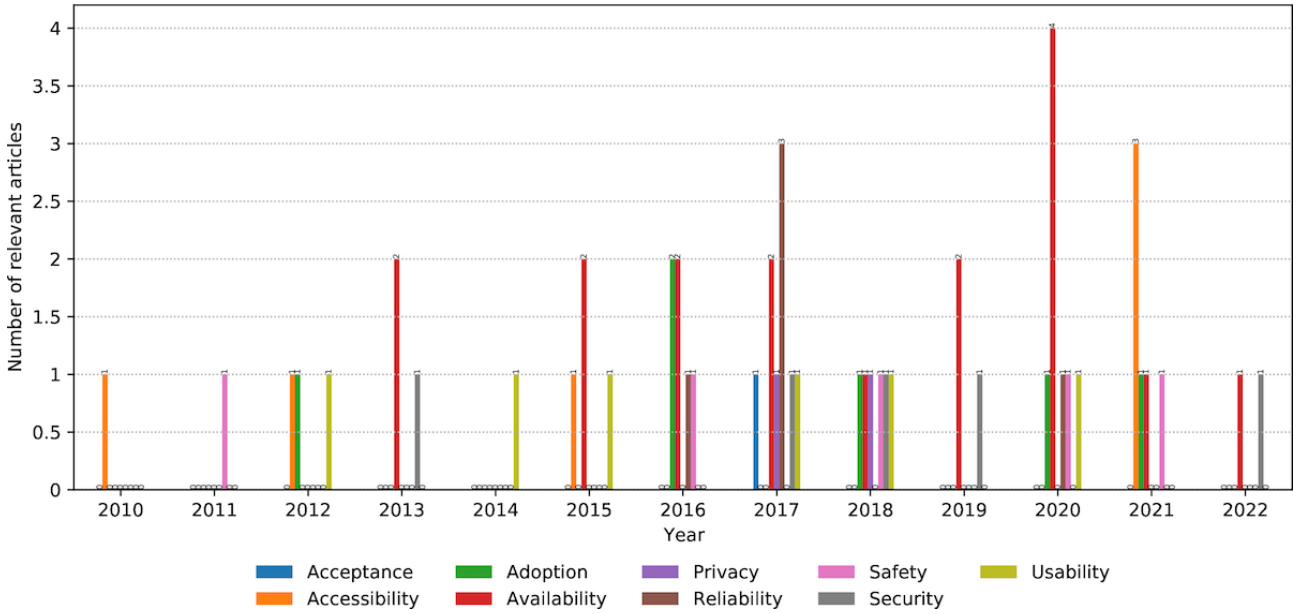
Accessibility, *adoption*, and *usability* appeared 6 times each. The accessibility of the AAL systems' services reflects the convenience of reaching them, such as the functions of an Internet of Things device that generates user profiles from their activities [74]. The adoption referred to a more sustained and stable use and integration of the introduced technology into the beneficiaries' routine, such as the use of widely adopted technologies (ie, smartphones) for HAR [45]. Usability was described as the ease of use of various AAL systems, including the percentage of successful task completion when using a medication management application [63].

Reliability, *safety*, and *security* had 5 occurrences each. Reliability describes the reliability levels of the AI models' outcomes from the beneficiaries' perspective, such as the perceived accuracy of activity trackers [56]. Safety is a requirement for AAL applications to prevent any harm to their users, such as detecting abnormal human behaviors to avoid dangerous situations [75]. Security refers to protecting users from external threats when using AAL technology; for example, when using users' appliance consumption data to infer their activities [43].

Privacy (2 occurrences) and *acceptance* (1 occurrence) received the least attention. Privacy manifested as the need to protect beneficiaries' data during collection, analysis, and use by the AAL system, such as protecting persons' identities [41]. Acceptance emerged as a desired quality of the AAL technology that facilitates the attitude, such as the unobtrusiveness of radar-based sensors for patient monitoring [61].

XSL•FO
RenderX

Figure 13. The number and annual distribution of relevant articles regarding the beneficiaries' concerns from January 2010 to July 2022.



Discussion

Overview

This section summarizes the results of the scoping review. The AI models are key drivers of AAL systems. In this respect, this study clarifies their role and significance over the previous decade by considering domains, technologies, and end users. At the same time, it highlights critical user concerns to identify gaps that require further research.

The overall goal was to provide an overview and synthesis of the research on AI classes and models in AAL (RQ1), domains in which they were applied (RQ2), technologies that used them (RQ3), and their beneficiaries along with use concerns (RQ4).

The following section discusses the principal findings concerning the evolution of the AI models and related categories and implications for different stakeholder groups, including well-being and health care, technology, and research.

Principal Findings

Evolution of Categories in AAL

The time frame has seen a variety of AI *model* contributions that target a range of *domains*, *technologies*, and *beneficiaries*. Semisupervised and unsupervised learning classes dominate the intelligent AAL landscape. Their prevalence is because of an increase in the variety of the health and living domain and the gradual appearance of labeled input data describing related ADL and IADL the learning aimed to support [37,38,143]. Supervised approaches have been used for classification tasks [41].

DL and NLP models have been mainly used throughout the search time frame. DL models combined neural network–based algorithms such as convolutional neural network and recurrent neural network [146]. These algorithms can be both supervised and unsupervised but were rarely considered explicitly in relevant articles. However, an in-depth manual article analysis showed that they were mainly supervised. These algorithms deal with multidimensional input data from heterogeneous sources. The data describe various human activities to support or infer health conditions [45,46]. In the first half of the frame, NLP models mainly recognized users' spoken input [47]. During the second half, they enabled conversations with users [144]. Models using reference examples (ie, instance-based learning) and clustering were used for classification tasks [48-51]. Ensemble approaches, by definition, combine separate models to compensate for individual drawbacks [52,53] and were used later in the time frame (from 2014). However, other approaches were notably less frequently used.

Activity assistance and recognition were the leading domains, with a generally growing trend. In most cases, the activity assistance assumed recognition (38/61, 62% occurrences), while the remaining instances focused on specific activities known in advance. A range of ADL and IADL were supported, where different indoor and outdoor mobility (ie, walking, physical exercise, and transportation) prevailed [59,60]. The interaction referred to the systems as seen by their end users [63]. Communication denoted internal interrelations among AAL system components [64]. Health monitoring concentrated on deviations in vital functions and detection of abnormal behaviors [61].

Ambient sensing and mobile technology are mainly used in AAL. Sensing uses different sensors to detect available signals that carry specific information on user behavior [68]. Mobile technologies are convenient to use (ie, market availability, affordability, and wide adoption) at the application level as lifestyle applications for health and well-being [65] and at the device level as a platform with integrated sensors [53]. Robots appeared as either assistive devices helping users carry out their activities [69] or companions for pleasurable activities [144].

The study found notably fewer wearables, followed by conversational and gaming technologies.

Older adults were the primary beneficiaries of AI models in AAL within the search time frame [70]. Patients coexisted with

XSL•FO
RenderX

older adults 9 of 26 times or 35% of found cases (Figure 6). This shows that other ages benefited from AI models [44,72]. Frail persons were less present and coappeared with older adults in 43% (9/12) of instances. Health care staff, caregivers, and families were underrepresented compared with the former and occasionally mentioned.

Availability prevails as a beneficiary concern. In general, off-the-shelf, affordable technology [53] is preferred compared with more expensive equipment concerning cost and deployment [54]. The remaining concerns had fewer cases.

Implications for Health Care and Well-being

On the basis of our observations concerning domains, beneficiaries, and concerns, we identified gaps in the existing literature and articulated the following directions for future work:

- Collaborative decision-making—current AAL systems make their decisions autonomously, driven by the models' algorithms and input data. The involvement of expert users (ie, health care staff and caregivers) in the decision-making process can improve its accuracy, facilitate automated learning about users, and reduce the burden on health care professionals.
- Augmenting caregivers and recipients—by definition, the AAL occurs outside health care facilities. In such a scenario, consideration of caregiving and caretaking is critical for adherence to health care services that should address the participants' concerns. Active participation of these beneficiaries is crucial for a successful digital health care intervention, from their AI model comprehension to a particular technology design and deployment.
- AAL interventions—studies included various technologies and platforms to support independent living. Our analysis did not reveal knowledge exchange among studies concerning their results and experiences. Technology-supported health care interventions have been designed for various medical domains. Systematized knowledge of models, domains, technologies, and beneficiaries can guide AAL interventions tailored to specific health care requirements. Such knowledge can reinforce best practices and mitigate potential risks.
- Regulations and compliance—at present, AAL design and deployment space are not regulated, nor is their compliance acknowledged and endorsed by regulatory authorities globally. AAL systems must comply with regulations at both national and international levels. This is crucial for their implementation in medical practice and general adoption. To meet this need, we advocate for a repository of evaluation methods and design guidelines that would support compliance and provide a clear view of how to incorporate critical aspects during AAL system design.

Implications for Technology

The analysis of the models, technologies, and concerns revealed unsolved matters that require more attention, including the following:

- Transparency and privacy—AI models, by their very nature, need, produce, and process large amounts of various user-related data, from intensive data collection and analysis to delivering their decisions as personalized recommendations to users. First, the technology should be transparent on why and how user data are collected, analyzed, and used. Second, it should respect a user's right to control their private data and communications and that they are free from intrusion. Satisfying these user needs is critical for building trust in AAL systems.
- Integration with health care services—AAL systems are usually built and deployed as stand-alone platforms, independent of institutional health care systems. Connecting with existing medical technological infrastructure and digital services can increase the efficiency and effectiveness of health care provision. These benefits are mutual. AI models can be fed with existing user medical records and procedures for improved decision-making. In turn, medical actors could be timely informed of emergencies or changes in users' behaviors that are difficult to observe in clinical settings.
- Inclusive AAL—AI models focus on individual users as a user-system relation. Group dynamics are not supported, such as user-system–physician relations or forming peer groups of similar users. Future intelligent AAL systems should equally engage and moderate multiple beneficiaries: patients, families, caregivers, and health care staff. This also represents a general implication for health care systems.

Implications for Research

Looking at the results as RQs' responses, following research directions emerged:

- Explainable decision-making—as capabilities of AI models increase, the absence of explanations behind automated behaviors raises uncertainty with users due to a lack of understanding of how specific decisions are made [147]. The explanatory behavior of the models can ingrain positive behaviors to maintain a healthy lifestyle [148]. Thus, a general requirement for future AI models is the provision of explanations understandable to beneficiaries without background or knowledge in AI (ie, nonexperts).
- Evaluation techniques—studies proposed evaluation techniques that could be broadly categorized into functional (ie, technical) and nonfunctional (ie, medical and usability). They used existing instruments to measure AI models' algorithms (accuracy and performance) or medical and user-related outcomes (standard scales for particular medical conditions, interviews, and questionnaires). Moreover, they focused on a single measure or several measures from the same category. To obtain a clear and valid assessment of the effectiveness and efficiency of the AI models used in AAL, we need a more comprehensive and coherent set of cross-category evaluation metrics to be proposed and verified in practice.
- Design recommendations—the discovery of design guidelines from relevant articles depends on how they are described. In the analysis phase, the identification and extraction of guidelines were not straightforward. The design contributions were mainly presented as suggestions derived from the studies conducted. Other forms included development and deployment practices concerning specific

XSL•FO
RenderX

models, domains, and technologies. These contributions are difficult to apply and reproduce, being a barrier to their uptake. Standard reporting procedures and knowledge bases could help address this issue and provide actionable guidelines to interested communities. Several independent studies will be needed to implement and validate the guidelines.

Limitations

We acknowledge that the proposed AI and machine learning (ML) class and model categorization, which served as a basis for our search, is not comprehensive, exhaustive, or exclusive. Although there are other taxonomies, our goal was to highlight the underlying mechanisms of these classes and models for the review to provide a proper understanding of their roles in AAL systems.

Moreover, the categories and associated keywords may have limited the search results. Thus, we included common synonyms found in the literature as keywords to capture more results at the cost of more nonrelevant articles. Still, we may have missed relevant materials using other terms or not using searched keywords explicitly.

Another limitation of our study was the necessity of setting a time frame for the articles included in the review. We chose to cover work by early reviews of AAL systems and advancements in AI learning algorithms. However, as with any date restriction, there is a risk of not considering potentially relevant work.

A further limitation concerns manual extraction and categorization of retrieved articles (for inclusion), which may introduce a subjective perception of coders. The risk was addressed by cross-analysis and discussion of each other's results for agreement. Relatedly, the findings on prevalence or trend may primarily represent the researchers' interest but not an objective sampling of all the stakeholders' perspectives, including that of the users.

Finally, this study considered 5 digital libraries, among others. Considering the size, coverage, and diversity of digital libraries regarding RQs, we believe that the obtained results sufficiently respond to them.

Comparison With Prior Work

In comparison with relevant work, we focus on previous reviews and metareviews on related topics and comparative studies, giving preference to AI models. The reviews', metareviews', and studies' scope was generally more constrained than ours.

The meta-review presented by Climent-Perez et al [16] examined video-based lifelogging technologies for AAL in older adults. Lifelogging assumes recording personal data of a user's daily life. It produces a data set as computational knowledge about a person (also known as quantified self) that could be used for different purposes, such as detecting emergencies and predicting user behavior. The target model was DL, domain HAR, and technology RGBD sensing devices. This study articulates ethical implications for these applications.

The review by Singh et al [149] analyzed existing fall-detection systems through the implementation of existing sensor technologies. It provides a descriptive framework to help choose appropriate sensors for particular deployment scenarios and locations. The main areas for technical improvement were unobtrusiveness, installation costs, and power requirements.

A survey by Demrozi et al [150] discussed ML and DL algorithms for sensor-based HAR of older adults concerning their accuracy and quantity (coverage of recognized activities). ML models require less data and computational resources, whereas DL models better recognize complex activities.

A review of mobile apps for dementia [151] showed that caregivers were the primary users, and the app content mainly provided information on dementia. The barrier to the availability of these apps is a lack of navigating the app marketplace and quality metrics for their dementia information.

A review of DL techniques used in smartphones and wearable sensor-based HAR systems [152] demonstrated that DL techniques outperform other ML ones. However, they were verified on preexisting data sets, not the data acquired in real time.

An in-depth analysis of DL algorithms for HAR using mobile and wearable sensor networks [146] raised the need for higher computational resources in mobile and wearable devices to enable web-based and real-time decision-making.

A more comprehensive review of assistive technologies for older adults classified technologies into clusters, such as general information and communication technology (eg, computer and internet applications), robotics, telemedicine, sensor technology, medication management applications, and video games [17].

A study analyzed randomized controlled trials on the effectiveness of assistive technology for memory support in people with dementia [153]. Measured outcomes included ADL, level of dependency, clinical and care-related outcomes, and perceived quality of life and well-being. The evidence was mixed and inconsistent and drew no generalized conclusions.

Another review investigated mobile health interventions for adults who had experienced stroke [154]. The interventions targeted different patient functions, mostly upper-extremity function, functional mobility, and language and speech skills. However, they were mainly preliminary, focused on technology development up to pilot testing, and lacked evidence from large-scale trials.

Off-the-shelf voice assistants were used by persons with motor, linguistic, and cognitive disabilities [155]. Although these systems are widespread, inexpensive, and nonstigmatizing compared with other assistive technologies, participants' performances depended on their level of cognitive and linguistic skills.

A comparative study of different ML algorithms for HAR [156] used existing data sets and indicated that sensor-based techniques were preferred over vision-based techniques because they better preserve user privacy. A similar study [157] examined particular algorithms, namely decision tree, k-nearest neighbor, support vector machines, naïve Bayes, linear discriminant analysis, and ensemble learning, in recognizing specific ADL (meal preparation, eating, housekeeping, etc). In

XSL•FO
RenderX

general, the algorithms performed equally well on the chosen data set.

Conclusions

We have described a scoping review based on systematic search and analysis, which identified research trends concerning AI models, domains, technologies, and beneficiaries along with their concerns. The AI models, domains, technologies, beneficiaries, and concerns extracted from the literature represent a knowledge base that can be consulted and used when developing and deploying AI-infused AAL systems. Its findings can (1) inform end users, health care professionals, and caregivers on available technologies and their target medical domains; (2) guide health care providers and engineers in implementing and deploying these technologies; and (3) help end users understand the benefits and trade-offs of the technologies.

Research activity has increased awareness of AI models in AAL and revealed gaps in the field. Further work is needed in making AAL systems more efficient, effective, and user friendly. In particular, hybrid physician-model decision-making, the inclusion of caregivers by technology design, and compliance with health-related regulations will lead to the uptake of AAL by a society. Moreover, improving transparency and privacy, integration with legacy systems, and the equal inclusion of different beneficiaries will improve the acceptance and availability of AAL systems. Finally, efforts to explain automated decision-making, adopt standard evaluation metrics, and verify design guidelines will recognize different AAL approaches to ensure them in digital health care.

Acknowledgments

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).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The list of the relevant articles retrieved from the 5 digital libraries.
[XLSX File (Microsoft Excel File), 120 KB - jmir_v24i11e36553_app1.xlsx]

References

1. Sun H, De Florio V, Gui N, Blondia C. Promises and challenges of ambient assisted living systems. In: Proceedings of the 6th International Conference on Information Technology: New Generations. 2009 Presented at: ITNG '09; April 27-29, 2009; Las Vegas, NV, USA p. 1201-1207. [doi: 10.1109/ITNG.2009.169]
2. Buyl R, Beogo I, Fobelets M, Deletroz C, Van Landuyt P, Dequanter S, et al. e-Health interventions for healthy aging: a systematic review. Syst Rev 2020 Jun 03;9(1):128 [FREE Full text] [doi: 10.1186/s13643-020-01385-8] [Medline: 32493515]
3. Taylor J, Walsh S, Kwok W, Pinheiro MB, de Oliveira JS, Hassett L, et al. A scoping review of physical activity interventions for older adults. Int J Behav Nutr Phys Act 2021 Jun 30;18(1):82 [FREE Full text] [doi: 10.1186/s12966-021-01140-9] [Medline: 34193157]
4. Rashidi P, Mihailidis A. A survey on ambient-assisted living tools for older adults. IEEE J Biomed Health Inform 2013 May;17(3):579-590. [doi: 10.1109/jbhi.2012.2234129] [Medline: 24592460]
5. Population structure and ageing. Eurostat. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing [accessed 2022-08-15]
6. Ageing and health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2022-08-15]
7. Pal D, Funilkul S, Charoenkitkarn N, Kanthamanon P. Internet-of-things and smart homes for elderly healthcare: an end user perspective. IEEE Access 2018 Feb 22;6:10483-10496. [doi: 10.1109/access.2018.2808472]
8. Calvaresi D, Cesarini D, Sernani P, Marinoni M, Dragoni AF, Sturm A. Exploring the ambient assisted living domain: a systematic review. J Ambient Intell Human Comput 2016 May 4;8(2):239-257. [doi: 10.1007/s12652-016-0374-3]
9. Health promotion and disease prevention through population-based interventions, including action to address social determinants and health inequity. World Health Organization. URL: http://www.emro.who.int/about-who/public-health-functions/health-promotion-disease-prevention.html [accessed 2022-08-15]
10. Garcés L, Ampatzoglou A, Avgeriou P, Nakagawa EY. Quality attributes and quality models for ambient assisted living software systems: a systematic mapping. Inf Soft Technol 2017 Feb;82:121-138. [doi: 10.1016/j.infsof.2016.10.005]
11. Jovanović M, De Angeli A, McNeill A, Coventry L. User requirements for inclusive technology for older adults. Int J Human Comput Interact 2021 May 12;37(20):1947-1965. [doi: 10.1080/10447318.2021.1921365]
12. Offermann-van Heek J, Ziefle M. Nothing else matters! Trade-offs between perceived benefits and barriers of AAL technology usage. Front Public Health 2019 Jun 12;7:134 [FREE Full text] [doi: 10.3389/fpubh.2019.00134] [Medline: 31245342]

XSL•FO
RenderX

13. Whelan S, Murphy K, Barrett E, Krusche C, Santorelli A, Casey D. Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: a literature review. Int J of Soc Robotics 2018 Mar 1;10(5):643-668. [doi: 10.1007/s12369-018-0471-x]
14. Skjæret N, Nawaz A, Morat T, Schoene D, Helbostad JL, Vereijken B. Exercise and rehabilitation delivered through exergames in older adults: an integrative review of technologies, safety and efficacy. Int J Med Inform 2016 Jan;85(1):1-16. [doi: 10.1016/j.ijmedinf.2015.10.008] [Medline: 26559887]
15. Ferrari F, Divan S, Guerrero C, Zenatti F, Guidolin R, Palopoli L, et al. Human–robot interaction analysis for a smart walker for elderly: the ACANTO interactive guidance system. Int J of Soc Robotics 2019 Jun 25;12(2):479-492. [doi: 10.1007/s12369-019-00572-5]
16. Climent-Pérez P, Spinsante S, Mihailidis A, Florez-Revuelta F. A review on video-based active and assisted living technologies for automated lifelogging. Expert Syst Appl 2020 Jan;139:112847. [doi: 10.1016/j.eswa.2019.112847]
17. Khosravi P, Ghapanchi AH. Investigating the effectiveness of technologies applied to assist seniors: a systematic literature review. Int J Med Inform 2016 Jan;85(1):17-26. [doi: 10.1016/j.ijmedinf.2015.05.014] [Medline: 26216463]
18. Danielsen A, Olofsen H, Bremdal BA. Increasing fall risk awareness using wearables: a fall risk awareness protocol. J Biomed Inform 2016 Oct;63:184-194 [FREE Full text] [doi: 10.1016/j.jbi.2016.08.016] [Medline: 27544413]
19. McGlynn SA, Kemple S, Mitzner TL, King CH, Rogers WA. Understanding the potential of PARO for healthy older adults. Int J Hum Comput Stud 2017 Apr;100:33-47 [FREE Full text] [doi: 10.1016/j.ijhcs.2016.12.004] [Medline: 28943748]
20. Orlandini A, Kristoffersson A, Almquist L, Björkman P, Cesta A, Cortellessa G, et al. ExCITE Project: a review of forty-two months of robotic telepresence technology evolution. Presence (Camb) 2016 Dec 01;25(3):204-221. [doi: 10.1162/pres_a_00262]
21. Khosla R, Nguyen K, Chu M. Human robot engagement and acceptability in residential aged care. Int J Human Comput Interact 2017;33(6):510-522. [doi: 10.1080/10447318.2016.1275435]
22. Sun F, Norman IJ, While AE. Physical activity in older people: a systematic review. BMC Public Health 2013 May 06;13:449 [FREE Full text] [doi: 10.1186/1471-2458-13-449] [Medline: 23648225]
23. Szanton SL, Walker RK, Roberts L, Thorpe Jr RJ, Wolff J, Agree E, et al. Older adults' favorite activities are resoundingly active: findings from the NHATS study. Geriatr Nurs 2015;36(2):131-135 [FREE Full text] [doi: 10.1016/j.gerinurse.2014.12.008] [Medline: 25619566]
24. Daly RM, Gianoudis J, Hall T, Mundell NL, Maddison R. Feasibility, usability, and enjoyment of a home-based exercise program delivered via an exercise app for musculoskeletal health in community-dwelling older adults: short-term prospective pilot study. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e21094 [FREE Full text] [doi: 10.2196/21094] [Medline: 33439147]
25. Cohen-Mansfield J, Muff A, Meschiany G, Lev-Ari S. Adequacy of web-based activities as a substitute for in-person activities for older persons during the COVID-19 pandemic: survey study. J Med Internet Res 2021 Jan 22;23(1):e25848 [FREE Full text] [doi: 10.2196/25848] [Medline: 33439851]
26. Kamali ME, Angelini L, Caon M, Carrino F, Röcke C, Guye S, et al. Virtual coaches for older adults' wellbeing: a systematic review. IEEE Access 2020 May 21;8:101884-101902. [doi: 10.1109/ACCESS.2020.2996404]
27. Zdravevski E, Lameski P, Trajkovik V, Kulakov A, Chorbev I, Goleva R, et al. Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. IEEE Access 2017 Mar 21;5:5262-5280. [doi: 10.1109/ACCESS.2017.2684913]
28. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 05;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
29. Cosco TD, Firth J, Vahia I, Sixsmith A, Torous J. Mobilizing mHealth data collection in older adults: challenges and opportunities. JMIR Aging 2019 Mar 19;2(1):e10019 [FREE Full text] [doi: 10.2196/10019] [Medline: 31518253]
30. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
31. Nurgalieva L, Jara Laconich JJ, Baez M, Casati F, Marchese M. A systematic literature review of research-derived touchscreen design guidelines for older adults. IEEE Access 2019 Feb 12;7:22035-22058. [doi: 10.1109/ACCESS.2019.2898467]
32. Brownlee J. Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch. San Francisco, CA, USA: Machine Learning Mastery; 2016.
33. Bonaccorso G. Machine Learning Algorithms: Popular Algorithms for Data Science and Machine Learning. 2nd edition. La Vergne, TN, USA: Ingram Academic; 2017.
34. Otter DW, Medina JR, Kalita JK. A survey of the usages of deep learning for natural language processing. IEEE Trans Neural Netw Learn Syst 2021 Feb;32(2):604-624. [doi: 10.1109/TNNLS.2020.2979670] [Medline: 32324570]
35. 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]
36. Zdravevski E, Lameski P, Trajkovik V, Chorbev I, Goleva R, Pombo N, et al. Automation in systematic, scoping and rapid reviews by an NLP toolkit: a case study in enhanced living environments. In: Ganchev I, Garcia NM, Dobre C,

XSL•FO
RenderX

Mavromoustakis CX, Goleva R, editors. Enhanced Living Environments: Algorithms, Architectures, Platforms, and Systems. Cham, Switzerland: Springer; 2019:1-18.
37. Ali H, Messina E, Bisiani R. Subject-dependent physical activity recognition model framework with a semi-supervised clustering approach. In: Proceedings of the 2013 European Modelling Symposium. 2013 Presented at: EMS '13; November 20-22, 2013; Manchester, UK p. 42-47. [doi: 10.1109/EMS.2013.7]
38. Appiah K, Hunter A, Lotfi A, Waltham C, Dickinson P. Human behavioural analysis with self-organizing map for ambient assisted living. In: Proceedings of the 2014 IEEE International Conference on Fuzzy Systems. 2014 Presented at: FUZZ-IEEE '14; July 6-11, 2014; Beijing, China p. 2430-2437. [doi: 10.1109/FUZZ-IEEE.2014.6891833]
39. Gayathri KS, Easwarakumar KS, Elias S. Contextual pattern clustering for ontology based activity recognition in smart home. In: Proceedings of the 2nd International Conference on Smart Secure Systems – IoT and Analytics Perspective. 2018 Presented at: ICIIT '17; December 20-22, 2017; Chennai, India p. 209-223. [doi: 10.1007/978-981-10-7635-0_16]
40. Zhang X, Chung FL, Wang S. An interpretable fuzzy DBN-based classifier for indoor user movement prediction in ambient assisted living applications. IEEE Trans Ind Inf 2020 Jan;16(1):42-53. [doi: 10.1109/TII.2019.2912625]
41. Paolanti M, Romeo L, Liciotti D, Cenci A, Frontoni E, Zingaretti P. Person re-identification with RGB-D camera in top-view configuration through multiple nearest neighbor classifiers and neighborhood component features selection. Sensors (Basel) 2018 Oct 15;18(10):3471 [FREE Full text] [doi: 10.3390/s18103471] [Medline: 30326647]
42. Akbari A, Jafari R. Personalizing activity recognition models through quantifying different types of uncertainty using wearable sensors. IEEE Trans Biomed Eng 2020 Sep;67(9):2530-2541. [doi: 10.1109/TBME.2019.2963816] [Medline: 31905130]
43. Ferrández-Pastor FJ, Mora-Mora H, Sánchez-Romero JL, Nieto-Hidalgo M, García-Chamizo JM. Interpreting human activity from electrical consumption data using reconfigurable hardware and hidden Markov models. J Ambient Intell Human Comput 2016 Nov 22;8(4):469-483. [doi: 10.1007/s12652-016-0431-y]
44. Haller E, Scarlat G, Mocanu I, Trăscău M. Human activity recognition based on multiple kinects. In: Proceedings of the International Competitions and Final Workshop on Evaluating AAL Systems Through Competitive Benchmarking. 2013 Presented at: EvAAL '13; September 24-26, 2013; Norrköping, Sweden p. 48-59. [doi: 10.1007/978-3-642-41043-7_5]
45. Abdel-Basset M, Hawash H, Chang V, Chakrabortty RK, Ryan M. Deep learning for heterogeneous human activity recognition in complex IoT applications. IEEE Internet Things J 2022 Apr 15;9(8):5653-5665. [doi: 10.1109/jiot.2020.3038416]
46. Amine Elforaici ME, Chaaraoui I, Bouachir W, Ouakrim Y, Mezghani N. Posture recognition using an RGB-D camera: exploring 3D body modeling and deep learning approaches. In: Proceedings of the 2018 IEEE Life Sciences Conference. 2018 Presented at: LSC '18; October 28-30, 2018; Montreal, Canada p. 69-72. [doi: 10.1109/lsc.2018.8572079]
47. Schroeder J, Wabnik S, van Hengel PW, Goetze S. Detection and classification of acoustic events for in-home care. In: Wichert R, Eberhardt B, editors. Ambient Assisted Living. Berlin, Germany: Springer; 2011:181-195.
48. Guan Q, Yin X, Guo X, Wang G. A novel infrared motion sensing system for compressive classification of physical activity. IEEE Sensors J 2016 Apr;16(8):2251-2259. [doi: 10.1109/jsen.2016.2514606]
49. Giannakeris P, Meditskos G, Avgerinakis K, Vrochidis S, Kompatsiaris I. Real-time recognition of daily actions based on 3D joint movements and fisher encoding. In: Proceedings of the 26th International Conference on MultiMedia Modeling. 2020 Presented at: MMM '20; January 5–8, 2020; Daejeon, South Korea p. 601-613. [doi: 10.1007/978-3-030-37734-2_49]
50. Wu C, Zhang J, Sener O, Selman B, Savarese S, Saxena A. Watch-n-patch: unsupervised learning of actions and relations. IEEE Trans Pattern Anal Mach Intell 2018 Feb;40(2):467-481. [doi: 10.1109/TPAMI.2017.2679054] [Medline: 28287959]
51. Tariq M, Majeed H, Beg MO, Khan FA, Derhab A. Accurate detection of sitting posture activities in a secure IoT based assisted living environment. Future Gen Comput Syst 2019 Mar;92:745-757. [doi: 10.1016/j.future.2018.02.013]
52. Rahman S, Irfan M, Raza M, Moyeezullah Ghori K, Yaqoob S, Awais M. Performance analysis of boosting classifiers in recognizing activities of daily living. Int J Environ Res Public Health 2020 Feb 08;17(3):1082 [FREE Full text] [doi: 10.3390/ijerph17031082] [Medline: 32046302]
53. Zia S, Khan AN, Zaidi KS, Ali SE. Detection of generalized tonic clonic seizures and falls in unconstraint environment using smartphone accelerometer. IEEE Access 2021 Mar 4;9:39432-39443. [doi: 10.1109/ACCESS.2021.3063765]
54. Slade P, Troutman R, Kochenderfer MJ, Collins SH, Delp SL. Rapid energy expenditure estimation for ankle assisted and inclined loaded walking. J Neuroeng Rehabil 2019 Jun 06;16(1):67 [FREE Full text] [doi: 10.1186/s12984-019-0535-7] [Medline: 31171003]
55. Pei D, Olikkal P, Adali T, Vinjamuri R. Reconstructing synergy-based hand grasp kinematics from electroencephalographic signals. Sensors (Basel) 2022 Jul 18;22(14):5349 [FREE Full text] [doi: 10.3390/s22145349] [Medline: 35891029]
56. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H. How accurate is your activity tracker? A comparative study of step counts in low-intensity physical activities. JMIR Mhealth Uhealth 2017 Aug 11;5(8):e106 [FREE Full text] [doi: 10.2196/mhealth.6321] [Medline: 28801304]
57. Muaaz M, Chelli A, Abdelgawwad AA, Mallofre AC, Patzold M. WiWeHAR: multimodal human activity recognition using Wi-Fi and wearable sensing modalities. IEEE Access 2020 Sep 7;8:164453-164470. [doi: 10.1109/ACCESS.2020.3022287]

XSL•FO
RenderX

58. Davis K, Owusu E, Bastani V, Marcenaro L, Hu J, Regazzoni C, et al. Activity recognition based on inertial sensors for ambient assisted living. In: Proceedings of the 19th International Conference on Information Fusion. 2016 Presented at: FUSION '16; July 5-8, 2016; Heidelberg, Germany p. 371-378.
59. Papageorgiou XS, Chalvatzaki G, Tzafestas CS, Maragos P. Hidden markov modeling of human pathological gait using laser range finder for an assisted living intelligent robotic walker. In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2015 Presented at: IROS '15; September 28-October 2, 2015; Hamburg, Germany p. 6342-6347. [doi: 10.1109/iros.2015.7354283]
60. Mukhiddinov M, Cho J. Smart glass system using deep learning for the blind and visually impaired. Electronics 2021 Nov 11;10(22):2756. [doi: 10.3390/electronics10222756]
61. Seifert AK, Amin MG, Zoubir AM. Toward unobtrusive in-home gait analysis based on radar micro-doppler signatures. IEEE Trans Biomed Eng 2019 Sep;66(9):2629-2640. [doi: 10.1109/TBME.2019.2893528] [Medline: 30668460]
62. Chalvatzaki GG, Pavlakos G, Maninis K, Papageorgiou XS, Pitsikalis V, Tzafestas CS, et al. Towards an intelligent robotic walker for assisted living using multimodal sensorial data. In: Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies. 2014 Presented at: MOBIHEALTH '14; November 3-5, 2014; Athens, Greece p. 156-159. [doi: 10.1109/MOBIHEALTH.2014.7015934]
63. Ligons FM, Mello-Thoms C, Handler SM, Romagnoli KM, Hochheiser H. Assessing the impact of cognitive impairment on the usability of an electronic medication delivery unit in an assisted living population. Int J Med Inform 2014 Nov;83(11):841-848 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.07.004] [Medline: 25153770]
64. Churchill R, Lorence D, Richards M. Advanced data capture in the assisted medical home: a model for distributed and multimedia technologies. J Med Syst 2010 Aug;34(4):685-693. [doi: 10.1007/s10916-009-9282-9] [Medline: 20703923]
65. Knight A, Fouyaxis J, Jarrad G, Beski K, Cho G, Bidargaddi N. Systems to harness digital footprint to elucidate and facilitate ageing in place. Stud Health Technol Inform 2018;246:91-101. [Medline: 29507262]
66. Serbedzija N. Adaptive assistance: smart home nursing. In: Proceedings of the 2nd International Conference on Wireless Mobile Communication and Healthcare. 2012 Presented at: MOBIHEALTH '11; October 5-7, 2011; Kos Island, Greece p. 240-247. [doi: 10.1007/978-3-642-29734-2_33]
67. Eren C, Karamzadeh S, Kartal M. Signal processing techniques for human vital signs sensing by short range radar. In: Proceedings of the 28th Signal Processing and Communications Applications Conference. 2020 Presented at: SIU '20; October 5-7, 2020; Gaziantep, Turkey p. 1-4. [doi: 10.1109/siu49456.2020.9302095]
68. Siriwardhana C, Madhuranga D, Madushan R, Gunasekera K. Classification of activities of daily living based on depth sequences and audio. In: Proceedings of the 14th Conference on Industrial and Information Systems. 2019 Presented at: ICIIS '19; December 18-20, 2019; Kandy, Sri Lanka p. 278-283. [doi: 10.1109/iciis47346.2019.9063306]
69. Gopal P, Gesta A, Mohebbi A. A systematic study on electromyography-based hand gesture recognition for assistive robots using deep learning and machine learning models. Sensors (Basel) 2022 May 11;22(10):3650 [FREE Full text] [doi: 10.3390/s22103650] [Medline: 35632058]
70. Saeed U, Shah SY, Shah SA, Ahmad J, Alotaibi AA, Althobaiti T, et al. Discrete human activity recognition and fall detection by combining FMCW RADAR data of heterogeneous environments for independent assistive living. Electronics 2021 Sep 12;10(18):2237. [doi: 10.3390/electronics10182237]
71. Ejupi A, Galang C, Aziz O, Park EJ, Robinovitch S. Accuracy of a wavelet-based fall detection approach using an accelerometer and a barometric pressure sensor. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2017 Presented at: EMBC '17; July 11-15, 2017; Jeju, South Korea p. 2150-2153. [doi: 10.1109/EMBC.2017.8037280]
72. Forkan AR, Khalil I. A probabilistic model for early prediction of abnormal clinical events using vital sign correlations in home-based monitoring. In: Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communications. 2016 Presented at: PerCom '16; March 14-19, 2016; Sydney, Australia p. 1-9. [doi: 10.1109/percom.2016.7456519]
73. Feng J, Zhang SW, Chen L, Alzheimer's Disease Neuroimaging Initiative (ADNI). Identification of Alzheimer's disease based on wavelet transformation energy feature of the structural MRI image and NN classifier. Artif Intell Med 2020 Aug;108:101940. [doi: 10.1016/j.artmed.2020.101940] [Medline: 32972667]
74. Tsirmpas C, Anastasiou A, Bountris P, Koutsouris D. A new method for profile generation in an internet of things environment: an application in ambient-assisted living. IEEE Internet Things J 2015 Dec;2(6):471-478. [doi: 10.1109/jiot.2015.2428307]
75. Mojarad R, Attal F, Chibani A, Amirat Y. A hybrid context-aware framework to detect abnormal human daily living behavior. In: Proceedings of the 2020 International Joint Conference on Neural Networks. 2020 Presented at: IJCNN '20; July 19-24, 2020; Glasgow, UK p. 1-8. [doi: 10.1109/ijcnn48605.2020.9206930]
76. Patel AD, Shah JH. Performance analysis of supervised machine learning algorithms to recognize human activity in ambient assisted living environment. In: Proceedings of the IEEE 16th India Council International Conference. 2019 Presented at: INDICON '19; December 13-15, 2019; Rajkot, India p. 1-4. [doi: 10.1109/INDICON47234.2019.9030353]

XSL•FO
RenderX

77. Hervás R, Garcia-Lillo A, Bravo J. Mobile augmented reality based on the semantic web applied to ambient assisted living. In: Proceedings of the 3rd International Workshop on Ambient Assisted Living. 2011 Presented at: IWAAL '11; June 8-10, 2011; Torremolinos-Málaga, Spain p. 17-24. [doi: 10.1007/978-3-642-21303-8_3]
78. Kunnappilly A, Marinescu R, Seceleanu C. Assuring intelligent ambient assisted living solutions by statistical model checking. In: Proceedings of the 8th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation. 2018 Presented at: ISoLA '18; November 5-9, 2018; Limassol, Cyprus p. 457-476. [doi: 10.1007/978-3-030-03421-4_29]
79. Mayer C, Morandell M, Gira M, Hackbarth K, Petzold M, Fagel S. AALuis, a user interface layer that brings device independence to users of AAL systems. In: Proceedings of the 13th International Conference on Computers Helping People with Special Needs. 2012 Presented at: ICCHP '12; July 11-13, 2012; Linz, Austria p. 650-657. [doi: 10.1007/978-3-642-31522-0_98]
80. García-Magariño I. Defining and transforming models of Parkinson patients in the development of assisted-living multi-agent systems with INGENIAS. In: Proceedings of the International Workshops of Highlights on Practical Applications of Agents and Multi-Agent Systems. 2013 Presented at: PAAMS '13; May 22-24, 2013; Salamanca, Spain p. 460-471. [doi: 10.1007/978-3-642-38061-7_43]
81. Ngankam HK, Pigot H, Giroux S. OntoDomus: a semantic model for ambient assisted living system based on smart homes. Electronics 2022 Apr 05;11(7):1143. [doi: 10.3390/electronics11071143]
82. Gavril AF, Trascau M, Mocanu I. Multimodal interface for ambient assisted living. In: Proceedings of the 21st International Conference on Control Systems and Computer Science. 2017 Presented at: CSCS '17; May 29-31, 2017; Bucharest, Romania p. 223-230. [doi: 10.1109/CSCS.2017.38]
83. Aman F, Vacher M, Rossato S, Portet F. Speech recognition of aged voice in the AAL context: detection of distress sentences. In: Proceedings of the 7th Conference on Speech Technology and Human - Computer Dialogue. 2013 Presented at: SpeD '13; October 16-19, 2013; Cluj-Napoca, Romania p. 1-8. [doi: 10.1109/SpeD.2013.6682669]
84. Fuentes-Fernández R, Gómez-Sanz JJ. Development of simulations for ambient assisted living through pattern repositories. In: Proceedings of the 2017 Federated Conference on Computer Science and Information Systems. 2017 Presented at: FedCSIS '17; September 3-6, 2017; Prague, Czech Republic p. 299-305. [doi: 10.15439/2017F272]
85. Pang D, Gao Y, Liao L, Ying X. Brain functional network alterations caused by a strong desire to void in healthy adults: a graph theory analysis study. Neurourol Urodyn 2020 Sep;39(7):1966-1976. [doi: 10.1002/nau.24445] [Medline: 32806881]
86. Sarni T, Pulli P. Activity support framework for people with dementia based on Markov decision process. In: Proceedings of the 2015 International Conference on Intelligent Environments. 2015 Presented at: IE '15; July 15-17, 2015; Prague, Czech Republic p. 25-32. [doi: 10.1109/IE.2015.12]
87. Bahram M, Hossein-Zadeh GA. Functional parcellations affect the network measures in graph analysis of resting-state fMRI. In: Proceedings of the 21th Iranian Conference on Biomedical Engineering. 2014 Presented at: ICBME '14; November 26-28, 2014; Tehran, Iran p. 263-268. [doi: 10.1109/ICBME.2014.7043933]
88. Shi J, Liu B. Stage detection of mild cognitive impairment via fMRI using Hilbert Huang transform based classification framework. Med Phys 2020 Jul;47(7):2902-2915. [doi: 10.1002/mp.14183] [Medline: 32302413]
89. Ortiz A, Munilla J, Górriz JM, Ramírez J. Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. Int J Neural Syst 2016 Dec;26(7):1650025. [doi: 10.1142/S0129065716500258] [Medline: 27478060]
90. Hsieh TH, Sun MJ, Liang SF. Diagnosis of schizophrenia patients based on brain network complexity analysis of resting-state fMRI. In: Proceedings of the 15th International Conference on Biomedical Engineering. 2014 Presented at: ICBME '13; December 4-7, 2013; Singapore, Singapore p. 203-206. [doi: 10.1007/978-3-319-02913-9_52]
91. Liu J, Li M, Lan W, Wu FX, Pan Y, Wang J. Classification of Alzheimer's disease using whole brain hierarchical network. IEEE/ACM Trans Comput Biol Bioinform 2018;15(2):624-632. [doi: 10.1109/TCBB.2016.2635144] [Medline: 28114031]
92. Osório E, Ferreira L, Abreu R, Sousa F. Interoperability in ambient assisted living using OpenEHR. In: Proceedings of the IEEE 15th International Conference on e-Health Networking, Applications and Services. 2013 Presented at: Healthcom '13; October 9-12, 2013; Lisbon, Portugal p. 394-398. [doi: 10.1109/HealthCom.2013.6720707]
93. He H, Tan Y, Huang J. Unsupervised classification of smartphone activities signals using wavelet packet transform and half-cosine fuzzy clustering. In: Proceedings of the 2017 IEEE International Conference on Fuzzy Systems. 2017 Presented at: FUZZ-IEEE '17; July 9-12, 2017; Naples, Italy p. 1-6. [doi: 10.1109/FUZZ-IEEE.2017.8015566]
94. Mowafey S, Gardner S. Towards ambient intelligence in assisted living: the creation of an Intelligent Home Care. In: Proceedings of the 2013 Science and Information Conference. 2013 Presented at: SAI '13; October 7-9, 2013; London, UK p. 51-60.
95. Jethanandani M, Perumal T, Liaw YC, Chang JR, Sharma A, Bao Y. Binary relevance model for activity recognition in home environment using ambient sensors. In: Proceedings of the 2019 IEEE International Conference on Consumer Electronics - Taiwan. 2019 Presented at: ICCE-TW '19; May 20-22, 2019; Yilan, Taiwan p. 1-2. [doi: 10.1109/ICCE-TW46550.2019.8991837]
96. Yapeng L, Qin Y, Chen X, Li W. Exploring alterations of brain networks of AD patients using WTC method. IEEE Intell Syst (forthcoming) 2014 Jul 11:1-1. [doi: 10.1109/MIS.2014.37]

XSL•FO
RenderX

97. Merkle N, Zander S. Agent-based assistance in ambient assisted living through reinforcement learning and semantic technologies. In: Proceedings of the Confederated International Conferences On the Move to Meaningful Internet Systems. 2017 Presented at: OTM '17; October 23-27, 2017; Rhodes, Greece p. 180-188. [doi: 10.1007/978-3-319-69459-7_12]
98. Boger J, Jackson P, Mulvenna M, Sixsmith J, Sixsmith A, Mihailidis A, et al. Principles for fostering the transdisciplinary development of assistive technologies. Disabil Rehabil Assist Technol 2017 Jul;12(5):480-490. [doi: 10.3109/17483107.2016.1151953] [Medline: 27052793]
99. Hassan M, Umar M, Bermak A, Ali AA, Amira A. Computationally efficient environmental monitoring with electronic nose: a potential technology for ambient assisted living. In: Proceedings of the 2016 IEEE International Symposium on Systems Engineering. 2016 Presented at: ISSE '16; October 3-5, 2016; Edinburgh, UK p. 1-4. [doi: 10.1109/SysEng.2016.7753122]
100. Liu S, Zeng Z, Zhang YD, Fan T, Shan T, Tao R. Automatic human fall detection in fractional Fourier domain for assisted living. In: Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. 2016 Presented at: ICASSP '16; March 20-25, 2016; Shanghai, China p. 799-803. [doi: 10.1109/ICASSP.2016.7471785]
101. Kazeminejad A, Golbabaei S, Soltanian-Zadeh H. Graph theoretical metrics and machine learning for diagnosis of Parkinson's disease using rs-fMRI. In: Proceedings of the 2017 Artificial Intelligence and Signal Processing Conference. 2017 Presented at: AISP '17; October 25-27, 2017; Shiraz, Iran p. 134-139. [doi: 10.1109/AISP.2017.8324124]
102. Espinilla M, Liu J, García-Chamizo JM. Special issue on “Recent Advancements in Ubiquitous Computing”. J Ambient Intell Human Comput 2017 Mar 22;8(4):467-468. [doi: 10.1007/s12652-017-0476-6]
103. Figueroa A, Peralta B, Nicolis O. Coming to grips with age prediction on imbalanced multimodal community question answering data. Information 2021 Jan 21;12(2):48. [doi: 10.3390/info12020048]
104. Gopinath DE, Argall BD. Active intent disambiguation for shared control robots. IEEE Trans Neural Syst Rehabil Eng 2020 Jun;28(6):1497-1506. [doi: 10.1109/TNSRE.2020.2987878] [Medline: 32305928]
105. Gibellini F, Higler S, Lucas J, Luli M, Stallmann M, Dotti D, et al. Towards approximating personality cues through simple daily activities. In: Proceedings of the 20th International Conference on Advanced Concepts for Intelligent Vision Systems. 2020 Presented at: ACIVS '20; February 10–14, 2020; Auckland, New Zealand p. 192-204. [doi: 10.1007/978-3-030-40605-9_17]
106. Muaaz M, Chelli A, Pätzold M. WiHAR: from Wi-Fi channel state information to unobtrusive human activity recognition. In: Proceedings of the IEEE 91st Vehicular Technology Conference. 2020 Presented at: VTC '20; May 25-28, 2020; Antwerp, Belgium p. 1-7. [doi: 10.1109/VTC2020-Spring48590.2020.9128418]
107. da Silva Lisboa Tigre MF, Santos GL, Lynn T, Sadok D, Kelner J, Endo PT. Modeling the availability of an e-health system integrated with edge, fog and cloud infrastructures. In: Proceedings of the 2018 IEEE Symposium on Computers and Communications. 2018 Presented at: ISCC '18; June 25-28, 2018; Natal, Brazil p. 416-421. [doi: 10.1109/ISCC.2018.8538589]
108. Farràs-Permanyer L, Mancho-Fora N, Montalà-Flaquer M, Gudayol-Ferré E, Gallardo-Moreno GB, Zarabozo-Hurtado D, et al. Estimation of brain functional connectivity in patients with mild cognitive impairment. Brain Sci 2019 Dec 30;9(12):350 [FREE Full text] [doi: 10.3390/brainsci9120350] [Medline: 31801260]
109. Kuo HS, Sung PH, Lee SC, Kuan WT, Wang JF. Auditory-based robust speech recognition system for ambient assisted living in smart home. In: Proceedings of the 2014 International Conference on Orange Technologies. 2014 Presented at: ICOT '14; September 20-23, 2014; Xi'an, China p. 169-172. [doi: 10.1109/ICOT.2014.6956626]
110. Xiao M, Ge H, Khundrakpam BS, Xu J, Bezgin G, Leng Y, et al. Attention performance measured by attention network test is correlated with global and regional efficiency of structural brain networks. Front Behav Neurosci 2016 Oct 10;10:194 [FREE Full text] [doi: 10.3389/fnbeh.2016.00194] [Medline: 27777556]
111. Kim Y, An J, Lee M, Lee Y. An activity-embedding approach for next-activity prediction in a multi-user smart space. In: Proceedings of the 2017 IEEE International Conference on Smart Computing. 2017 Presented at: SMARTCOMP '17; May 29-31, 2017; Hong Kong, China p. 1-6. [doi: 10.1109/SMARTCOMP.2017.7946985]
112. Valenti AP, Chita-Tegmark M, Tickle-Degnen L, Bock AW, Scheutz MJ. Using topic modeling to infer the emotional state of people living with Parkinson's disease. Assist Technol 2021 May 04;33(3):136-145 [FREE Full text] [doi: 10.1080/10400435.2019.1623342] [Medline: 31194649]
113. Wang Q, Wang X. Proceedings of the 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference. 2020 Presented at: UEMCON '20; October 28-31, 2020; New York, NY, USA p. 558-561. [doi: 10.1109/UEMCON51285.2020.9298181]
114. Marini M, Viganò M, Corbo M, Zettin M, Simoncini G, Fattori B, et al. IDEA: an Italian dysarthric speech database. In: Proceedings of the 2021 IEEE Spoken Language Technology Workshop. 2021 Presented at: SLT '21; January 19-22, 2021; Shenzhen, China p. 1086-1093. [doi: 10.1109/SLT48900.2021.9383467]
115. Pavão SL, Lima CR, Rocha NA. Association between sensory processing and activity performance in children with cerebral palsy levels I-II on the gross motor function classification system. Braz J Phys Ther 2021;25(2):194-202 [FREE Full text] [doi: 10.1016/j.bjpt.2020.05.007] [Medline: 32540329]
116. Tyagi A, Patel A, Shah P. Text like classification of skeletal sequences for human action recognition. In: Proceedings of the 5th Asian Conference on Pattern Recognition. 2019 Presented at: ACPR '19; November 26–29, 2019; Auckland, New Zealand p. 335-350. [doi: 10.1007/978-3-030-41299-9_26]

XSL•FO
RenderX

117. Ehrenmüller I, Hasenauer R, Belviso C. Social assistive robots for elderly care: the new efficiency in the context of triple bottom line and digitization. In: Proceedings of the 2019 Portland International Conference on Management of Engineering and Technology. 2019 Presented at: PICMET '19; August 25-29, 2019; Portland, OR, USA p. 1-14. [doi: 10.23919/PICMET.2019.8893926]
118. Khazaee A, Ebrahimzadeh A, Babajani-Feremi A. Automatic classification of Alzheimer's disease with resting-state fMRI and graph theory. In: Proceedings of the 21th Iranian Conference on Biomedical Engineering. 2014 Presented at: ICBME '14; November 26-28, 2014; Tehran, Iran p. 252-257. [doi: 10.1109/ICBME.2014.7043931]
119. Demiris G, Thompson HJ. Mobilizing older adults: harnessing the potential of smart home technologies. contribution of the IMIA working group on smart homes and ambient assisted living. Yearb Med Inform 2012;7:94-99. [Medline: 22890348]
120. Stanford J, Peach B. Scene detection using convolutional neural networks. In: Proceedings of the 2nd IET International Conference on Technologies for Active and Assisted Living. 2016 Presented at: TechAAL '16; October 24-25, 2016; London, UK p. 1-6. [doi: 10.1049/ic.2016.0062]
121. Zhu C, Sheng W. Realtime human daily activity recognition through fusion of motion and location data. In: Proceedings of the 2010 IEEE International Conference on Information and Automation. 2010 Presented at: ICINFA '10; June 20-23, 2010; Harbin, China p. 846-851. [doi: 10.1109/ICINFA.2010.5512451]
122. Lu J, Guo S, Chen M, Wang W, Yang H, Guo D, et al. Generate the scale-free brain music from BOLD signals. Medicine (Baltimore) 2018 Jan;97(2):e9628 [FREE Full text] [doi: 10.1097/MD.0000000000009628] [Medline: 29480872]
123. Kyrarini M, Zheng Q, Haseeb MA, Graser A. Robot learning of assistive manipulation tasks by demonstration via head gesture-based interface. IEEE Int Conf Rehabil Robot 2019 Jun;2019:1139-1146. [doi: 10.1109/ICORR.2019.8779379] [Medline: 31374783]
124. Li N, Crane M, Ruskin HJ. Visual experience for recognising human activities. In: Proceedings of the International Competitions and Final Workshop on Evaluating AAL Systems Through Competitive Benchmarking. 2012 Presented at: EvAAL '12; September 24, 2012; Eindhoven, The Netherlands p. 173-185. [doi: 10.1007/978-3-642-37419-7_14]
125. Alnaeb A, Abdullah RS, Salah AA, Sali A, Rashid NE, Pasya I. Detection and classification real-time of fall events from the daily activities of human using forward scattering radar. In: Proceedings of the 20th International Radar Symposium. 2019 Presented at: IRS '19; June 26-28, 2019; Ulm, Germany p. 1-10. [doi: 10.23919/IRS.2019.8768130]
126. Alharbi F, Ouarbya L, Ward JA. Comparing sampling strategies for tackling imbalanced data in human activity recognition. Sensors (Basel) 2022 Mar 11;22(4):1373 [FREE Full text] [doi: 10.3390/s22041373] [Medline: 35214275]
127. Eldib M, Philips W, Aghajan H. Discovering human activities from binary data in smart homes. Sensors (Basel) 2020 May 29;20(9):2513 [FREE Full text] [doi: 10.3390/s20092513] [Medline: 32365545]
128. Costa N, Domingues P, Fdez-Riverola F, Pereira A. A mobile Virtual Butler to bridge the gap between users and ambient assisted living: a Smart Home case study. Sensors (Basel) 2014 Aug 06;14(8):14302-14329 [FREE Full text] [doi: 10.3390/s140814302] [Medline: 25102342]
129. GeethaRamani R, Sivaselvi K. Data mining technique for identification of diagnostic biomarker to predict Schizophrenia disorder. In: Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research. 2014 Presented at: ICCIC '14; December 18-20, 2014; Coimbatore, India p. 1-8. [doi: 10.1109/ICCIC.2014.7238525]
130. Zhang C, Gruver WA. Distributed agent system for behavior pattern recognition. In: Proceedings of the 2010 International Conference on Machine Learning and Cybernetics. 2010 Presented at: ICMLC '10; July 11-14, 2010; Qingdao, China p. 204-209. [doi: 10.1109/ICMLC.2010.5581067]
131. Lee YT, Chiu CS, Jhang LH, Santiago C. A self-reliance assistive tool for disable people. In: Proceedings of the 3rd International Conference on Control and Robotics Engineering. 2018 Presented at: ICCRE '18; April 20-23, 2018; Nagoya, Japan p. 26-30. [doi: 10.1109/ICCRE.2018.8376428]
132. Amin N, Gross T, Rosenthal S, Borschbach M. Impact of microphone orientation and distance on BSS quality within interaction devices. In: Proceedings of the 2nd IET International Conference on Intelligent Signal Processing. 2015 Presented at: ISP '15; December 1-2, 2015; London, UK p. 1-6. [doi: 10.1049/cp.2015.1787]
133. Pisotta I, Perruchoud D, Ionta S. Hand-in-hand advances in biomedical engineering and sensorimotor restoration. J Neurosci Methods 2015 May 15;246:22-29. [doi: 10.1016/j.jneumeth.2015.03.003] [Medline: 25769276]
134. Leong K, Leung C, Miao C, Chen YC. Detection of anomalies in activity patterns of lone occupants from electricity usage data. In: Proceedings of the 2016 IEEE Congress on Evolutionary Computation. 2016 Presented at: CEC '16; July 24-29, 2016; Vancouver, Canada p. 1361-1369. [doi: 10.1109/CEC.2016.7743947]
135. Freeman C, Exell T, Meadmore K, Hallewell E, Hughes AM. Computational models of upper-limb motion during functional reaching tasks for application in FES-based stroke rehabilitation. Biomed Tech (Berl) 2015 Jul;60(3):179-191 [FREE Full text] [doi: 10.1515/bmt-2014-0011] [Medline: 25355246]
136. Alchihabi A, Kivilicim BB, Ekmekci O, Newman SD, Vural FT. Decoding cognitive subtasks of complex problem solving using fMRI signals. In: Proceedings of the 26th Signal Processing and Communications Applications Conference. 2018 Presented at: SIU '18; May 2-5, 2018; Izmir, Turkey p. 1-4. [doi: 10.1109/SIU.2018.8404666]
137. Teachasrisaksakul K, Zhang ZQ, Yang GZ, Lo B. Imitation of dynamic walking with BSN for humanoid robot. IEEE J Biomed Health Inform 2015 May;19(3):794-802. [doi: 10.1109/JBHI.2015.2425221] [Medline: 25935051]

XSL•FO
RenderX

138. Yue L, Gong X, Li J, Ji H, Li M, Nandi AK. Hierarchical feature extraction for early Alzheimer's disease diagnosis. IEEE Access 2019 Jul 3;7:93752-93760. [doi: 10.1109/ACCESS.2019.2926288]
139. Kabary IA, Büchler M, Schuldt H. TOUCHify: bringing pen-based touch screen functionality to flat panel display screens. In: Proceedings of the 2013 International Conference on Information Society. 2013 Presented at: i-Society '13; June 24-26, 2013; Toronto, Canada p. 111-116.
140. Jiang B, Xu F, Huang Y, Yang C, Huang W, Xia J. Adaptive adversarial latent space for novelty detection. IEEE Access 2020 Nov 11;8:205088-205098. [doi: 10.1109/ACCESS.2020.3037346]
141. Anastou AC, Delibasis KK, Boulogeorgos AA, Sandalidis HG, Vavoulas A, Tasoulis SK. A low complexity indoor visible light positioning method. IEEE Access 2021 Apr 9;9:57658-57673. [doi: 10.1109/ACCESS.2021.3072348]
142. Hiraoka A, Murakami T. A standing-up assist control method considering movement speed. In: Proceedings of the IECON 46th Annual Conference of the IEEE Industrial Electronics Society. 2020 Presented at: IECON '20; October 18-21, 2020; Singapore, Singapore p. 71-76. [doi: 10.1109/IECON43393.2020.9254919]
143. Ghadi YY, Waheed M, Gochoo M, Alsuhibany SA, Chelloug SA, Jalal A, et al. A graph-based approach to recognizing complex human object interactions in sequential data. Appl Sci 2022 May 20;12(10):5196. [doi: 10.3390/app12105196]
144. Zsiga K, Tóth A, Pilissy T, Péter O, Dénes Z, Fazekas G. Evaluation of a companion robot based on field tests with single older adults in their homes. Assist Technol 2018;30(5):259-266. [doi: 10.1080/10400435.2017.1322158] [Medline: 28628395]
145. Clapés A, Pardo À, Pujol Vila O, Escalera S. Action detection fusing multiple Kinects and a WIMU: an application to in-home assistive technology for the elderly. Mach Vision Appl 2018 May 3;29(5):765-788. [doi: 10.1007/s00138-018-0931-1]
146. Nweke HF, Teh YW, Al-garadi MA, Alo UR. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. Expert Syst Appl 2018 Sep;105:233-261 [FREE Full text] [doi: 10.1016/j.eswa.2018.03.056]
147. Vilone G, Longo L. Notions of explainability and evaluation approaches for explainable artificial intelligence. Inf Fusion 2021 Dec;76:89-106. [doi: 10.1016/j.inffus.2021.05.009]
148. Dragoni M, Donadello I, Eccher C. Explainable AI meets persuasiveness: translating reasoning results into behavioral change advice. Artif Intell Med 2020 May;105:101840. [doi: 10.1016/j.artmed.2020.101840] [Medline: 32505427]
149. Singh A, Rehman SU, Yongchareon S, Chong PH. Sensor technologies for fall detection systems: a review. IEEE Sensors J 2020 Jul 1;20(13):6889-6919. [doi: 10.1109/JSEN.2020.2976554]
150. Demrozi F, Pravadelli G, Bihorac A, Rashidi P. Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey. IEEE Access 2020;8:210816-210836 [FREE Full text] [doi: 10.1109/access.2020.3037715] [Medline: 33344100]
151. Chelberg GR, Neuhaus M, Mothershaw A, Mahoney R, Caffery LJ. Mobile apps for dementia awareness, support, and prevention - review and evaluation. Disabil Rehabil 2022 Aug;44(17):4909-4920. [doi: 10.1080/09638288.2021.1914755] [Medline: 34034601]
152. Ramanujam E, Perumal T, Padmavathi S. Human activity recognition with smartphone and wearable sensors using deep learning techniques: a review. IEEE Sensors J 2021 Jun 15;21(12):13029-13040. [doi: 10.1109/jsen.2021.3069927]
153. Van der Roest HG, Wenborn J, Pastink C, Dröes RM, Orrell M. Assistive technology for memory support in dementia. Cochrane Database Syst Rev 2017 Jun 11;6(6):CD009627 [FREE Full text] [doi: 10.1002/14651858.CD009627.pub2] [Medline: 28602027]
154. Burns SP, Terblanche M, Perea J, Lillard H, DeLaPena C, Grinage N, et al. mHealth intervention applications for adults living with the effects of stroke: a scoping review. Arch Rehabil Res Clin Transl 2021 Mar;3(1):100095 [FREE Full text] [doi: 10.1016/j.arrct.2020.100095] [Medline: 33778470]
155. Masina F, Orso V, Pluchino P, Dainese G, Volpato S, Nelini C, et al. Investigating the accessibility of voice assistants with impaired users: mixed methods study. J Med Internet Res 2020 Sep 25;22(9):e18431 [FREE Full text] [doi: 10.2196/18431] [Medline: 32975525]
156. Jain V, Khurana N, Bhardwaj S. A review on human behavior using machine learning for ambient assisted living. In: Proceedings of 3rd International Conference on Computing Informatics and Networks. 2020 Presented at: ICCIN '20; July 29-30, 2020; New Delhi, India p. 333-345. [doi: 10.1007/978-981-15-9712-1_28]
157. Xu Z, Wang G, Guo X. Comparative studies on activity recognition of elderly people living alone. In: Proceedings of 2019 Chinese Intelligent Systems Conference. 2019 Presented at: CISC '19; October 26-27, 2019; Haikou, China p. 276-291. [doi: 10.1007/978-981-32-9682-4_29]

Abbreviations

AAL: ambient assisted living
ADL: activities of daily living
AI: artificial intelligence
DL: deep learning
HAR: human activity recognition
IADL: instrumental activities of daily living

XSL•FO
RenderX

ML: machine learning
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RGBD: red green blue depth
RQ: research question

Edited by G Eysenbach; submitted 17.01.22; peer-reviewed by JY Choi, J Fan; comments to author 05.07.22; revised version received 15.08.22; accepted 23.09.22; published 04.11.22.

Please cite as:
Jovanovic M, Mitrov G, Zdravevski E, Lameski P, Colantonio S, Kampel M, Tellioglu H, Florez-Revuelta F
Ambient Assisted Living: Scoping Review of Artificial Intelligence Models, Domains, Technology, and Concerns
J Med Internet Res 2022;24(11):e36553
URL: https://www.jmir.org/2022/11/e36553
doi:10.2196/36553
PMID:36331530

©Mladjan Jovanovic, Goran Mitrov, Eftim Zdravevski, Petre Lameski, Sara Colantonio, Martin Kampel, Hilda Tellioglu, Francisco Florez-Revuelta. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Examining Analytic Practices in Latent Dirichlet Allocation Within Psychological Science: Scoping Review

Lauryn J Hagg[1], BAppSc, GDipPsych; Stephanie S Merkouris[1], PhD; Gypsy A O'Dea[1], BPsych (Hons); Lauren M Francis[1], BPsych (Hons); Christopher J Greenwood[1,2,3], PhD; Matthew Fuller-Tyszkiewicz[1], PhD; Elizabeth M Westrupp[1,4], PhD; Jacqui A Macdonald[1,2,3], PhD; George J Youssef[1,2], PhD

[1]Centre for Social and Early Emotional Development (SEED), School of Psychology, Deakin University, Geelong, Australia

[2]Centre for Adolescent Health, Murdoch Children's Research Institute, Melbourne, Australia

[3]Department of Paediatrics, University of Melbourne, Melbourne, Australia

[4]Judith Lumley Centre, La Trobe University, Melbourne, Australia

Corresponding Author:
Lauryn J Hagg, BAppSc, GDipPsych
Centre for Social and Early Emotional Development (SEED)
School of Psychology
Deakin University
1 Gheringhap St
Geelong, 3220
Australia
Phone: 61 9251 7344
Email: lauryn.hagg@research.deakin.edu.au

Abstract

Background: Topic modeling approaches allow researchers to analyze and represent written texts. One of the commonly used approaches in psychology is latent Dirichlet allocation (LDA), which is used for rapidly synthesizing patterns of text within "big data," but outputs can be sensitive to decisions made during the analytic pipeline and may not be suitable for certain scenarios such as short texts, and we highlight resources for alternative approaches. This review focuses on the complex analytical practices specific to LDA, which existing practical guides for training LDA models have not addressed.

Objective: This scoping review used key analytical steps (data selection, data preprocessing, and data analysis) as a framework to understand the methodological approaches being used in psychology research using LDA.

Methods: A total of 4 psychology and health databases were searched. Studies were included if they used LDA to analyze written words and focused on a psychological construct or issue. The data charting processes were constructed and employed based on common data selection, preprocessing, and data analysis steps.

Results: A total of 68 studies were included. These studies explored a range of research areas and mostly sourced their data from social media platforms. Although some studies reported on preprocessing and data analysis steps taken, most studies did not provide sufficient detail for reproducibility. Furthermore, the debate surrounding the necessity of certain preprocessing and data analysis steps is revealed.

Conclusions: Our findings highlight the growing use of LDA in psychological science. However, there is a need to improve analytical reporting standards and identify comprehensive and evidence-based best practice recommendations. To work toward this, we developed an LDA Preferred Reporting Checklist that will allow for consistent documentation of LDA analytic decisions and reproducible research outcomes.

(J Med Internet Res 2022;24(11):e33166) doi:10.2196/33166

KEYWORDS

latent Dirichlet allocation; LDA; review; analysis; methodology

XSL•FO
RenderX

Introduction

Background

The past 25 years have seen an enormous increase in the availability of so called "big data," a broad term describing very large, but typically unstructured data sets [1]. One example of big data is textual data, which describes any source of data that contains written words or words that are transcribed from speech. The big data era [1] has seen increasing availability of large textual data sets derived from a variety of sources including web-based forums (eg, Reddit), social microblogging platforms (eg, Twitter, Facebook, and Instagram), formal documentation (eg, discharge summaries and clinical notes), qualitative data sets, Google Books, and scientific literature. Big data sets have been used in a variety of research areas such as travel [2], digital humanities [3], and marketing [4]. Given that textual data sets may provide important insights into trends and associations relating to human behavior and attitudes, it is not surprising that the use of these data sets is increasing in the psychological sciences.

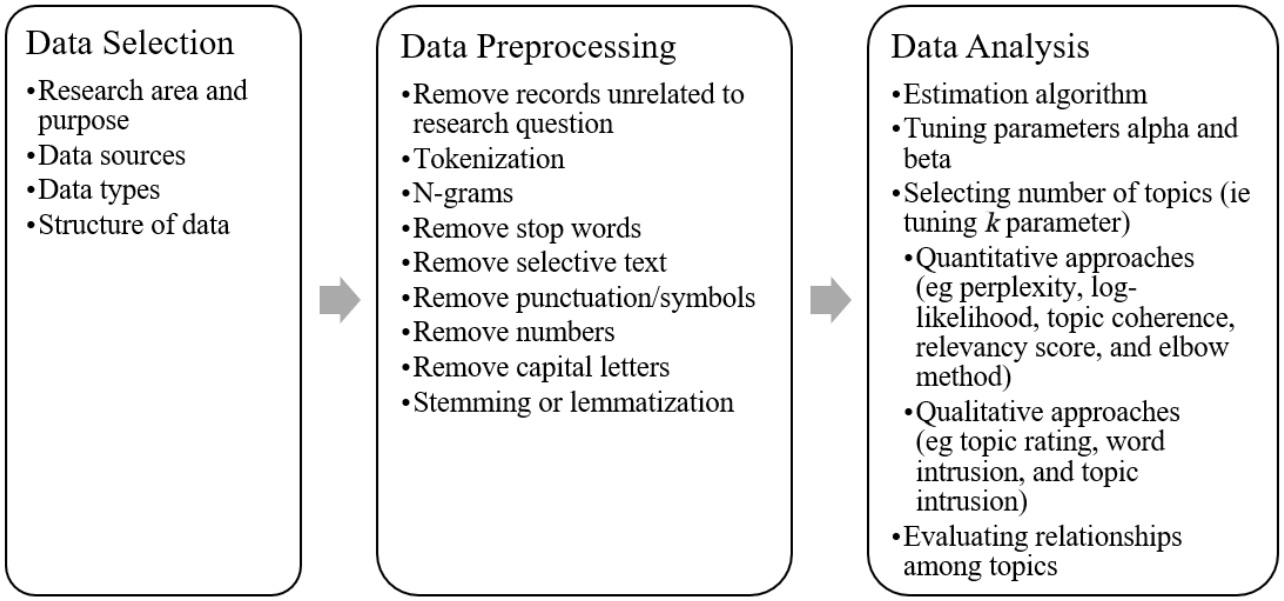
Considering the potential size and complexity of big textual data sets, psychology researchers have begun to rely on natural language processing (NLP) techniques. These computational methods are used to analyze and represent written text [5,6]. Topic modeling approaches are largely automated and allow researchers to effectively and efficiently engage with big textual data sets in ways that cannot be practically achieved with nonautomated techniques for synthesizing (ie, literature reviews) and analyzing (ie, qualitative approaches) textual data.

There are a range of topic modeling approaches available [7]; for example, latent semantic analysis is a nonprobabilistic method that can be used to draw meaning from textual data [8], and Dirichlet multinomial mixture–based methods may perform better for smaller texts [9]. However, one commonly used NLP technique used in health research is latent Dirichlet allocation (LDA), which is a machine learning methodology that uses Bayesian probability–based algorithms to discover latent (unobserved) "topics" based on co-occurrence of words from within a body of text (ie, corpus). Although detailed explanations of these algorithms can be found in the studies by Blei et al [10] and Griffiths and Steyvers [11], in simple terms, LDA identifies latent topics within a corpus by estimating both *document-topic probabilities* (ie, the probability that each document is generated by any specific topic) and *word-topic probabilities* (ie, the probability that any word is generated by a specific topic; [12,13]). LDA assumes that documents comprise many latent topics and that latent topics comprise many words [12]. Briefly, the LDA algorithm first requires the user to specify the number of latent topics (*k*) expected within the corpus. Initially, the algorithm iterates through each document (ie, unit of text) and words within the document and randomly assigns the words to one of the latent topics. This results in a distribution of document-topic probabilities (ie, the probability of the words in any document assigned to each of the *k* topics) and word-topic probabilities (ie, the proportion of times a word has been assigned to each of the *k* topics) based on random allocation. This random allocation is then optimized by iterating through each document and words within the documents, recalculating the probability of a word belonging to a topic given a particular document, and then updating the word-topic probabilities across all documents. In addition to the number of topics (*k*), the LDA algorithm is influenced by 2 other parameters (also known as hyperparameters) that can be specified by the researcher and affects how topics are represented across documents and by words. Alpha influences how documents contribute to topics, with larger alpha values resulting in documents comprising many topics (ie, smaller alpha values suggest that documents comprise a small number of topics; [14]). Beta (also known as delta) influences how words create topics, with large values resulting in topics represented by a greater number of words (ie, smaller beta values suggest topics will be represented by fewer words; [14]). Once the LDA model is optimized, analysts can examine both the words and documents that are most probabilistically related to each topic to derive topic meaning and understanding of the larger textual data set.

As implied in the brief explanation above, training an LDA model is a complex task that involves decision-making and consideration of multiple factors that have the potential to influence the outcomes of the analysis. Several practical guides have been published [14-17] that broadly outline several different ways to approach LDA, using a variety of packages. Broadly, training an LDA model involves 3 major steps: *data selection*, *data preprocessing*, and *data analysis* (Figure 1). However, these are not prescriptive, and individual applications of LDA may involve iterations of these steps.

XSL•FO
RenderX

Figure 1. Summary of latent Dirichlet allocation (LDA) data selection, preprocessing, and analysis steps. Note: Tokenization is a required preprocessing step that ensures that the data are appropriately structured for analysis. All other preprocessing steps are optional.



Data Selection

The analyst must first make decisions regarding the textual data to be analyzed. The 4 major decisions in this step include determining (1) the research area and the purpose of the research being conducted, (2) the source of textual data, (3) the data types within these sources used for analysis, and (4) how data will be structured for analysis. Specifically, the research area and purpose of the research influences decisions made about the source of textual data (eg, social media, formal documentation, and scientific literature), the data types within that source that will be used for analysis (eg, original posts, comments, paragraphs, sentences, words, and other specific sections of text), and how these data will be structured (eg, by post, by user, by citation, and by paragraph) into documents (ie, units of text) for analysis.

Data Preprocessing

Once a data set has been identified, the second major step involves preprocessing the text for analysis. Preprocessing is the process of preparing the data with the aim of increasing fidelity so that the results are meaningfully representative of the data [15,18] and relevant to the research question. Textual data sets have the potential to contain a substantial amount of noise and irrelevant textual information [18]. As outlined in numerous sources [15-17], textual data may require a range of general preprocessing steps depending on the research question. These may include, for example, converting to lower case, replacing entities (eg, people, places, and numbers) with placeholder using named entity recognition, and removal of punctuation and symbols, numbers, selective text that minimally contributes toward research questions and varies among studies, and stop words that are words thought to add no meaning to the data (eg, “and,” “it,” and “to”; [19]) and can be implemented using various stop word lists [20,21]. Furthermore, 2 processes of transforming words include stemming (ie, shortening words to a similar root form, without needing to have meaning; eg, “explore,” “exploratory,” and “exploration” into “explor”) and lemmatization (ie, transforming words to a canonical [lemma] form; eg, “explore,” “exploratory,” and “exploration” into “explore” [16]). Notably, although some research suggests using stemming or lemmatization cautiously because of the potential impact on results [16], the necessity of using this preprocessing step has also been called into question [22]. Finally, other preprocessing steps are undertaken to describe the way data are used in the analysis. Specifically, tokenization is when words are broken down into n-grams denoting single words (unigrams) or a series of words that are presented in the same order (2 words=bigram; 3 words=trigram [16]). Tokenization and n-grams are advantageous for disambiguating meaning in the context of surrounding words. For example, grouping “cognitive,” “behavioral,” “therapy” as a trigram allows researchers to observe how this construct contributes to a topic rather than how the individual words do.

Data Analysis

Following preprocessing, the LDA analysis is typically conducted as the third step. There are 4 decision-making points during this step, including (1) the LDA estimation algorithm (eg, sampling approaches based on Markov Chain Monte Carlo [23,24], such as Gibbs sampling [11], and optimization approaches based on variational Bayes (VB) approximations [23,24], such as the variational EM algorithm [10]); (2) tuning parameters such as the alpha parameter [25], which influences how documents contribute to topics [14], and less importantly the beta parameter [25], which influences how words create topics [14]; (3) tuning the *k* parameter, that is, the process of selecting the number of latent topics that represent the data set, which can be done using quantitative (eg, perplexity [10], log-likelihood [14], topic coherence [26], relevancy score [27], and elbow method that is used to visually identify the optimal number of topics when plotting the results of quantitative metrics [28]) or qualitative approaches (eg, topic rating [29], word intrusion [30], and topic intrusion [30]); and (4) the process of evaluating relationships among topics.

XSL•FO
RenderX

LDA is a burgeoning approach with an increasing number of studies published in the psychological sciences. Several practical guides on LDA exist providing high-level advice, but they are inconsistent and not comprehensive. Therefore, the next steps in this research are to evaluate how LDA is being conducted by researchers in psychology and how this compares to synthesized advice from the existing guides, informing the development of best practice guidelines. Our aim was to conduct a scoping review to describe the methodological practices used in studies using LDA throughout the psychological literature. Scoping reviews focus on examining the nature of research activity and can be used specifically to survey how methodological approaches are implemented within an area of research [31-33]. Thus, a scoping review is particularly well-suited to examining the methodological practices of studies using LDA in psychology. Calvo et al [34] and Shatte et al [35] have previously conducted scoping reviews on broader machine learning techniques. Although these reviews examined the mental health literature and described different sources of textual data, they did not focus on the analytical decisions that were specific to LDA. This scoping review focuses on the key steps of *data selection*, *data preprocessing*, and *data analysis* as a framework to understand the methodological approaches being used in psychology research using LDA.

Methods

Transparency and Openness

This scoping review adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews; [36]) and reports on search strategy, eligibility criteria, and data charting processes detailed in the following sections. This study was not preregistered.

Search Strategy

Four electronic databases were searched using the following search strategy: “latent dirichlet” OR “topic* model*” OR “latent topic*.” MEDLINE Complete, CINAHL Complete, and EMBASE were searched up to April 15, 2020, with searches limited to the English language and research based on humans, with a peer-review limiter also applied to CINAHL Complete. PsycINFO was searched up to April 30, 2020, with English language and peer-review limiters applied.

Eligibility Criteria and Selection of Sources of Evidence

Following the recommended practices for conducting scoping reviews [32], we used an iterative, team-based approach to finalize inclusion and exclusion criteria. Studies were included if they (1) were published in English, (2) were published in a peer-reviewed journal, (3) used LDA to analyze textual data, and (4) focused on a psychological construct or issue (eg, mental health issues, substance use, gender differences, and social issues such as same-sex marriage and environmental issues). Studies were excluded if they (1) were a commentary, letter, thesis, conference abstract or slides, or a methods paper; (2) used data that were not written words or words transcribed from speech (eg, genetic codes, mental health codes, and information derived from images); and (3) focused on constructs or issues that were nonpsychological in nature (eg, medical [37-40], marketing [4], and humanities [3]).

Titles and abstracts of all records were reviewed independently by 3 investigators (LJH, LMF, and GAO). All full-text records were assessed by a single investigator (LJH). In addition, 10% (71/712) of the articles were independently screened at the full-text level by another reviewer (LF or GAO) as part of the iterative process for refining inclusion criteria in accordance with recommended practices for conducting scoping reviews [32]. Disagreements during title and abstract screening and full-text assessment were resolved through discussion and consensus agreement by the research team.

Data Charting Process, Data Items, and Synthesis of Results

A data charting (extraction) template based on common data selection, preprocessing, and data analysis steps was constructed and used to collate all relevant information from the included articles. The development of this data charting template was an iterative process that was continuously updated and refined during the data charting process.

In addition to study characteristics (ie, author, year, and journal of publication), the data charting process included the extraction of the (1) topic area (eg, mental health, depression, autism, self-harm, treatment, discrimination, and global climate) and purpose of research (ie, broadly what the study was aiming to achieve), (2) data sources (eg, social media, scientific literature, and formal documentation) and data types (eg, posts or comments, abstracts or titles, and selective words), (3) structure of the analyzed documents (eg, by user, post, patient, and citation), (4) data preprocessing steps conducted (eg, stop words, stemming, and lower casing), (5) LDA estimation algorithms used, (6) estimation parameters used, (7) relationships among topics, and (8) programs and packages used.

All charted data relating to study characteristics, topic area, purpose of research, data sources, and data types were tabulated according to the study, and all charted data relating to preprocessing and data analysis were tabulated according to the type of preprocessing step and methodological approach.

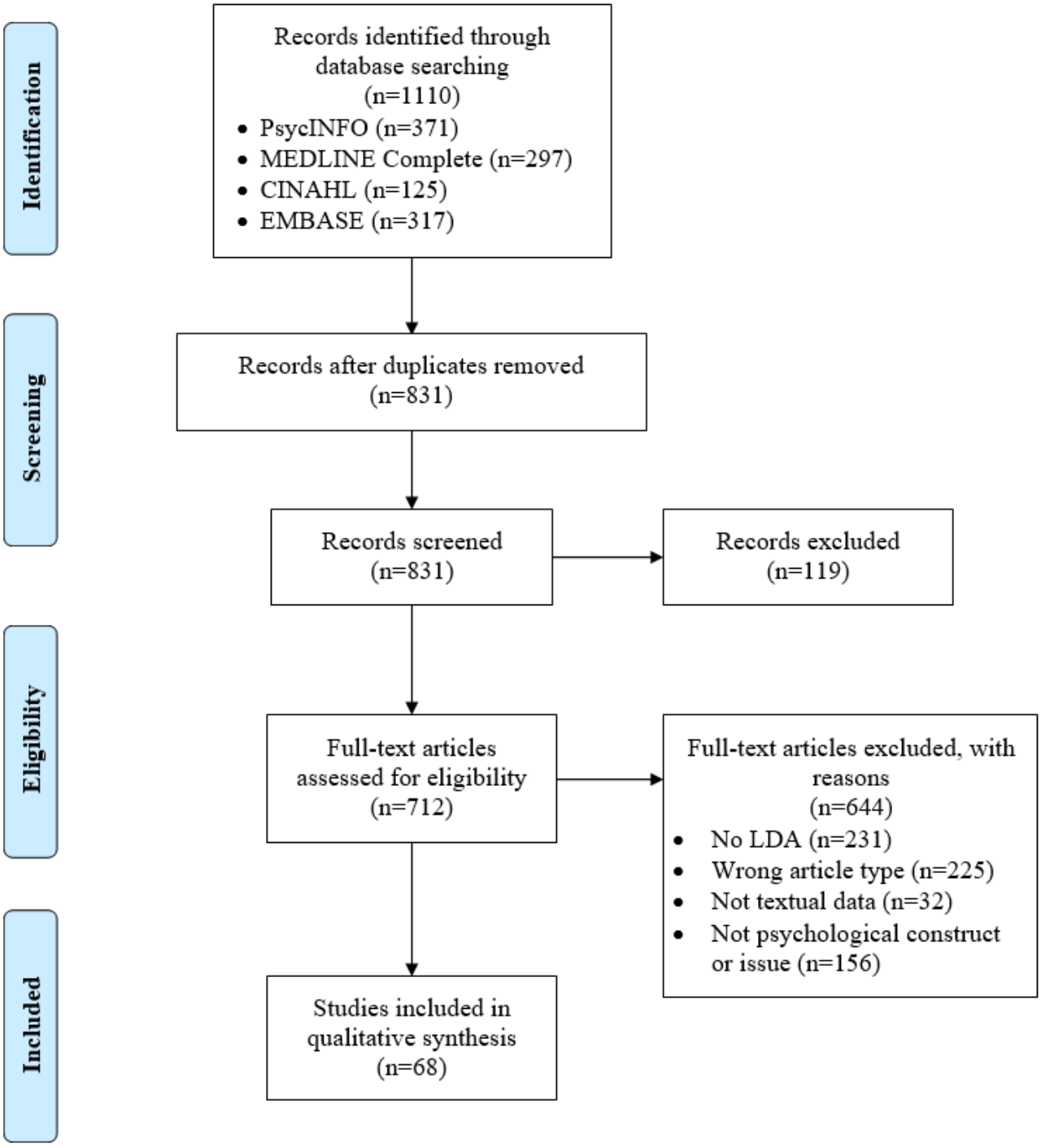
Results

Selection of Sources of Evidence

A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the systematic search results is shown in Figure 2 [41]. After removing duplicates (n=279), the search identified 831 articles for title and abstract screening. Of these, the full texts of 85.7% (712/831) potentially eligible articles were assessed, and 9.6% (68/712) of these articles were included in this scoping review.

XSL•FO
RenderX

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart detailing study inclusion and exclusion process [41]. LDA: latent Dirichlet allocation.



Characteristics of Sources of Evidence

Table 1 presents the characteristics of the included studies. The 68 studies that met the inclusion criteria were published between 2014 and 2020, with the application of LDA to psychological constructs increasing from 1 publication in 2014 to 11 in 2018 and 23 in 2019. A total of 13 articles were published in 2020 at the time of searching. Of the 55 different journals publishing these articles, the most frequent publication sources were the Journal of Medical Internet Research (7/68, 10%), PLOS One (3/68, 4%), and International Journal of Environmental Research and Public Health (3/68, 4%).

XSL•FO
RenderX

Table 1. Summary of study characteristics and data selection.

Author	Journal	Topic area	Purpose of research	Source of data	Data type nested within document level	Documents, n	Words per document (before or after preprocessing)
Abdellaoui et al [42]	*Journal of Medical Internet Research*	Substance use	Detect cases of noncompliance to drug treatment in patient forum posts	Social media; forum	Posts (escitalopram); post Posts (aripiprazole); post	Escitalopram=3649; aripiprazole=2164	NR[a]
Afshar et al [43]	*PLOS One*	Substance use	Identify subtypes in patients with opioid misuse	Formal documentation; clinical notes	Selective words; NR	NR	NR
Alam et al [44]	*Behaviour & Information Technology*	Social issues	Improve situational awareness of humanitarian organizations about disaster events	Social media; Twitter	Posts; NR	NR	NR
Barry et al [45]	*American Journal of Health Education*	Substance use	Examine advertising practices of alcohol brands	Social media; Twitter	Posts; NR	NR	NR
Bittermann and Fischer [46]	*Zeitschrift fur Psychologie*	Scientific topics	Identify hot topics in psychology	Scientific literature	Controlled keyword terms; citation	314,573	NR
Carpenter et al [47]	*Journal of Medical Internet Research*	Mental health	Assessing efficacy of internet well-being interventions	Social media; other—Happify	Free text response; task	NR	Mean 51.23 (before)
Carron-Arthur et al [48]	*BMC Psychiatry*	Mental health	Topics of discussion in mental health support groups	Social media; forum	Posts; post	131,004	Range 70-110 (after)
Chen et al [49]	*Journal of Medical Internet Research*	Substance use	Understanding electronic cigarette and hookah use	Social media; forum	Posts; NR	NR	NR
Choi and Seo [50]	*Issues in Mental Health Nursing*	Mental health	Provide an overview of depression of caregivers	Scientific literature	Abstracts; citation	426	NR
Choudhury et al [51]	*Strategic Management Journal*	Social issues	Investigate managerial cognitive capabilities and CEO[b] communication	Other: interview transcripts	Interview transcripts; response to interview question	69	Mean 8234 (before; SD 3458)
Cohan et al [52]	*Journal of the Association for Information Science & Technology*	Mental health	Determining mental health based on indications for self-harm ideation	Social media; forum	Posts; NR	NR	NR
Feldhege et al [53]	*Journal of Affective Disorders*	Mental health	Investigate topics in a web-based depression community	Social media; forum	Posts and comments; user	20,037	NR
Franz et al [54]	*Suicide and Life-Threatening Behavior*	Mental Health	Identify self-injurious thoughts and behaviors and related themes on the web	Social media; forum	Posts; post	2355	Mean 43.21 (before; SD 42.99)
Gerber [55]	*Decision Support Systems*	Forensic	Predicting crime	Social media; Twitter	Selective tweets; neighborhood	NR	NR
Giorgi et al [56]	*Organization Science*	Social issues	Examine relationship between films and their legal environment via a cultural contingency perspective	Formal documentation; congressional hearings and annual reports; other; newspaper articles	Annual reports, congressional hearings, and newspaper articles; annual report, congressional hearing, and newspaper article	Annual report=84; congressional hearing=25; newspaper article=950	NR
Guo et al [57]	*PLOS One*	Social issues	Map the topic landscape of social class an inequality	Scientific literature	Selective words in titles, keywords, and abstracts; NR	NR	NR

XSL•FO
RenderX

Author	Journal	Topic area	Purpose of research	Source of data	Data type nested within document level	Documents, n	Words per document (before or after preprocessing)
Hemmatian et al [58]	*Behavior Research Methods*	Social issues	Demonstrate how change in the framing of same-sex marriage in public discourse relates to changes in public opinion	Social media; forum	Selective comments; NR	NR	NR
Hwang et al [59]	*Journal of Medical Internet Research*	Mental health	Analyze behavior patterns of emotional eaters	Social media; forum	Posts and comments; NR	NR	NR
Jaworska and Nanda [60]	*Applied Linguistics*	Social issues	Examine thematic patterns and their changes over time of corporate social responsibility reports in the oil sector	Formal documentation; social responsibility reports	Reports; NR	NR	NR
Jung and Suh [61]	*Decision Support Systems*	Mental health	Identifying job satisfaction	Other; company review website	Reviews; NR	NR	NR
Kagashe et al [62]	*Journal of Medical Internet Research*	Substance use	Understanding the use of medicinal drugs during seasonal influenza	Social media; Twitter	Posts; post	459,043	NR
Karami et al [63]	*Psychology of Violence*	Social issues	Understand experiences of sexism and sexual harassment in the workplace	Social media; Forum	Posts; post	2362	NR
Kee et al [64]	*Mindfulness*	Mental health	Identify topics relevant to mindfulness research	Scientific literature	Titles and abstracts; NR	NR	NR
Kigerl [65]	*Social Science Computer Review*	Social issues	Further understand cybercrime carding forums	Social media: forum	Posts; user	30,469	NR
Kreitzberg et al [66]	*Addictive Behaviors*	Substance use	Examine tobacco promotion	Social media; Instagram	Posts; post	4629	NR
Landstrøm et al [67]	*Sexualities*	Social issues	Explore how norms for appropriate behavior between parents and children are constructed	Other; various webpages	Posts; NR	NR	NR
Lee et al [68]	*Evolution and Human Behavior*	Evolution	Investigate mating-relevant self-concepts and mate preference	Social media; other—web-based dating profiles	Written descriptions; profile	7973	Mean 69.65 (before; SD 106.83)
Lee et al [69]	*European Child and Adolescent Psychiatry*	Mental health	Identify characteristics of Korean student suicide	Formal documentation; teacher reports	Selective words; NR	NR	NR
Liang et al [70]	*Journal of Health Communication*	Physical health	Identify associations between regional prevalence of obesity and overweight and regional information and social environments	Social media; Twitter	Tweets; NR	NR	NR
Liu et al [71]	*International Journal of Medical Informatics*	Social issues	Investigate gender difference in web-based health communities	Social media; forum	Post; NR	NR	NR
Liu et al [72]	*Journal of Biomedical Informatics*	Mental health	Determine symptom-based patient subgroups in mental illness	Formal documentation; clinical notes	Selective words; patient	1746	NR
Liu et al [73]	*Psychology, Health & Medicine*	Scientific topics	Identify hot topics in published review articles in clinical psychology	Scientific literature	Titles and abstracts; NR	NR	NR

XSL•FO
RenderX

Author	Journal	Topic area	Purpose of research	Source of data	Data type nested within document level	Documents, n	Words per document (before or after preprocessing)
Liu et al [74]	*International Journal of Environmental Research and Public Health*	Emotions; mental health; physical health	Study differences in the emotions of patients with physiological and psychological diseases	Social media; forum	Posts; post	17,891	NR
Lou et al [75]	*Journal of Interactive Advertising*	Social issues	Investigate how influencer vs brand-promoted advertisements affect consumer engagement, sentiment, and topics of comment	Social media; Instagram	Advertisement; NR	NR	NR
Louvigné and Rubens [76]	*Behaviormetrika*	Education	Classification of goal-based messages	Social media; Twitter	Tweets; learning goal	NR	NR
Magua et al [77]	*Journal of Women's Health*	Social issues	Investigate disadvantages of being a woman in renewing grants	Formal documentation; summary statements	Summary statements; NR	NR	NR
McCoy [78]	*Psychosomatics: Journal of Consultation and Liaison Psychiatry*	Mental health	Map delirium literature	Scientific literature	Titles and abstracts; citation	3231	NR
Merrill and Åkerlund [79]	*Journal of Computer-Mediated Communication*	Social issues	Investigate how racism contributes to group discussion of immigration and how Facebook allows this	Social media; Facebook	Posts and comments; identical post	23,939	NR
Murdock et al [80]	*Cognition*	Development	Study exploration and exploitation trade-off	Other; nonfiction books	Books; NR	NR	NR
Oh et al [81]	*Journal of Counselling Psychology*	Scientific topics	Identify topics in *Journal of Counselling Psychology*	Scientific literature	Abstracts; NR	NR	NR
Pandrekar et al [82]	*American Medical Informatics Association annual symposium proceedings; American Medical Informatics Association symposium*	Substance use	Investigate opioid-related discussions	Social media; forum	Posts; NR	NR	NR
Pantti et al [83]	*European Journal of Communication*	Social issues	Investigate how racism is used in Finnish public debate	Social media: forum; other: news media content	Discussion forum content and news content; NR	NR	NR
Pappa et al [84]	*Journal of Medical Internet Research*	Physical health	Identifying factors associated with weight change	Social media; forum	Posts and comments; NR	NR	NR
Park and Conway [85]	*American Medical Informatics Association annual symposium proceedings; American Medical Informatics Association symposium*	Substance use; physical health	Track health-related discussions (ie, Ebola, e-cigarettes, influenza, and marijuana)	Social media; forum	Selective words from posts and comments; post	114,320,798	NR

XSL•FO
RenderX

Author	Journal	Topic area	Purpose of research	Source of data	Data type nested within document level	Documents, n	Words per document (before or after preprocessing)
Ray et al [86]	*Journal of Strategic Marketing*	Education	Explore values affecting behavioral intention in e-learning	Social media: Twitter; other: reviews	Review and tweets; review	Reviews=139,581; tweets=1442	NR
Ruiz et al [87]	*Attachment & Human Development*	Development	Investigate reflective functioning in fathers of children born preterm and at term	Other: survey data	Text response to 8 survey items; NR	NR	NR
Rumshisky et al [88]	*Translational Psychiatry*	Mental health	Predicting psychiatric readmission	Formal documentation; health records	Selective words; NR	NR	NR
Santos et al [89]	*Systems Research and Behavioural Science*	Social issues	Investigate the impact of social media and traditional media on democratic systems	Social media: Twitter; other: various webpages	Tweets and webpages; NR	NR	NR
Shahin and Dai [90]	*American Behavioral Scientist*	Social issues	Understand public engagement with global aid agencies	Social media; Twitter	Selective tweets; inbound data set	NR	NR
Shin et al [91]	*Frontiers in Psychology*	Education	Create distractor items	Other; open-source data set	Student responses; NR	NR	NR
Sieweke and Santoni [92]	*The Leadership Quarterly*	Social issues	Review research using natural experimental designs to infer causal relationships about leadership	Scientific literature	Abstracts; citation	1156	NR
Son et al [93]	*International Journal of Information Management*	Social issues	Investigate how Twitter's representational features influence average retweet time and how effects differed based on type of disaster communication	Social media; Twitter	Tweets; NR	NR	NR
Sorour et al [94]	*Journal of Educational Technology & Society*	Education	Predict student performance	Other; student feedback	Selective words in comments; NR	NR	NR
Sperandeo et al [95]	*Frontiers in Psychiatry*	Mental health; personality	Investigate nature of research regarding personality and mental health	Scientific literature	Abstracts; NR	NR	NR
Szekely and Vom Brocke [96]	*PLOS One*	Social issues	Derive propositions for research and practice from corporate sustainability reports	Formal documentation; sustainability reports	Reports; NR	NR	NR
Törnberg and Törnberg [97]	*Discourse & Society*	Social issues	Analyzing discursive connections between Islamophobia and antifeminism	Social media; forum	Posts; user	576,801	1000 (before)
Tran et al [98]	*International Journal of Environmental Research and Public Health*	Mental health	Understand artificial intelligence application in the management of depressive disorders	Scientific literature	Abstracts; citation	NR	NR
Tran et al [99]	*Complementary Therapies in Medicine*	Mental health	Map mind-body interventions to improve quality of life	Scientific literature	Abstracts; NR	NR	NR
Turrentine et al [100]	*Journal of the American College of Surgeons*	Social issues	Examine gender differences in surgical residency applicants; recommendation letters	Formal documentation; letters of recommendation	Letters of recommendation; letter	332	Mean 404 (after)

XSL•FO
RenderX

Author	Journal	Topic area	Purpose of research	Source of data	Data type nested within document level	Documents, n	Words per document (before or after preprocessing)
Wang et al [101]	*BMC Public Health*	Substance use; mental health	Identifying topics about adolescent substance use and depression	Scientific literature	Abstracts; NR	NR	NR
Weij et al [102]	*International Journal of Consumer Studies*	Social issues	Discussion of attention to contemporary protesting artists among Western audiences	Social media; Twitter	Tweets; NR	NR	NR
Westmaas et al [103]	*Nicotine & Tobacco Research*	Substance use	Determine context of discussions surrounding cessation treatment for cancer survivors who smoke	Social media; forum	Posts; post	3998	NR
Wu et al [104]	*Journal of Educational Technology & Society*	Education	Investigate learner interest in open learning environments	Social media; other—Learning Cell Knowledge Community	Learning cell; learner	3538	NR
Yoon [105]	*Journal of the American Psychiatric Nurses Association*	Mental health	Identifying mental health needs for people with dementia	Social media; Twitter	Tweets and retweets; NR	NR	NR
Zhan et al [106]	*Journal of Medical Internet Research*	Substance use	Understanding how consumers and policy makers use social media to track e-cigarette–related content	Social media; Twitter and forum	Posts; NR	NR	NR
Zhao et al [107]	*International Journal of Environmental Research and Public Health*	Disability	Understand how autism-affected users use support groups on Facebook	Social media; Facebook	Interactions and content from 5 Facebook groups; NR	NR	NR
Zheng and Shahin [108]	*Information, Communication & Society*	Social issues	Examine social media use in pollical campaigns	Social media; Twitter	Tweets; NR	NR	NR
Zou [109]	*Expert opinion on drug safety*	Substance use	Analyze trends on drug safety	Scientific literature	Titles and abstracts; NR	NR	NR

[a]NR: not reported.

[b]CEO: chief executive officer.

Data Selection

Research Area and Purpose

Table 1 shows that the most prominent areas of research were social issues (23/68, 34%; eg, racism, sexism, same-sex marriage, and global climate), mental health (19/68, 28%), and substance use (12/68, 26%). There was great variation among studies regarding the purpose of their research, which ranged from simply understanding behaviors (eg, e-cigarette and hookah use) and experiences (eg, sexism and sexual harassment) to assessing the efficacy of interventions (eg, internet well-being and mind-body interventions), identifying social discourse (eg, same-sex marriage, racism, and feminism), and analyzing trends (eg, drug safety).

Data Sources and Data Types

Table 1 highlights the key sources of the data used in LDA (Multimedia Appendix 1 [42-109] provides more details of data selection, data preprocessing, and data analysis) and the types of data used within these sources. The most common sources of data were social media platforms (35/68, 51%), which were most often derived from forums (eg, Reddit: 7/35, 20%) or microblogging platforms (eg, Twitter: 11/35, 31%; Facebook: 2/35, 6%; and Instagram: 2/35, 6%). Other social media sources included a knowledge community space (1/35, 3%) and web-based dating profiles (1/35, 3%). Studies typically sourced their data from one social media platform, with only 3% (1/35) of studies using multiple social media platforms as their source of data (ie, forum and Twitter). Of the studies that used data from forums and microblogging platforms, all indicated that they used some form of web-based posts (eg, original posts and comments) in their analyses. Some were explicit in that they specified the use of posts and comments or retweets (5/33, 15%), although some also included selective criterion (4/33, 12%; eg, selective comments containing negative and positive words or phrases [58] and selective words with specific term

XSL•FO
RenderX

frequency–inverse document frequency scores [88]). Most studies, however, simply mentioned the use of "posts" or "tweets," or "interactions online" or "discussion forum content" and did not describe their precise selection criteria (24/33, 73%).

Scientific literature was the next most common source of textual data (13/68, 19%), for which data were derived from searches of databases including Web of Science (5/13, 38%), MEDLINE (2/13, 15%), PubMed (2/13, 15%), and PSYINDEX (1/13, 8%). However, 23% (3/13) of the studies used scientific literature derived from specific journals. All studies using scientific literature specified the data used for analysis. Specifically, some studies only used data from abstracts (7/13, 54%), whereas others used data from titles and abstracts (4/13, 31%), controlled key terms (1/13, 8%), and selective words from titles, keywords, and abstracts (1/13, 8%).

Formal documentation was another common source of textual data (8/68, 12%), where data were derived from different forms of documentation such as sustainability, social responsibility, teacher reports (3/8, 37%), clinical notes (2/8, 25%), health records (1/8, 12%), summary statements (1/8, 12%), and letters of recommendation (1/8, 12%). These studies either used selective words from the documentation (4/8, 50%) or used the documentation in its entirety for analytic purposes (4/8, 50%).

Other uncategorized sources of textual data included nonfiction books (1/68, 1%), student feedback (1/68, 1%), survey data (1/68, 1%), interview transcripts (1/68, 1%), an open-source data set (1/68, 1%), a company review website (1/68, 1%), a web platform (1/68, 1%), and various webpages (1/68, 1%). The data types used in these studies are listed in Table 1.

Finally, although most studies used data from a single source, 6% (4/68) of the studies derived data from multiple sources. Of these, 75% (3/4) of the studies used data from social media microblogging platforms (eg, Twitter and forums) and other uncategorized sources including reviews, various webpages, and news media content. Moreover, of the 4 studies, 1 (25%) study used data from various formal documentation sources (eg, annual reports and congressional hearings) and an uncategorized source (newspaper articles).

Structure of Textual Data

Overall, 43% (29/68) of the studies reported how textual data were structured into documents for the purpose of analysis (Table 1). The remaining 57% (39/68) of the studies did not provide any methodological details on how the textual data were structured. Of the studies that reported on how they structured their data, those that derived data from social media commonly defined documents as individual posts (10/19, 53%) or a user's history of posts (3/19, 16%). Studies that derived data from the scientific literature defined each document as text from individual publications (5/5, 100%), and studies that used data derived from formal documentation structured their data by patient (1/3, 33%), letter (1/3, 33%), or annual report or congressional hearing (1/3, 33%). Overall, 35% (24/68) of the studies reported sample sizes (ie, number of documents, which ranged from 69 documents to 114,320,798 documents (Median 3998, IQR 2164-30469). Finally, 10% (7/68) of the studies reported the number of words (or average number of words or range of words) per document (Median 90, IQR 60.44-702), and of those that did, 2 studies reported this value after preprocessing.

Data Preprocessing

Overall, 86% (59/68) of the studies reported preprocessing their data. Table 2 highlights various preprocessing steps undertaken when preparing textual data for an LDA (Multimedia Appendix 1 describes preprocessing steps broken down by study). Specifically, the most frequently used steps included removing: stop words (46/59, 78%), punctuation, symbols or special characters (31/59, 53%), selective text (eg, hyperlinks, names, frequent words; 29/59, 49%), numbers (20/59, 34%), and invalid records (eg, records that do not provide relevant text; 17/59, 29%). Furthermore, 36% (21/59) of the studies undertook stemming or lemmatization, whereas 7% (4/59) studies explicitly stated that this step was not conducted [49,79,80,97]. Few studies reported conducting tokenization (15/59, 25%) and 15% (9/59) of the studies specified which n-grams were applied. Other preprocessing steps that were identified but less commonly used included removing capital letters, clearing whitespace, and correcting misspelled words (which can be conducted using automated spell checkers such as hunspell [110]). Overall, 10% (7/68) of the studies did not report data preprocessing, and 3% (2/68) of the studies indicated that data were preprocessed but provided no further details. Regarding the use of programs or packages for preprocessing data, 51% (35/68) of the studies did not comment on the tools used, 28% (19/68) highlighted the program or package used for all preprocessing undertaken, and 21% (14/68) specified the program or package for some preprocessing steps but not all (Multimedia Appendix 1).

XSL•FO
RenderX

Table 2. Summary of study engagement in data preprocessing, selection of *k*, and use of programs or packages.

Preprocessing steps (n[a])	Selection of *k* (n)	Program; LDA[b] package (n)
Stop words (46)	Quantitative approach (28)	Java; MALLET[c] (15)
Punctuation, symbols, special characters (31)	Perplexity (11); [10]	R; Topicmodels package (13)
Selective text (29)	Harmonic mean of model log-likelihoods (5); [11]	R; MALLET package (2)
Stemming or lemmatization (21)	Topic coherence (4); [26]	R; stm package (1)
Numbers (20)	Log-likelihood (3); [14]	R; maptpx package (1)
Invalid records (17)	Kullback-Leibler divergence (3); [111]	R; KoNLP[d] package (1)
Tokenization (15)	Jensen-Shannon divergence (3); [112]	R; dfrtopics package (1)
N-grams (9)	Exclusivity (1); [113]	R; LDA tuning package (1)
Unigrams (8)	Hierarchical Dirichlet process (HDP-LDA; 1); [114]	R; NR[e] (4)
Bigrams (5)	Log Bays factor (1); [115]	Python; Gensim package (7)
Trigrams (1)	Per-document topic distributions (1); [62]	Python; LDA package (1)
Lower casing (16)	Topic probability (1); [116]	Python; Natural Language Toolkit package (1)
Whitespace (7)	Observing average *F*-measure (1); [94]	Python; NR (2)
Spelling (5)	Optimal_k function (1); [117]	Stata (2)
Unclear (2)	Minimization fit metric (1); [118]	Big text Tool (1)
NR (7)	*t*-distributed stochastic neighbor embedding (1); [91]	MeCab (1)
N/A[f]	Qualitative approach (10)	NR (17)
N/A	Quantitative and qualitative approach (5)	N/A
N/A	Topic coherence (4)	N/A
N/A	Perplexity (1)	N/A
N/A	Specificity (1); [119]	N/A
N/A	Kullback-Leibler divergence (1)	N/A
N/A	Sample size (1); [73]	N/A
N/A	Jensen-Shannon divergence (1)	N/A
N/A	Unclear (1)	N/A
N/A	NR (24)	N/A

[a]n: number of studies. Further details and references are provided in Multimedia Appendix 1.

[b]LDA: latent Dirichlet allocation.

[c]MALLET: Machine Learning for Language Toolkit.

[d]KoNLP: Korean natural language processing.

[e]NR: not reported.

[f]N/A: not applicable.

Data Analysis

LDA Estimation Algorithms

As shown in Table 2, 75% (51/68) of the studies specified the program or package used to train the LDA model, with the most common implementation being Machine Learning for Language Toolkit (MALLET; 15/51, 29%), topic models in R (13/51, 25%), and Gensim in Python (7/51, 14%). Among the studies that used Gensim in Python, it was unclear whether Gensim's implementation of LDA or Gensim's LDA MALLET wrapper was used. Multimedia Appendix 1 provides the programs and packages used broken down by study.

Only 26% (18/68) of the studies explicitly reported the estimation algorithms used to train the LDA model (Multimedia Appendix 1). Most of these studies used a Gibbs sampling method (16/18, 89%). Overall, 74% (50/68) of the studies did not explicitly provide the estimation algorithms used. Of these 50 studies, 25 (50%) referred readers to algorithm-specific documentation (eg, the studies by Blei et al [10] for the variational EM algorithm and Griffiths and Steyvers [11] for Gibbs sampling), and 19 (38%) studies specified the programs and packages used for analysis, for which the default algorithms can be determined (eg, program or package documentation) and were likely used.

XSL•FO
RenderX

Selection of Alpha and Beta Parameters

Only 13% (9/68) of the studies (Multimedia Appendix 1) specified the selection of alpha and beta parameters. Specifically, the most consistently selected alpha parameters were 0.1 (3/9, 33%) and 50/*k* (3/9, 33%), and the most common beta parameter was 0.01 (5/9, 56%).

Selecting the Number of Topics (k Parameter)

An essential parameter that must be specified when training an LDA model is the number of topics. Table 2 highlights various approaches that have been applied to determine the optimal number of topics (Multimedia Appendix 1 provides an approach to determine the optimal number of topics broken down by study). Overall, the most common approaches were quantitative in nature (28/68, 41%). The most predominant approach was perplexity (11/28, 39%), which is a common method of evaluating model fit in LDA models [10,120], where models with lower perplexity are considered the best fitting. Another commonly used method for evaluating model fit was topic coherence (4/28, 14%), which allows for a comparison of topics by measuring the degree of semantic similarity among words that contribute the most to that topic [26]. Log-likelihood was also used (3/28, 11%), whereby the best-fitting model was considered to occur at the maximum log-likelihood value. These data suggest that perplexity and coherence remain popular approaches. Perplexity, which uses the log-likelihood, attempts to quantify how well an estimated model generalizes to a new data set. Although this is helpful for understanding the optimal number of topics in a data set, this approach can lead to uninterpretable topics; therefore, combining quantitative and qualitative measures should be used to assess the quality of the topics. Consequently, coherence metrics attempt to quantify the semantic relatedness of the words that are most strongly related to a topic. A model in which the *k* number of topics all have high coherence suggests that the topics will be more interpretable by researchers. Finally, a range of minimization and maximization fit metrics were used to determine the optimal number of topics (eg, harmonic mean of the model log-likelihoods, Kullback-Leibler divergence, and Jensen-Shannon divergence). A qualitative approach to determining the appropriate number of topics was used by 15% (10/68) of the studies, which involved using human judgment and researcher expertise to specify the number of topics. Furthermore, 7% (5/68) of the studies used a mixed methods approach to determine the optimal number of topics, and 1% (1/68) of studies suggested that LDA tuning was undertaken but did not specify how. Finally, 35% (24/68) of the studies did not report on how the optimal number of topics was determined.

Evaluating Relationships Among Topics

Another consideration when training an LDA model is evaluating the relationships or overlap among topics (Multimedia Appendix 1). Overall, 85% (58/68) of the studies did not report the relationships among topics, and 7% (4/58) of these studies acknowledged this as a limitation of their research. The remaining 15% (10/68) of the studies that reported relationships among topics did so using hierarchical clustering analyses (3/10, 30%) or other study-specific methods including visualization techniques (4/10, 40%; eg, LDAvis).

Discussion

Principal Findings

Our aim was to conduct a scoping review to describe the methodological practices used in LDA studies throughout the psychological literature. We focused on the steps of data *selection*, *data preprocessing*, and *data analysis* as a framework to understand the methodological approaches being used in psychology research that use LDA. The inclusion of 68 empirical studies, all of which were published since 2014, demonstrates that psychology researchers are adopting LDA to draw insights from big data sets; however, we identified considerable variability in the reporting of the steps outlined in the available practical guides, ranging from 10% for the number of words per document to 86% for any preprocessing.

Data Selection

Research Area and Purpose

The literature shows that the research areas evaluated using LDA included both narrow and broad foci. The areas of focus included behavioral, cognitive, and affective constructs, which can be categorized into the following research areas: mental health, social issues (eg, racism, sexism, same-sex marriage, and global climate), substance use, physical health, education, identification of scientific topics, human development (eg, exploratory behavior, and parenting), personality, emotions, forensics, disability, and evolution. Although the areas in which LDA has been applied fall within the range of research areas highlighted earlier, the purpose for which LDA is used in psychological research varies widely and includes understanding behaviors (eg, e-cigarette and hookah use) and concepts (eg, sexism), assessing the efficacy of interventions (eg, internet well-being and mind-body interventions), identifying social discourse (eg, same-sex marriage, racism, and feminism), and analyzing trends (eg, drug safety).

Data Sources, Data Types, and Structure of Data

The findings of this review demonstrate that the common sources of big data used in psychological LDA research are social media (eg, forums, Twitter, Facebook, and Instagram), scientific literature, and formal documentation (eg, reports, clinical notes, health records, summary statements, and letters of recommendation). Given that the content often examined in psychological research is of a sensitive nature (eg, mental health issues and personal experiences), it may be particularly relevant to consider the ethical implications of using publicly available data (eg, social media), which might be linked to a person's identity. We encourage researchers to consult ethics boards when determining whether approval is needed to use such data, even if it is publicly available [121,122]. Furthermore, social media data can be more prone to grammatical errors and increased ambiguity (eg, owing to spelling errors and slang) compared with scientific literature and formal documentation and may require more in-depth preprocessing depending on the nature of the research question. Where required, social media data can be preprocessed using packages such as TweetTokenizer from the Natural Language Tool Kit [123]. Despite the potential challenges associated with social media

XSL•FO
RenderX

data, most included studies (35/68, 51%) used social media data and were more likely to report the structure of textual data, and the length of included documents, compared with studies using scientific literature, formal documentation, and other uncategorized sources of textual data. However, the scientific literature was slightly more likely to report the sample size.

The results also demonstrate that LDA provides researchers with unique flexibility in selecting the type of textual data that can best answer their research questions. The selection of textual data for analysis plays an influential role in analysis outcomes; therefore, it is imperative that authors clearly specify their data inclusion and exclusion criteria to ensure reproducibility. For instance, researchers can use "original posts" alone, to obtain a broad overview of topics within a forum or group, or "original posts" plus the subsequent comments, which allows for the analysis of topics in discourse. Although all studies specified the type of data used for analysis, most studies that used social media data did not describe their precise data selection criteria and simply mentioned the use of "posts" or "interactions online." Taken together, the literature demonstrates that more transparency is needed in reporting practices.

This review identified that less than half of the included studies (29/68, 43%) reported how textual data are structured into documents (ie, units of text). This is an extension of data-type selection decisions, as it is important to consider that the same set of selected data could be structured in multiple ways. This underreporting of document structures can have a potentially important influence on contextualizing results [16,124]. For example, the decision to use titles and abstracts as the set of data for analysis answers different research questions if documents are structured according to a citation or journal. Consequently, not reporting document structure clouds interpretation of any topics that have been derived. Furthermore, only a small number of studies reported sample size (ie, number of documents) and the length of the included documents. This minimal reporting may be linked to inconsistent evidence regarding the optimal sample size and length of documents for LDA. For instance, some evidence argues for a larger number of documents, as it may be theoretically impossible to identify meaningful topics from a smaller number of documents; however, it also suggests that there is a threshold whereby increasing the number will not affect the performance of the LDA [124]. Others indicate that the sample size is dependent upon theoretical and methodological considerations related to the research question [16]. In addition, documents that are too long or too short can produce results that are difficult to interpret [124]. In the context of short pieces of textual data (eg, Twitter posts), LDA may not perform well, as this approach assumes that there are multiple topics per document. Qiang et al [9] reviewed a range of alternative methods for the modeling of short text documents, which are more likely to comprise a single topic or have a lower ability to find co-occurrence patterns, although there is some evidence that LDA may also perform adequately with such texts [125]. Furthermore, Mehrotra et al [126] and Ito et al [127] identified that pooling textual data, and therefore making documents longer, leads to improved LDA topic models. In contrast, Sbalchiero et al [128] highlighted the potential effects of different length texts on results and complexities associated with topic modeling in long texts, which warrants further investigation. At this time, it is suggested that the best way to determine the appropriate length of a document is to observe the optimal model fit for samples of different text lengths [128] but to use other approaches such as qualitative or, as discussed, other NLP methods (see the study by Qiang et al [9] for a review of methods for analyzing short texts and a GitHub resource that supports the comparison of different algorithms for short text documents) when dealing with smaller texts. Given that the structure of textual data into documents, sample size, and document length may influence the LDA, it is important that researchers training an LDA model clearly report this information and that future empirical studies investigate how these factors may affect results.

Data Preprocessing

In contrast to the suggested practices in existing guides, studies do not routinely report on data preprocessing steps, with 13% (9/68) of studies not reporting this. Given that preprocessing steps work to increase the fidelity of data to ensure that results are meaningfully representative of the data, this underreporting is problematic as it may influence analyses and compromise the interpretability and subsequent conclusions [129]. Studies that reported preprocessing of data typically conducted a common set of processes including removing stop words, selective text (eg, hyperlinks, names, and frequent words), punctuation or symbols, invalid records, and numbers, and conducting stemming or lemmatization. Furthermore, few studies have clearly reported the use of tokenization and n-grams; however, some studies have highlighted the use of tokenization but did not specify the n-grams applied. The overall scarce reporting of tokenization and n-grams even more so highlights that the focus of researchers has been on reporting preprocessing steps that aim to increase data fidelity (eg, stop words, punctuation or symbols, and numbers), and less so on reporting preprocessing steps that describe how data are organized for analysis (eg, tokenization and n-grams). A need for transparency surrounding the presentation of data is demonstrated by literature that suggests the suitability of both unigrams and bigrams [16]; however, methodological studies have suggested that bigrams may not improve categorization into topics [130]. This indicates the need for further research exploring best practices for preprocessing steps that describe how data are presented for analysis.

Although a number of studies chose to conduct stemming or lemmatization, some explicitly stated that to facilitate topic interpretation, this step was not conducted [49,79,80,97]. This is consistent with the findings of Yang et al [131], which suggest that although topic models with and without stemming provide similar results, the stemmed results may be more difficult to interpret. Similarly, other studies have suggested that stemming or lemmatization provides no meaningful improvement to the quantitative measures of model fit and has the potential to reduce topic stability [132]. Despite methodological studies erring toward not engaging in stemming or lemmatization [132], a number of studies in the psychological sciences continue to engage in this practice. We recommend that future studies reflect the necessity of stemming, given the existing evidence. In addition, research may evaluate the effects of different types of

XSL•FO
RenderX

stemming or lemmatization [132,133] on the results. Future research should consider reporting results with and without stemming or lemmatization to demonstrate the potential effects on results, which can be used to inform best practice recommendations.

Data Analysis

LDA Programs and Packages, LDA Estimation Algorithms, Selecting Alpha and Beta Parameters, and Selecting Number of Topics (k Parameter)

Although results revealed that many programs or packages were used to train the LDA model, among the most commonly used were Java, R, and Python. The open-source nature of each of these programs emphasizes that LDA is an accessible analysis type for researchers in psychology. As such, we recommend that these open-source programs continue to be used in practice; however, the different estimation algorithms used in each program should be considered.

The results indicated that Gibbs sampling was the most commonly used estimation algorithm. However, the selection of estimation algorithms is underreported (ie, reported by only 18/68, 26% studies), which may reflect a lack of understanding about the potential implications of selecting these algorithms. Although there are some conflicting methodological studies investigating these estimation algorithms (eg, see VB algorithms for evidence of appropriateness [134-136]), Gibbs sampling appears to be a generally robust approach as defined by better prediction of the optimal number of topics [11,137], as well as strong performance even when compared with newer algorithms [29]. Although decisions surrounding which estimation algorithms to use are often guided by practicality related to ease of implementation in analysis programs (ie, availability in widely used statistical packages), we suggest that the wide availability of Gibbs sampling within packages makes this approach a strong contender for use in psychological studies.

Although estimation algorithms are underreported, by mentioning the programs and packages used, it is possible for the reader to assume that the default algorithms highlighted in the associated documentation were likely used; however, packages often change default settings, and therefore, package and version numbers should be documented. Furthermore, although the literature has highlighted that programming languages provide default implementations of LDA [14], there is evidence suggesting that tuning of the alpha (but not beta) parameter is an important consideration [25]. Of the studies that specified alpha and beta, 78% (7/9) of studies overrode defaults and specifically tuned alpha (as 0.1 and 50/*k*) and beta (as 0.01).

A parameter that is tuned consistently throughout the literature is the *k* parameter, which is the selection of the number of topics derived from the model [138]. Throughout the psychological literature, it is evident that approaches used to determine the number of topics shift between qualitative and quantitative methodologies, which is reflective of inconsistencies in practical guides, where some advocate for the use of quantitative approaches (eg, perplexity, log-likelihood, and topic coherence; [14]), which can be conducted in multiple ways (eg, [139]), whereas others suggest using qualitative approaches (eg, human judgment and expertise [16]). Quantitative approaches are beneficial, as they can be faster, systematic, and can be validated using cross-validation [15], which is the process of randomly splitting data into portions and training the model on all but one of those portions and then validating the model on the remaining portion. Although qualitative approaches are more time consuming, they too can also be systematic and cross-validated. In addition, research has demonstrated that quantitative methods do not replace human judgment when deciding a model's interpretability and that qualitative methods allow researchers to explore textual data in ways that model fit statistics do not [30]. Some human judgment approaches include *topic rating* that refers to viewing a topic and assigning a quality score [29], *word intrusion* that is the qualitative process of identifying out-of-place words within a topic to understand a topic's coherence [30], and *topic intrusion* that evaluates a topic model's distribution of documents into topics compared with human judgment of a document's content [30]. There are benefits and drawbacks associated with these 2 different methods of determining number of topics, and Asmussen et al [15] posited that as akin to factor analytic models where interpretability of factors is as important as statistical model fit, the number of topics should be determined by a balance between a usable number of topics and appropriate model fit. Moving beyond topic modeling alone, the literature has begun to analyze textual data sets by conducting qualitative coding and comparing these results to topic models [54]. Considering the conflicting literature, it is interesting to note that very few studies in psychology have used a combination of these techniques [48,56,58,73,75]. Overall, there are various ways of determining the number of topics, and although several different authors have proposed recommended approaches [29,140,141], this is an area of ongoing research, as recommended approaches do not necessarily converge on the same value for *k* selected.

Evaluating Relationships Among Topics

The results indicate that evaluating the relationships among topics is not a common practice in LDA studies conducted in the psychological sciences. Specifically, evaluating the relationships among topics involves observing the overlap among topics and understanding how topics are similar or different. One of the ways this can be achieved is by visualizing topics using tools such as LDAvis in R [27] and pyLDAvis in Python [142]. Increased evaluation of the relationships among topics will allow for richer findings and the potential to identify unexpected links among topics.

Limitations

This is the first study to evaluate the decision-making processes in psychological research studies that use LDA, thus providing researchers in this space with an introduction to some of the key considerations when training an LDA model. The findings from this review should be considered in light of certain limitations. First, the points of decision-making within the analytic pipeline discussed in this review should be considered by all researchers; however, there are other points of decision-making that fall within *data selection*, *data preprocessing*, and *data analysis* that were not included in this review, as they are discretionary depending on the research

XSL•FO
RenderX

question. For example, stratified analyses by potential theoretical or methodological moderators can help identify whether there is consistency in latent topics identified across the strata [16], but the use of such moderators is dependent upon the research question being asked. In addition, researchers may find it useful to develop specific inclusion and exclusion criteria and extract data in a way that is driven by clearly developed working definitions. For example, researchers may develop dictionaries of words that can be used to identify relevant content, which are carefully constructed based on theoretical and expert opinions to reflect important aspects of the constructs of interest for a study [16]. However, it is important to consider that this may not always be appropriate because, for example, social media users may not use the same language as experts; therefore, the extracted data may not be representative. A data-driven approach may be useful in that it can capture a greater breadth of data; however, this can be time consuming. Second, of the studies that did not provide methodological details on how textual data were structured into documents (ie, units of text), inferences could be made for some of these studies based on the language used throughout the article. This may be considered a limitation, as this information was not included in the interpretation of results; however, we argue that this is an illustration of the primary issues surrounding the lack of reporting within this literature. Third, this review focused on mapping the literature rather than appraising its quality; therefore, it is important to note that the intensity of engagement with the 3 steps discussed throughout this review does not necessarily reflect the quality or accuracy of the results as they relate to the constructs under investigation. Fourth, this review only included studies that applied LDA to a construct or issue; therefore, studies providing insights into the LDA methodology have not been reviewed. Fifth, this review specifically focused on traditional applications of LDA rather than modifications thereof, as these are increasingly being used in psychology research. Although the LDA used by studies in this review was unsupervised, a supervised LDA approach [143] may be useful, particularly if the aim of the research is prediction. The supervised LDA permits the user to label each document with known properties that can be used for model fitting. Jacobucci et al [144] provided a recent example of supervised LDA, where they included information on whether the author of each document used in their model had a known history of suicide risk. The study by Šperková [145] provides further information about variations of LDA (eg, sentiment LDA and factorial LDA). Finally, this review focuses on one topic modeling approach rather than an overview of multiple topic modeling approaches. When conducting topic modeling, we encourage researchers to consider the suitability of other approaches; the study by Terragni et al [7] provides further information about other topic modeling approaches (eg, latent semantic analysis and embedded topic models).

Conclusions

This review demonstrates that LDA is an accessible and flexible technique that provides researchers with the opportunity to reap the benefits of big textual data sets, and as such, we advocate for its continued use in the psychological sciences. Although some studies explicitly highlight engaging in data selection, data preprocessing, and data analysis, this was not always the case, thus reducing the capacity for reproducibility and evaluation of alignment with suggested practices. Therefore, we encourage researchers to be thorough and transparent in their reporting standards. To assist with reporting processes and to work toward best practice recommendations, we have developed an LDA Preferred Reporting Checklist (Table 3) outlining the key data selection, data preprocessing, and data analysis steps that researchers should report on where appropriate, or at the very least consider, when training an LDA model.

Furthermore, this review revealed that there is still an ongoing debate surrounding the necessity of certain preprocessing steps, the most appropriate estimation algorithms, and the most appropriate methods for determining the number of topics, with limited investigation into how these decisions may influence results. Given this, we recommend that future research be conducted across all stages of LDA to identify comprehensive and evidence-based best practice recommendations.

XSL•FO
RenderX

Table 3. Latent Dirichlet allocation (LDA) Preferred Reporting Checklist.

Section and topic	Item	Checklist item	Reported on page
Data selection			
Research area and purpose	1	Develop research questions, aims, objectives, and hypotheses as to which topics are likely to emerge.	
Research area and purpose	2	Consider the suitability of LDA; is this the most appropriate methodology to answer the research question (eg, consider if another topic modeling approach, especially for short texts, or traditional qualitative or quantitative approaches may be more suitable to the research question)?	
Inclusion and exclusion criteria	3	State inclusion and exclusion criteria for textual data to be used in LDA analysis (eg, based on researcher-developed dictionaries or data-driven approaches)	
Data sources	4	Indicate source of evidence (eg, social media, formal documentation, scientific literature, survey responses, and books) and comment on quality of writing. Consider ethical obligations associated with the use of a chosen data source.	
Data types	5	Specify the data types (eg, original posts or comments, titles, abstracts, or keywords) from within data sources that will be used for analyses.	
Structure of data	6	State the document level (eg, structured by citation, paragraph, post, and user).	
Structure of data	7	Specify number of documents.	
Structure of data	8	Specify length of documents (eg, range, mean, and SD).	
Data preprocessing			
Program, package, and version	9	Specify the program, package, and version used for preprocessing and analysis.	
Cleaning	10	List the preprocessing steps conducted (eg, punctuation, symbols and remove unrelated records, numbers, and whitespace).	
Stop words and selective text	11	Specify which stop word lists were applied and whether selective text was removed (eg, frequently or infrequently used words, hyperlinks, and names).	
N-grams and tokenization	112	Indicate the use of tokenization and specify the n-gram (eg, unigram, bigram, or trigram).	
Stemming or lemmatization	13	Indicate use of stemming, lemmatization, or neither and provide a rationale for decision.	
Stemming or lemmatization	14	Consider reporting results with and without stemming or lemmatization.	
Data analysis			
Estimation algorithms	15	State estimation algorithm used for analysis (eg, Gibbs sampling and variational EM[a] algorithm).	
Tuning parameters (alpha, beta, and k)	16	Specify alpha (eg, 0.01), beta (eg, 0.1, 50/k), and *k* (number of topics) parameters.	
Tuning parameters (alpha, beta, and k)	17	Detail iterative approach and specify metrics (eg, qualitative or quantitative such as coherence, perplexity, and log-likelihood) used to optimize parameters (ie, number of topics). Include an explanation of qualitative or quantitative cross-validation approaches.	
Evaluating relationships among topics	18	Evaluate and comment on relationships among topics (eg, visualization of topic modeling).	
Reporting results	19	Include examples of prototypical documents for each topic. If top words within topics have little coherence, use the label “uninterpretable” to describe those topics.	
Reproducibility: share deidentified data, code, and documentation	20	Publicly release deidentified data (when permitted), code, and documentation on platforms such as Open Science Framework to allow for reproducibility.	

[a]EM: expectation maximization.

Acknowledgments

LJH received funding from an Australian Government Research Training Scholarship.

XSL•FO
RenderX

Authors' Contributions

LJH, SSM, and GJY planned and developed the study protocol. LJH, GAO'D, and LMF collected the data. LJH collated the data. LJH, SSM, GAO'D, LMF, CJG, MF-T, EMW, JAM, and GJY interpreted results. LJH wrote the manuscript, and SSM, GAO'D, LMF, CJG, MFT, EMW, JAM, and GJY critically revised the manuscript for important intellectual content. All authors have contributed to the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of data selection, preprocessing, and analysis broken down by study.
[DOCX File , 48 KB - jmir_v24i11e33166_app1.docx]

References

1. Chen M, Mao S, Liu Y. Big data: a survey. Mobile Netw Appl 2014 Jan 22;19(2):171-209. [doi: 10.1007/s11036-013-0489-0]
2. Vu HQ, Li G, Law R. Discovering implicit activity preferences in travel itineraries by topic modeling. Tour Manag 2019 Dec;75:435-446. [doi: 10.1016/j.tourman.2019.06.011]
3. Puschmann C, Bastos M. How digital are the Digital Humanities? An analysis of two scholarly blogging platforms. PLoS One 2015 Feb 12;10(2):e0115035 [FREE Full text] [doi: 10.1371/journal.pone.0115035] [Medline: 25675441]
4. Cho Y, Fu P, Wu C. Popular research topics in marketing journals, 1995–2014. J Interact Market 2022 Jan 31;40(1):52-72. [doi: 10.1016/j.intmar.2017.06.003]
5. Cambria E, White B. Jumping NLP curves: a review of natural language processing research [review article]. IEEE Comput Intell Mag 2014 May;9(2):48-57. [doi: 10.1109/mci.2014.2307227]
6. Liddy ED. Enhanced text retrieval using natural language processing. Bul Am Soc Info Sci Tech 2005 Jan 31;24(4):14-16. [doi: 10.1002/bult.91]
7. Terragni S, Fersini E, Galuzzi B, Tropeano P, Candelieri A. OCTIS: comparing and optimizing topic models is simple!. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations. 2021 Presented at: 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations; Apr, 2021; Online. [doi: 10.18653/v1/2021.eacl-demos.31]
8. Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. Discourse Processe 1998 Jan;25(2-3):259-284. [doi: 10.1080/01638539809545028]
9. Qiang J, Qian Z, Li Y, Yuan Y, Wu X. Short text topic modeling techniques, applications, and performance: a survey. IEEE Trans Knowl Data Eng 2022 Mar 1;34(3):1427-1445. [doi: 10.1109/tkde.2020.2992485]
10. Blei D, Ng A, Jordan M. Latent dirichllocation. J Mach Learn Res 2003;3:993-1022.
11. Griffiths TL, Steyvers M. Finding scientific topics. Proc Natl Acad Sci U S A 2004 Apr 06;101 Suppl 1(suppl_1):5228-5235 [FREE Full text] [doi: 10.1073/pnas.0307752101] [Medline: 14872004]
12. Silge J, Robinson D. Text Mining With R A Tidy Approach. Sebastopol, California, United States: O'Reilly Media; 2017.
13. 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]
14. Kosinski M, Wang Y, Lakkaraju H, Leskovec J. Mining big data to extract patterns and predict real-life outcomes. Psychol Methods 2016 Dec;21(4):493-506. [doi: 10.1037/met0000105] [Medline: 27918179]
15. Asmussen CB, Møller C. Smart literature review: a practical topic modelling approach to exploratory literature review. J Big Data 2019 Oct 19;6(1). [doi: 10.1186/s40537-019-0255-7]
16. Banks GC, Woznyj HM, Wesslen RS, Ross RL. A review of best practice recommendations for text analysis in R (and a user-friendly app). J Bus Psychol 2018 Jan 11;33(4):445-459. [doi: 10.1007/s10869-017-9528-3]
17. Chen EE, Wojcik SP. A practical guide to big data research in psychology. Psychol Methods 2016 Dec;21(4):458-474. [doi: 10.1037/met0000111] [Medline: 27918178]
18. Haddi E, Liu X, Shi Y. The role of text pre-processing in sentiment analysis. Procedia Comput Sci 2013;17:26-32. [doi: 10.1016/j.procs.2013.05.005]
19. Lo R, He B, Ounis I. Automatically building a stopword list for an information retrieval system. J Digit Inf Manag 2005;3(1):3-8.
20. Multilingual Stopword Lists in R. GitHub. URL: https://github.com/quanteda/stopwords [accessed 2022-02-10]
21. NLTK's list of english stopwords. GitHub. URL: https://gist.github.com/sebleier/554280 [accessed 2022-02-10]
22. Schofield A, Magnusson M, Mimno D. Pulling out the stops: rethinking stopword removal for topic models. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. 2017 Presented at: 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers; Apr, 2017; Valencia, Spain. [doi: 10.18653/v1/e17-2069]
23. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]

XSL•FO
RenderX

24. Hoffman M, Bach F, Blei D. Online learning for Latent Dirichlet Allocation. In: Proceedings of the Advances in Neural Information Processing Systems 23 (NIPS 2010). 2010 Presented at: Advances in Neural Information Processing Systems 23 (NIPS 2010); Dec 6-11, 2010; Vancouver, British Columbia, Canada.
25. Wallach H, Mimno D, McCallum A. Rethinking LDA: why priors matter. In: Proceedings of the Advances in Neural Information Processing Systems 22 (NIPS 2009). 2009 Presented at: Advances in Neural Information Processing Systems 22 (NIPS 2009); Dec 7-10, 2009; British Columbia, Canada.
26. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring topic coherence over many models and many topics. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012 Presented at: EMNLP-CoNLL '12: 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; Jul 12 - 14, 2012; Jeju Island Korea.
27. Sievert C, Shirley K. LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces. 2014 Presented at: Workshop on Interactive Language Learning, Visualization, and Interfaces; Jun, 2014; Baltimore, Maryland, USA. [doi: 10.3115/v1/w14-3110]
28. Khalid H, Wade V. Topic detection from conversational dialogue corpus with parallel dirichllocation model and elbow method. arXiv 2020. [doi: 10.5121/csit.2020.100508]
29. Hoyle A, Goel P, Hian-Cheong A, Peskov D, Boyd-Graber J, Resnik P. Is automated topic model evaluation broken?: the incoherence of coherence. In: Proceedings of 35th Conference on Neural Information Processing Systems (NeurIPS 2021). 2021 Presented at: 35th Conference on Neural Information Processing Systems (NeurIPS 2021); Dec 6-14, 2021; Virtual.
30. Chang J, Gerrish S, Wang C, Boyd-Graber J, Blei D. Reading tea leaves: how humans interpret topic models. In: Proceedings of the 22nd International Conference on Neural Information Processing Systems. 2009 Presented at: NIPS'09: 22nd International Conference on Neural Information Processing Systems; Dec 7 - 10, 2009; Vancouver British Columbia Canada.
31. 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]
32. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
33. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
34. CALVO RA, MILNE DN, HUSSAIN MS, CHRISTENSEN H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng 2017 Jan 30;23(5):649-685. [doi: 10.1017/S1351324916000383]
35. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
36. 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]
37. Baghaei Lakeh A, Ghaffarzadegan N. Global trends and regional variations in studies of HIV/AIDS. Sci Rep 2017 Jun 23;7(1):4170 [FREE Full text] [doi: 10.1038/s41598-017-04527-6] [Medline: 28646150]
38. Cesare N, Oladeji O, Ferryman K, Wijaya D, Hendricks-Muñoz KD, Ward A, et al. Discussions of miscarriage and preterm births on Twitter. Paediatr Perinat Epidemiol 2020 Sep 08;34(5):544-552 [FREE Full text] [doi: 10.1111/ppe.12622] [Medline: 31912544]
39. Tang C, Zhou L, Plasek J, Rozenblum R, Bates D. Comment topic evolution on a cancer institution's Facebook page. Appl Clin Inform 2017 Aug 23;8(3):854-865 [FREE Full text] [doi: 10.4338/ACI-2017-04-RA-0055] [Medline: 28832069]
40. Vaughn DA, van Deen WK, Kerr WT, Meyer TR, Bertozzi AL, Hommes DW, et al. Using insurance claims to predict and improve hospitalizations and biologics use in members with inflammatory bowel diseases. J Biomed Inform 2018 May;81:93-101 [FREE Full text] [doi: 10.1016/j.jbi.2018.03.015] [Medline: 29625187]
41. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
42. Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S. Detection of cases of noncompliance to drug treatment in patient forum posts: topic model approach. J Med Internet Res 2018 Mar 14;20(3):e85 [FREE Full text] [doi: 10.2196/jmir.9222] [Medline: 29540337]
43. Afshar M, Joyce C, Dligach D, Sharma B, Kania R, Xie M, et al. Subtypes in patients with opioid misuse: a prognostic enrichment strategy using electronic health record data in hospitalized patients. PLoS One 2019;14(7):e0219717 [FREE Full text] [doi: 10.1371/journal.pone.0219717] [Medline: 31310611]
44. Alam F, Ofli F, Imran M. Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria. Behav Inform Technol 2019 May 14;39(3):288-318. [doi: 10.1080/0144929X.2019.1610908]

XSL•FO
RenderX

45. Barry AE, Valdez D, Padon AA, Russell AM. Alcohol advertising on Twitter—a topic model. Am J Health Educ 2018 Jun 29;49(4):256-263. [doi: 10.1080/19325037.2018.1473180]
46. Bittermann A, Fischer A. How to identify hot topics in psychology using topic modeling. Zeitschrift für Psychologie 2018 Jan;226(1):3-13. [doi: 10.1027/2151-2604/a000318]
47. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Seeing the "Big" picture: big data methods for exploring relationships between usage, language, and outcome in internet intervention data. J Med Internet Res 2016 Aug 31;18(8):e241 [FREE Full text] [doi: 10.2196/jmir.5725] [Medline: 27580524]
48. Carron-Arthur B, Reynolds J, Bennett K, Bennett A, Griffiths KM. What's all the talk about? Topic modelling in a mental health internet support group. BMC Psychiatry 2016 Oct 28;16(1):367 [FREE Full text] [doi: 10.1186/s12888-016-1073-5] [Medline: 27793131]
49. Chen AT, Zhu SH, Conway M. What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. J Med Internet Res 2015 Sep 29;17(9):e220 [FREE Full text] [doi: 10.2196/jmir.4517] [Medline: 26420469]
50. Choi S, Seo JY. An exploratory study of the research on caregiver depression: using bibliometrics and LDA topic modeling. Issues Ment Health Nurs 2020 Jul;41(7):592-601. [doi: 10.1080/01612840.2019.1705944] [Medline: 32286089]
51. Choudhury P, Wang D, Carlson NA, Khanna T. Machine learning approaches to facial and text analysis: discovering CEO oral communication styles. Strat Mgmt J 2019 Aug 06;40(11):1705-1732. [doi: 10.1002/smj.3067]
52. Cohan A, Young S, Yates A, Goharian N. Triaging content severity in online mental health forums. J Assoc Inform Sci Technol 2017 Sep 25;68(11):2675-2689. [doi: 10.1002/asi.23865]
53. Feldhege J, Moessner M, Bauer S. Who says what? Content and participation characteristics in an online depression community. J Affect Disord 2020 Feb 15;263:521-527. [doi: 10.1016/j.jad.2019.11.007] [Medline: 31780138]
54. Franz PJ, Nook EC, Mair P, Nock MK. Using topic modeling to detect and describe self-injurious and related content on a large-scale digital platform. Suicide Life Threat Behav 2020 Feb;50(1):5-18. [doi: 10.1111/sltb.12569] [Medline: 31264733]
55. Gerber MS. Predicting crime using Twitter and kernel density estimation. Decision Support Syst 2014 May;61(3):115-125. [doi: 10.1016/j.dss.2014.02.003]
56. Giorgi S, Maoret M, J. Zajac E. On the relationship between firms and their legal environment: the role of cultural consonance. Organization Sci 2019 Jul;30(4):803-830. [doi: 10.1287/orsc.2018.1250]
57. Guo L, Li S, Lu R, Yin L, Gorson-Deruel A, King L. The research topic landscape in the literature of social class and inequality. PLoS One 2018;13(7):e0199510 [FREE Full text] [doi: 10.1371/journal.pone.0199510] [Medline: 29965983]
58. Hemmatian B, Sloman SJ, Cohen Priva U, Sloman SA. Think of the consequences: a decade of discourse about same-sex marriage. Behav Res Methods 2019 Aug;51(4):1565-1585. [doi: 10.3758/s13428-019-01215-3] [Medline: 30859479]
59. Hwang Y, Kim HJ, Choi HJ, Lee J. Exploring abnormal behavior patterns of online users with emotional eating behavior: topic modeling study. J Med Internet Res 2020 Mar 31;22(3):e15700 [FREE Full text] [doi: 10.2196/15700] [Medline: 32229461]
60. Jaworska S, Nanda A. Doing well by talking good: a topic modelling-assisted discourse study of corporate social responsibility. Applied Linguistics 2016 Jun 06;229(6):amw014-amw013. [doi: 10.1093/applin/amw014]
61. Jung Y, Suh Y. Mining the voice of employees: a text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. Decision Support Syst 2019 Aug;123(6):113074-113078. [doi: 10.1016/j.dss.2019.113074]
62. Kagashe I, Yan Z, Suheryani I. Enhancing seasonal influenza surveillance: topic analysis of widely used medicinal drugs using twitter data. J Med Internet Res 2017 Sep 12;19(9):e315 [FREE Full text] [doi: 10.2196/jmir.7393] [Medline: 28899847]
63. Karami A, Swan SC, White CN, Ford K. Hidden in plain sight for too long: using text mining techniques to shine a light on workplace sexism and sexual harassment. Psychol Violence 2019 Jun 27;229(6):1641-1648. [doi: 10.1037/vio0000239]
64. Kee YH, Li C, Kong LC, Tang CJ, Chuang K. Scoping review of mindfulness research: a topic modelling approach. Mindfulness 2019 Apr 15;10(8):1474-1488. [doi: 10.1007/s12671-019-01136-4]
65. Kigerl A. Profiling Cybercriminals. Social Sci Comput Rev 2017 Sep 20;36(5):591-609. [doi: 10.1177/0894439317730296]
66. Kreitzberg DS, Murthy D, Loukas A, Pasch KE. Heat not burn tobacco promotion on instagram. Addict Behav 2019 Apr;91:112-118. [doi: 10.1016/j.addbeh.2018.09.003] [Medline: 30241774]
67. Landstrøm EK, Jeppesen SH, Demant J. Paedophilia discourses in Denmark: towards a mixed method digital discourse approach. Sexualities 2017 Nov 20;22(3):381-400. [doi: 10.1177/1363460717741791]
68. Lee AJ, Jones BC, DeBruine LM. Investigating the association between mating-relevant self-concepts and mate preferences through a data-driven analysis of online personal descriptions. Evolution Human Behav 2019 May;40(3):325-335. [doi: 10.1016/j.evolhumbehav.2019.01.005]
69. Lee K, Lee D, Hong HJ. Text mining analysis of teachers' reports on student suicide in South Korea. Eur Child Adolesc Psychiatry 2020 Apr 20;29(4):453-465. [doi: 10.1007/s00787-019-01361-1] [Medline: 31222535]
70. Liang BO, Wang YE, Tsou MH. A "fitness" theme may mitigate regional prevalence of overweight and obesity: evidence from Google Search and Tweets. J Health Commun 2019;24(9):683-692. [doi: 10.1080/10810730.2019.1657526] [Medline: 31469057]

XSL•FO
RenderX

71. Liu X, Sun M, Li J. Research on gender differences in online health communities. Int J Med Inform 2018 Mar;111:172-181. [doi: 10.1016/j.ijmedinf.2017.12.019] [Medline: 29425630]
72. Liu Q, Woo M, Zou X, Champaneria A, Lau C, Mubbashar MI, et al. Symptom-based patient stratification in mental illness using clinical notes. J Biomed Inform 2019 Oct;98:103274 [FREE Full text] [doi: 10.1016/j.jbi.2019.103274] [Medline: 31499185]
73. Liu S, Zhang RY, Kishimoto T. Analysis and prospect of clinical psychology based on topic models: hot research topics and scientific trends in the latest decades. Psychol Health Med 2021 Apr;26(4):395-407. [doi: 10.1080/13548506.2020.1738019] [Medline: 32156155]
74. Liu J, Kong J, Zhang X. Study on differences between patients with physiological and psychological diseases in online health communities: topic analysis and sentiment analysis. Int J Environ Res Public Health 2020 Feb 26;17(5):1508 [FREE Full text] [doi: 10.3390/ijerph17051508] [Medline: 32111045]
75. Lou C, Tan S, Chen X. Investigating consumer engagement with influencer- vs. brand-promoted ads: the roles of source and disclosure. J Interactive Advertising 2019 Oct 15;19(3):169-186. [doi: 10.1080/15252019.2019.1667928]
76. Louvigné S, Rubens N. Meaning-making analysis and topic classification of SNS goal-based messages. Behaviormetrika 2016 Jan 1;43(1):65-82. [doi: 10.2333/bhmk.43.65]
77. Magua W, Zhu X, Bhattacharya A, Filut A, Potvien A, Leatherberry R, et al. Are female applicants disadvantaged in national institutes of health peer review? Combining algorithmic text mining and qualitative methods to detect evaluative differences in r01 reviewers' critiques. J Womens Health (Larchmt) 2017 May;26(5):560-570 [FREE Full text] [doi: 10.1089/jwh.2016.6021] [Medline: 28281870]
78. McCoy TH. Mapping the delirium literature through probabilistic topic modeling and network analysis: a computational scoping review. Psychosomatics 2019;60(2):105-120. [doi: 10.1016/j.psym.2018.12.003] [Medline: 30686485]
79. Merrill M, Åkerlund M. Standing up for Sweden? The racist discourses, architectures and affordances of an anti-immigration Facebook group. J Comput Mediated Commun 2018;23(6):332-353. [doi: 10.1093/jcmc/zmy018]
80. Murdock J, Allen C, DeDeo S. Exploration and exploitation of Victorian science in Darwin's reading notebooks. Cognition 2017 Feb;159:117-126. [doi: 10.1016/j.cognition.2016.11.012] [Medline: 27939837]
81. Oh J, Stewart AE, Phelps RE. Topics in the journal of counseling psychology, 1963-2015. J Couns Psychol 2017 Nov;64(6):604-615. [doi: 10.1037/cou0000218] [Medline: 29154573]
82. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz J, et al. Social media based analysis of opioid epidemic using reddit. AMIA Annu Symp Proc 2018;2018:867-876 [FREE Full text] [Medline: 30815129]
83. Pantti M, Nelimarkka M, Nikunen K, Titley G. The meanings of racism: public discourses about racism in Finnish news media and online discussion forums. Eur J Commun 2019 Sep 17;34(5):503-519. [doi: 10.1177/0267323119874253]
84. 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]
85. Park A, Conway M. Tracking health related discussions on reddit for public health applications. AMIA Annu Symp Proc 2017;2017:1362-1371 [FREE Full text] [Medline: 29854205]
86. Ray A, Bala PK, Dwivedi YK. Exploring values affecting e-Learning adoption from the user-generated-content: a consumption-value-theory perspective. J Strategic Market 2020 Apr 07;29(5):430-452. [doi: 10.1080/0965254X.2020.1749875]
87. Ruiz N, Witting A, Ahnert L, Piskernik B. Reflective functioning in fathers with young children born preterm and at term. Attach Hum Dev 2020 Feb 21;22(1):32-45. [doi: 10.1080/14616734.2019.1589059] [Medline: 30898032]
88. Rumshisky A, Ghassemi M, Naumann T, Szolovits P, Castro VM, McCoy TH, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. Transl Psychiatry 2016 Oct 18;6(10):e921 [FREE Full text] [doi: 10.1038/tp.2015.182] [Medline: 27754482]
89. Santos T, Louçã J, Coelho H. The digital transformation of the public sphere. Syst Res Behav Sci 2019 Nov 11;36(6):778-788. [doi: 10.1002/sres.2644]
90. Shahin S, Dai Z. Understanding public engagement with global aid agencies on Twitter: a technosocial framework. Am Behav Sci 2019 Mar 06;63(12):1684-1707. [doi: 10.1177/0002764219835248]
91. Shin J, Guo Q, Gierl MJ. Multiple-choice item distractor development using topic modeling approaches. Front Psychol 2019 Dec;10(6):825-828 [FREE Full text] [doi: 10.3389/fpsyg.2019.00825] [Medline: 31133911]
92. Sieweke J, Santoni S. Natural experiments in leadership research: an introduction, review, and guidelines. Leadersh Q 2020 Feb;31(1):101338-101338. [doi: 10.1016/j.leaqua.2019.101338]
93. Son J, Lee HK, Jin S, Lee J. Content features of tweets for effective communication during disasters: a media synchronicity theory perspective. Int J Inform Manag 2019 Apr;45(6):56-68. [doi: 10.1016/j.ijinfomgt.2018.10.012]
94. Sorour S, Goda K, Mine T. Comment data mining to estimate student performance considering consecutive lessons. Educ Technol Soc 2017;20(1):73-86.
95. Sperandeo R, Messina G, Iennaco D, Sessa F, Russo V, Polito R, et al. What does personality mean in the context of mental health? A topic modeling approach based on abstracts published in Pubmed over the last 5 years. Front Psychiatry 2019 Jan 9;10:938 [FREE Full text] [doi: 10.3389/fpsyt.2019.00938] [Medline: 31998157]

XSL•FO
RenderX

96. Székely N, Vom Brocke J. What can we learn from corporate sustainability reporting? Deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique. PLoS One 2017 Apr 12;12(4):e0174807 [FREE Full text] [doi: 10.1371/journal.pone.0174807] [Medline: 28403158]
97. Törnberg A, Törnberg P. Combining CDA and topic modeling: analyzing discursive connections between Islamophobia and anti-feminism on an online forum. Discourse Soc 2016 Mar 28;27(4):401-422. [doi: 10.1177/0957926516634546]
98. Tran BX, McIntyre RS, Latkin CA, Phan HT, Vu GT, Nguyen HL, et al. The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis. Int J Environ Res Public Health 2019 Jun 18;16(12):2150 [FREE Full text] [doi: 10.3390/ijerph16122150] [Medline: 31216619]
99. Tran BX, Harijanto C, Vu GT, Ho RC. Global mapping of interventions to improve quality of life using mind-body therapies during 1990-2018. Complement Ther Med 2020 Mar;49:102350. [doi: 10.1016/j.ctim.2020.102350] [Medline: 32147055]
100. Turrentine FE, Dreisbach CN, St Ivany AR, Hanks JB, Schroen AT. Influence of gender on surgical residency applicants' recommendation letters. J Am Coll Surg 2019 Apr;228(4):356-65.e3. [doi: 10.1016/j.jamcollsurg.2018.12.020] [Medline: 30630084]
101. Wang S, Ding Y, Zhao W, Huang Y, Perkins R, Zou W, et al. Text mining for identifying topics in the literatures about adolescent substance use and depression. BMC Public Health 2016 Mar 19;16:279 [FREE Full text] [doi: 10.1186/s12889-016-2932-1] [Medline: 26993983]
102. Weij F, Berkers P, Engelbert J. Western solidarity with Pussy Riot and the Twittering of cosmopolitan selves. Int J Consum Stud 2015;39(5):489-494. [doi: 10.1111/ijcs.12215]
103. Westmaas JL, McDonald BR, Portier KM. Topic modeling of smoking- and cessation-related posts to the American cancer society's cancer survivor network (CSN): implications for cessation treatment for cancer survivors who smoke. Nicotine Tob Res 2017 Aug 01;19(8):952-959. [doi: 10.1093/ntr/ntx064] [Medline: 28340059]
104. Wu P, Yu S, Wang D. Using a learner-topic model for mining learner interests in open learning environments. Educ Technol Soc 2018;21(2):192-204.
105. Yoon S. What can we learn about mental health needs from tweets mentioning dementia on world Alzheimer's day? J Am Psychiatr Nurses Assoc 2016 Nov;22(6):498-503 [FREE Full text] [doi: 10.1177/1078390316663690] [Medline: 27803262]
106. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. J Med Internet Res 2017 Jan 20;19(1):e24 [FREE Full text] [doi: 10.2196/jmir.5780] [Medline: 28108428]
107. Zhao Y, Zhang J, Wu M. Finding users' voice on social media: an investigation of online support groups for autism-affected users on Facebook. Int J Environ Res Public Health 2019 Nov 29;16(23):4804 [FREE Full text] [doi: 10.3390/ijerph16234804] [Medline: 31795451]
108. Zheng P, Shahin S. Live tweeting live debates: how Twitter reflects and refracts the US political climate in a campaign season. Inform Commun Soc 2018 Aug 06;23(3):337-357. [doi: 10.1080/1369118X.2018.1503697]
109. Zou C. Analyzing research trends on drug safety using topic modeling. Expert Opin Drug Saf 2018 Jun;17(6):629-636. [doi: 10.1080/14740338.2018.1458838] [Medline: 29621918]
110. Ooms J. hunspell: High-Performance Stemmer, Tokenizer and Spell Checker. R package version 3. 2020 Dec 9. URL: https://CRAN.R-project.org/package=hunspell [accessed 2022-02-05]
111. Arun R, Suresh V, Veni MC, Murthy N. On finding the natural number of topics with latent dirichlet allocation: some observations. In: Advances in Knowledge Discovery and Data Mining. Berlin, Heidelberg: Springer; 2010.
112. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Document numérique 2014 Apr 30;17(1):61-84. [doi: 10.3166/dn.17.1.61-84]
113. Airoldi EM, Bischof JM. Improving and evaluating topic models and other models of text. J Am Stat Assoc 2017 Jan 04;111(516):1381-1403. [doi: 10.1080/01621459.2015.1051182]
114. Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical dirichlet processes. J Am Stat Assoc 2012 Jan 01;101(476):1566-1581. [doi: 10.1198/016214506000000302]
115. Taddy M. On estimation and selection for topic models. In: Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics. 2012 Presented at: Fifteenth International Conference on Artificial Intelligence and Statistics; Apr 21 - 23, 2012; Canary Islands.
116. Steyvers M, Griffiths T. Probabilistic topic models. In: Handbook of Latent Semantic Analysis. Mahwah, MJ: Lawrence Erlbaum Associates Publishers; 2007.
117. topicmodels_learning. GitHub. URL: https://github.com/trinker/topicmodels_learning/blob/master/functions/optimal_k.R [accessed 2022-02-10]
118. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. Neurocomputing 2009 Mar;72(7-9):1775-1781. [doi: 10.1016/j.neucom.2008.06.011]
119. AlSumait L, Barbará D, Gentle J, Domeniconi C. Topic significance ranking of LDA generative models. In: Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer; 2009.
120. Bao Y, Datta A. Simultaneously discovering and quantifying risk types from textual risk disclosures. Manag Sci 2014 Jun;60(6):1371-1391. [doi: 10.1287/mnsc.2014.1930]

XSL•FO
RenderX

121. Ford E, Shepherd S, Jones K, Hassan L. Toward an ethical framework for the text mining of social media for health research: a systematic review. Front Digit Health 2020 Jan 26;2:592237 [FREE Full text] [doi: 10.3389/fdgth.2020.592237] [Medline: 34713062]
122. Gilbert J, Ng V, Niu J, Rees EE. A call for an ethical framework when using social media data for artificial intelligence applications in public health research. Can Commun Dis Rep 2020 Jun 04;46(6):169-173 [FREE Full text] [doi: 10.14745/ccdr.v46i06a03] [Medline: 32673381]
123. Bird S, Klein E, Loper E. Natural Language Processing with Python. Sebastopol, California, United States: O'Reilly Media; 2009.
124. Tang J, Meng Z, Nguyen X, Mei Q, Zhang M. Understanding the limiting factors of topic modeling via posterior contraction analysis. In: Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32. 2014 Presented at: 31st International Conference on International Conference on Machine Learning - Volume 32; Jun 21 - 26, 2014; Beijing China.
125. Albalawi R, Yeap TH, Benyoucef M. Using topic modeling methods for short-text data: a comparative analysis. Front Artif Intell 2020;3:42 [FREE Full text] [doi: 10.3389/frai.2020.00042] [Medline: 33733159]
126. Mehrotra R, Sanner S, Buntine W, Xie L. Improving LDA topic models for microblogs via tweet pooling and automatic labeling. In: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. 2013 Presented at: SIGIR '13: The 36th International ACM SIGIR conference on research and development in Information Retrieval; Jul 28- Aug 1, 2013; Dublin Ireland.
127. Ito J, Song J, Toda H, Koike Y, Oyama S. Assessment of tweet credibility with LDA features. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW '15: 24th International World Wide Web Conference; May 18 - 22, 2015; Florence Italy.
128. Sbalchiero S, Eder M. Topic modeling, long texts and the best number of topics. Some Problems and solutions. Qual Quant 2020 Feb 17;54(4):1095-1108. [doi: 10.1007/s11135-020-00976-w]
129. Denny MJ, Spirling A. Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it. Polit Anal 2018 Mar 19;26(2):168-189. [doi: 10.1017/pan.2017.44]
130. Bekkerman R, Allan J. Using bigrams in text categorization. Technical Report IR-408, Center of Intelligent Information Retrieval. URL: https://maroo.cs.umass.edu/getpdf.php?id=552 [accessed 2022-02-10]
131. Yang T, Torget A, Mihalcea R. Topic modeling on historical newspapers. In: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. 2011 Presented at: ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities; Jun 24, 2011; Portland, OR, USA.
132. Schofield A, Mimno D. Comparing apples to apple: the effects of stemmers on topic models. Transact Assoc Computat Linguistic 2016 Dec;4:287-300. [doi: 10.1162/tacl_a_00099]
133. Singh J, Gupta V. A systematic review of text stemming techniques. Artif Intell Rev 2016 Aug 1;48(2):157-217. [doi: 10.1007/s10462-016-9498-2]
134. Asuncion A, Welling M, Smyth P, Teh Y. On smoothing and inference for topic models. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. 2012 Presented at: UAI '09: 25 conference on Uncertainty in Artificial Intelligence; Jun 18 - 21, 2009; Montreal Quebec Canada.
135. Blei D, Jordan M. Variational methods for the Dirichlet process. In: Proceedings of the twenty-first international conference on Machine learning. 2004 Presented at: ICML '04: Proceedings of the twenty-first international conference on Machine learning; Jul 4 - 8, 2004; Banff Alberta Canada. [doi: 10.1145/1015330.1015439]
136. Braun M, McAuliffe J. Variational inference for large-scale models of discrete choice. J Am Stat Assoc 2012 Jan 01;105(489):324-335. [doi: 10.1198/jasa.2009.tm08030]
137. Zubir W, Aziz I, Jaafar J, Hasan M. Inference algorithms in latent Dirichlet allocation for semantic classification. In: Applied Computational Intelligence and Mathematical Methods. Cham: Springer; 2017.
138. Agrawal A, Fu W, Menzies T. What is wrong with topic modeling? And how to fix it using search-based software engineering. Inform Softw Technol 2018 Jun;98:74-88. [doi: 10.1016/j.infsof.2018.02.005]
139. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 Presented at: WSDM 2015: Eighth ACM International Conference on Web Search and Data Mining; Feb 2 - 6, 2015; Shanghai China. [doi: 10.1145/2684822.2685324]
140. Lau J, Newman D, Baldwin T. Machine reading tea leaves: automatically evaluating topic coherence and topic model quality. In: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. 2014 Presented at: 14th Conference of the European Chapter of the Association for Computational Linguistics; Apr, 2014; Gothenburg, Sweden.
141. Wallach H, Murray I, Salakhutdinov R, Mimno D. Evaluation methods for topic models. In: Proceedings of the 26th Annual International Conference on Machine Learning. 2009 Presented at: ICML '09: The 26th Annual International Conference on Machine Learning held in conjunction with the 2007 International Conference on Inductive Logic Programming; Jun 14 - 18, 2009; Montreal Quebec Canada. [doi: 10.1145/1553374.1553515]
142. pyLDAvis homepage. pyLDAvis. URL: https://pyldavis.readthedocs.io/en/latest/readme.html [accessed 2022-02-19]

XSL•FO
RenderX

143. Mcauliffe J, Blei D. Supervised topic models. In: Proceedings of the Advances in Neural Information Processing Systems 20 (NIPS 2007). 2007 Presented at: Advances in Neural Information Processing Systems 20 (NIPS 2007); 2007; Vancouver, British Columbia.
144. Jacobucci R, Ammerman BA, Tyler Wilcox K. The use of text-based responses to improve our understanding and prediction of suicide risk. Suicide Life Threat Behav 2021 Feb 24;51(1):55-64. [doi: 10.1111/sltb.12668] [Medline: 33624877]
145. Sperkova L. Review of latent Dirichlet allocation methods usable in voice of customer analysis. Acta Informatica Pragensia 2018 Dec 31;7(2):152-165. [doi: 10.18267/j.aip.120]

Abbreviations

LDA: latent Dirichlet allocation
MALLET: Machine Learning for Language Toolkit
NLP: natural language processing
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
VB: variational Bayes

Edited by R Kukafka; submitted 26.08.21; peer-reviewed by D Mimno, D Low, J Plasek; comments to author 24.10.21; revised version received 18.02.22; accepted 30.05.22; published 08.11.22.

Please cite as:
Hagg LJ, Merkouris SS, O'Dea GA, Francis LM, Greenwood CJ, Fuller-Tyszkiewicz M, Westrupp EM, Macdonald JA, Youssef GJ
Examining Analytic Practices in Latent Dirichlet Allocation Within Psychological Science: Scoping Review
J Med Internet Res 2022;24(11):e33166
URL: https://www.jmir.org/2022/11/e33166
doi:10.2196/33166
PMID:36346659

©Lauryn J Hagg, Stephanie S Merkouris, Gypsy A O'Dea, Lauren M Francis, Christopher J Greenwood, Matthew Fuller-Tyszkiewicz, Elizabeth M Westrupp, Jacqui A Macdonald, George J Youssef. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Determination of Markers of Successful Implementation of Mental Health Apps for Young People: Systematic Review

Holly Alice Bear[1], PhD; Lara Ayala Nunes[1], PhD; John DeJesus[2], BSc; Shaun Liverpool[3], PhD; Bettina Moltrecht[4,5], PhD; Lakshmi Neelakantan[1], PhD; Elinor Harriss[6], MSc; Edward Watkins[7], PhD; Mina Fazel[1], DM

[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[2]Department of Psychology, York University, Toronto, ON, Canada

[3]Faculty of Health, Social Care and Medicine, Edge Hill University, Ormskirk, United Kingdom

[4]Research Department for Clinical, Educational and Health Psychology, University College London, London, United Kingdom

[5]Centre for Longitudinal Studies, University College London, London, United Kingdom

[6]Bodleian Health Care Libraries, University of Oxford, Oxford, United Kingdom

[7]School of Psychology, University of Exeter, Exeter, United Kingdom

Corresponding Author:
Holly Alice Bear, PhD
Department of Psychiatry
University of Oxford
Warneford Hospital
Oxford, OX3 7JX
United Kingdom
Phone: 44 01865613127
Email: holly.bear@psych.ox.ac.uk

Abstract

Background: Smartphone apps have the potential to address some of the current issues facing service provision for young people's mental health by improving the scalability of evidence-based mental health interventions. However, very few apps have been successfully implemented, and consensus on implementation measurement is lacking.

Objective: This review aims to determine the proportion of evidence-based mental health and well-being apps that have been successfully adopted and sustained in *real-world* settings. A secondary aim is to establish if key implementation determinants such as coproduction, acceptability, feasibility, appropriateness, and engagement contribute toward successful implementation and longevity.

Methods: Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, an electronic search of 5 databases in 2021 yielded 18,660 results. After full-text screening, 34 articles met the full eligibility criteria, providing data on 29 smartphone apps studied with individuals aged 15 to 25 years.

Results: Of 34 studies, only 10 (29%) studies were identified that were evaluating the effectiveness of 8 existing, commercially available mental health apps, and the remaining 24 (71%) studies reported the development and evaluation of 21 newly developed apps, of which 43% (9/21) were available, commercially or otherwise (eg, in mental health services), at the time of enquiry. Most studies addressed some implementation components including adoption, acceptability, appropriateness, feasibility, and engagement. Factors including high cost, funding constraints, and lengthy research processes impeded implementation.

Conclusions: Without addressing common implementation drivers, there is considerable redundancy in the translation of mobile mental health research findings into practice. Studies should embed implementation strategies from the outset of the planned research, build collaborations with partners already working in the field (academic and commercial) to capitalize on existing interventions and platforms, and modify and evaluate them for local contexts or target problems and populations.

Trial Registration: PROSPERO CRD42021224365; https://tinyurl.com/4umpn85f

(J Med Internet Res 2022;24(11):e40347) doi:10.2196/40347

KEYWORDS

adolescent mental health; smartphones; mobile apps; apps; implementation science; mobile phone

XSL•FO
RenderX

Introduction

Background

There has been a proliferation in the number of smartphone apps being developed, both commercially and in academic research programs, which aim to improve mental health and well-being. Recent estimates suggest that anywhere from 10,000 [1] to 22,750 mental health apps exist [2]. Although many of these apps can be accessed directly by individuals in the commercial app marketplace as self-care tools, they are also playing an increasing role in clinical services, supplementing or enhancing traditional interventions [3]. The rapid expansion in the research and development of mental health and well-being apps highlights how much interest and potential there is thought to be in the mobile health arena.

Most common mental health disorders, including depression and anxiety, have their onset during adolescence and, if not successfully resolved, can lead to negative impacts well into adulthood [4,5]. Given the increasing number of young people (between the ages of 15 and 25 years) using digital technologies, smartphone-based interventions provide a scalable solution to support this group to manage their mental health and well-being [6]. Apps have the potential to address some of the accessibility issues in service provision for young people's mental health, especially for underserved populations. Emerging evidence suggests that some apps may produce significant symptom improvement across multiple outcomes compared with waitlist or control conditions [7-9]. Despite this promise, empirical research often fails to translate into meaningful and sustained implementation in "real-world" settings [10,11]. This can be attributed, in part, to the complex and lengthy process of implementing and maintaining evidence-based approaches in practice, as well as the commercial and regulatory complexities of scaling up mobile technologies in health services [12,13]. Besides innovation and efficacy, other factors, including user engagement, usability, acceptability, accessibility, and low cost, are key prerequisites for adoption, scalability, and uptake [14-17].

Given these multifaceted challenges, it is important to identify what facilitates and inhibits the implementation of mobile mental health interventions [10]. Currently, our understanding of how mental health apps are implemented in real-world settings is limited in several ways. Foremost, implementation processes and outcomes from research trials are seldom recorded or reported, and implementation efforts often lack a solid theoretical or model-based approach, making it difficult to understand and explain how and why implementation succeeds or fails [18,19]. In the context of mobile mental health apps, successful implementation can be measured by the extent to which the intervention has been embedded into service provision, the number of app users, the frequency of app use, app engagement, and evidence of sustained use following the end of a research trial [20,21]. Assessing implementation outcomes using a conceptually grounded framework allows for a systematic assessment of outcomes while also supporting the rigor and reproducibility of implementation research and providing building blocks for the implementation of future interventions.

Existing scoping and systematic reviews have focused on reviewing and critically appraising the methodological rigor and quality of implementation effectiveness studies, reporting implementation outcomes as their primary outcomes [22,23]. To the best of our knowledge, no review has taken a systematic approach to assessing the successful implementation and sustainment of all evidence-based mental health apps for young people. In this review, we assessed the factors influencing implementation success according to a set of implementation outcome criteria based on a modified version of the implementation framework by Proctor et al [24]. In total, 10 implementation variables were examined, of which 8 were from the Proctor model: acceptability (perceived usefulness and satisfaction with a technology), appropriateness (fitness for purpose), feasibility (extent to which a technology was successfully used), fidelity (implementation as intended), cost (financial impact of technology implementation), adoption (technology uptake and use), penetration (spread or reach of the technology), and sustainability (sustained uptake by users or maintenance or integration of a technology within a health care service). Two additional relevant outcomes were added: coproduction (user involvement in intervention development and evaluation) and engagement (adherence and dropout) [25,26].

Although conceptually distinct constructs, the implementation variables listed above are dynamically interrelated and sequentially contingent on one another [24,27]. For an app to be engaging, widely adopted, and well sustained, it must first be acceptable, appropriate, and feasible. It may be that an app-based intervention is deemed highly relevant and applicable to young people's needs (high appropriateness) but may be costly to download and time intensive (low feasibility). Similarly, an intervention may be considered by a mental health service as a good fit to address young people's needs (high appropriateness); nevertheless, the service user may be reluctant to use it if they dislike a certain feature of the intervention (low acceptability). Given the potential benefits of smartphone apps in supporting the mental health and well-being of young people, it is critical that researchers and app developers place greater emphasis on enhancing the engagement, implementation, and scalability of efficacious interventions in local contexts or specific populations.

Research Questions

The aim of this review was to determine how successfully evidence-based mobile apps, which aim to promote well-being and mental health outcomes in young people, are adopted, scaled up, and sustained in real-world settings. The research questions of interest were as follows:

1. What proportion of evidence-based mental health apps are sustained and adopted after development?
2. What components are needed for successful implementation outcomes and what are the common barriers?

XSL•FO
RenderX

Methods

The systematic review protocol was registered in PROSPERO (CRD42021224365).

Literature Search and Search Strategy

An information specialist (EH) performed an electronic search of the following databases from January 1, 2011, to the search date on February 2, 2021: Ovid Embase, Ovid MEDLINE, Ovid PsycINFO, Cochrane Database of Systematic Reviews, and Cochrane Central Register of Controlled Trials. The search strategies used text words and relevant indexing to capture the concepts of studies on the effectiveness or trials of mental health apps for young people. The search strategy was guided by similar reviews exploring digital mental health interventions for young people [25], and the terms for apps were derived from Cochrane reviews [28,29]. The full search strategies are available in Multimedia Appendix 1. Duplicates were removed following the method described by Falconer [30], and records were then screened by titles and abstracts to complete the process manually. The reference lists of included studies and relevant systematic reviews were assessed for additional relevant studies. All references were exported to Endnote X9 (Clarivate) and then to the systematic review software Rayyan [31].

Inclusion and Exclusion Criteria

The focus of this review was the implementation of app-based interventions that aim to promote mental health and well-being, prevent mental health problems, and treat existing mental health problems in young people. Screened articles were included if (1) the study targeted young people with a mean age of 15 to 25 years, with or without a formal mental health or physical health diagnosis (eg, targeting anxiety in adolescents with diabetes); (2) the intervention was an efficacious "native" mobile app (ie, not on a web browser), whose primary aim was to promote well-being, prevent mental health problems, or treat existing mental health problems; (3) the primary outcome was a measure of mental health or well-being, including change in anxiety and depressive symptoms, diagnosis, problem severity, problem improvement, recovery, remission, or more general change in mental health or well-being across at least 2 time points (eg, baseline and after the intervention or follow-up); and (4) the intervention was efficacious, that is, it had beneficial mental health or well-being outcomes compared with any other type of digital intervention, usual care (eg, psychotherapy), waitlist control group, or no-intervention control group. Both randomized and nonrandomized studies were considered for inclusion.

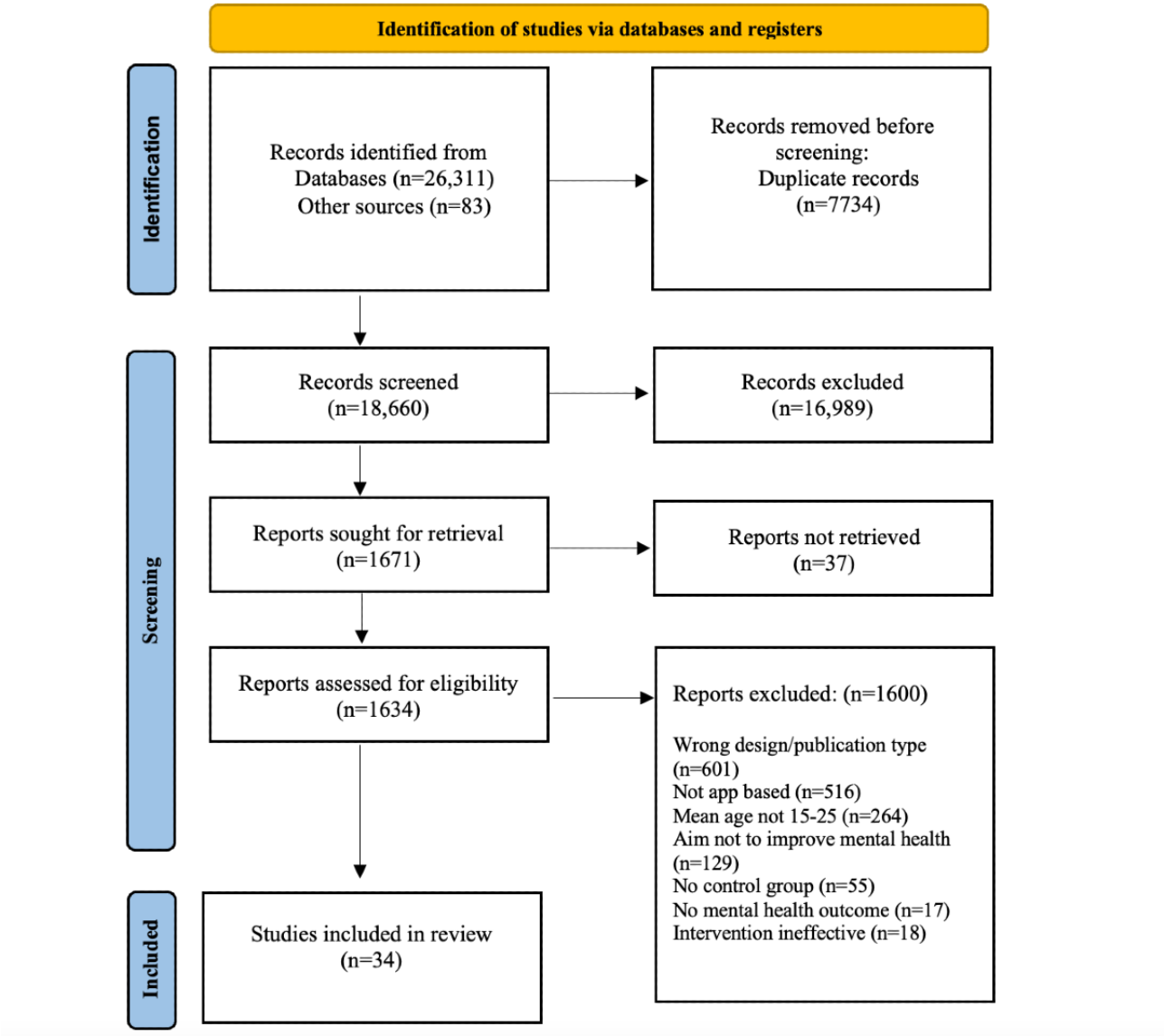
Articles were excluded if (1) the mean age of participants was not 15 to 25 years; (2) the intervention was not a mobile app, such as other digital interventions, including therapy delivered by phone, SMS text message, video platforms, or PC (eg, computer-based cognitive behavioral therapy); (3) the apps used were not efficacious (ie, there was no significant improvement in mental health or well-being compared with a control group); (4) the apps used did not include an intervention component (ie, primarily focused on diagnosis or assessment); (5) the studies did not report mental health outcomes or the primary outcome was physical (eg, blood sugar levels or exercise); and (6) the studies did not have a control group. Gray literature was not included in the search.

Study Selection

In accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [32], the flowchart presented in Figure 1 provides step-by-step details of the study selection procedure. The search strategy identified 18,660 citations after the removal of duplicates. Of these 18,660 studies, 1634 (8.75%) were considered potentially relevant based on their titles and abstracts. Three members of the review team (HB, LAN, and JD) screened the titles and abstracts against the inclusion criteria. The remaining full texts were screened by 6 members of the review team (HB, LAN, JD, SL, BM, and LN). At this stage, 20% of the texts were screened by at least 2 reviewers independently to ensure interrater reliability (intraclass correlation coefficient range 0.88-0.96). Any disagreements between the 2 reviewers were resolved through discussion with the wider review team.

XSL•FO
RenderX

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



Data Extraction

Data were extracted by 1 reviewer (HB or LAN) and reviewed for accuracy and completeness by another. After verifying all the extracted data, discrepancies were resolved through discussion or adjudication by another party (MF). Extracted data included information on study characteristics (eg, authors, publication year, country, study design, and study population), intervention characteristics (eg, characteristics of the technology, app name, therapeutic modality, and intervention outcomes), and implementation constructs (eg, implementation objectives and implementation results). Textbox 1 provides a description of the implementation outcome criteria.

XSL•FO
RenderX

Textbox 1. Description of implementation variables.

Outcome and definition

- Coproduction: user involvement in the development or evaluation of the intervention through coproduction or another patient and public involvement activities.
- Acceptability: perception among stakeholders that a given evidence-based practice is useful, agreeable, palatable, or satisfactory.
- Appropriateness: appropriateness is the perceived fit, relevance, suitability, or compatibility of an innovation with a practice setting or context.
- Feasibility: actual fit, utility, or suitability and the extent to which an evidence-based practice can be successfully used or conducted within a given context.
- Fidelity: extent to which an evidence-based practice is being delivered as intended. This includes adherence and the quality of program delivery.
- Adoption: intention, decision, or initiation to use or uptake an evidence-based practice.
- Engagement: user enrollment, attendance, session participation, homework completion, adherence, and dropout.
- Penetration: spread, reach, and integration of an evidence-based practice in "real-world" settings.
- Implementation cost: costs associated with implementing an evidence-based practice. This includes cost-effectiveness and cost-benefit.
- Sustainability: uptake by users and the extent to which a newly implemented evidence-based practice is maintained and continued within a service setting's ongoing, stable operations.

Quality Assessment

The mixed methods appraisal tool (version 2018) was used to assess the methodological quality of the included studies [33]. It was developed by combining the core relevant methodological criteria found in different well-known and widely used qualitative and quantitative critical appraisal tools. It consists of 2 screening questions applicable to all types of study design and a further 5 questions applicable to specific study designs. Responses were rated on a categorical scale as "no," "unclear," or "yes" to any of the methodological quality criteria. Quality assessments were made by 1 of 3 reviewers (SL, BM, and LN). We did not exclude any studies based on quality assessment scores.

Data Synthesis and Analysis

The extracted data were collated and summarized to produce a narrative summary of the study, sample, and intervention characteristics. To determine the proportion of apps that were sustained or adopted after development, we contacted the corresponding author of the included articles to complete a brief survey about the development and implementation of the app described in their study (Multimedia Appendix 2). If they did not respond, the first or lead author was then contacted. In addition, we searched to check the availability and discoverability of the app in the Apple App Store (iPhone or Mac) and Google Play Store. A codebook approach was used to code and synthesize implementation data from all available sources according to the 10 implementation outcome categories [34].

Results

The systematic search identified 34 studies published between 2011 and 2021, corresponding to 29 unique apps that reported a beneficial intervention effect when compared with a control group. Figure 1 provides additional details on the screening and inclusion processes.

Study Characteristics

The characteristics of the included studies are presented in Multimedia Appendix 3 [35-68]. Of the 34 studies, 11 (32%) studies were conducted in the United States; 6 (18%) in the United Kingdom; 3 (9%) in Australia; 2 (6%) each in Italy, Japan, New Zealand, and Spain; and 1 (3%) each in Canada, Iceland, Iran, Israel, South Korea, and Sweden. Most studies (31/34, 91%) were published between 2018 and 2021. Regarding the evaluated sample populations, most studies (24/34, 71%) primarily recruited university students, followed by clinical samples (5/34, 15%), general population samples (5/34, 15%), school students (3/34, 9%), primary care patients (1/34, 3%), and those attending youth organizations (1/34, 3%). In terms of study design, 74% (25/34) of studies were randomized controlled trials (RCTs), 15% (5/34) were pilot RCTs, 6% (2/34) were quasi-experimental, and 6% (2/34) were feasibility trials. Notably, 29% (10/34) of the included studies evaluated the effectiveness of a commercially available app that was not developed by the evaluation study team.

Intervention Characteristics

Multimedia Appendix 4 [35-68] outlines the format and delivery of interventions assessed in the included studies. Of the 34 studies, 15 (44%) studies aimed to treat mental health problems or reduce symptoms, 14 (41%) aimed to promote well-being or mental health, and 5 (15%) aimed to prevent the onset of mental health problems. Most apps studied (29/34, 85%) were stand-alone, and the remainder (5/34, 15%) accompanied other therapeutic interventions.

Successful Implementation

The measurement of successful implementation, such as sustained use following the end of the research trial either commercially (eg, discoverable in app stores) or otherwise (including available to young people via schools or mental health services) varied based on unique features of the app itself, its recipients, and its context. To ascertain the proportion of apps that were sustained after development, we contacted the authors of the included articles to request additional information

XSL•FO
RenderX

about the implementation of the app reported in their study. We did not contact the authors of the 18% (6/34) of studies testing existing apps that we knew were commercially available and discoverable to the public at the point of the review (eg, Headspace and Calm); that is, being successfully sustained. Of those contacted (28/34, 82%), we collected 23 survey responses (Multimedia Appendix 5 [35-43,48,50,51,54,55,57-61,63,65-67]), and 5 authors did not respond. However, on reviewing the survey responses, 17% (4/23) were evaluating other apps already available. In the absence of survey data for 5 studies, we checked if they were commercially available; 2 (40%) were discoverable on either the Apple App Store (iPhone or Mac) or Google Play Store (Google Inc) [45,62], and 3 (60%) were not available [52,53,64].

In summary, 10 articles evaluated the effectiveness of an existing app that was available for use at the time of the evaluation study [38,44,46,47,49,54,56,59,60,68]. Of the 10 articles, 8 (80%) evaluated an app still available on the market: Headspace [44,46,47,68], Calm [49], Pacifica/Sanvello [38], Smiling Mind [47], Stop Breathe Think [56], and Thrive [59]. The 2 major smartphone app markets (ie, App Store and Google Play) publicly list app ratings out of 5 on their store pages. Google Play also provides download count estimates. The consumer app ratings on a 5-point scale, from the App Store and Google Play, respectively, are as follows: Headspace (rating 4.9 and 4.25) and Calm (rating 4.8 and 4.25) were the most popular (>10 million downloads), followed by Sanvello (rating 4.8 and 4.5), Stop Breathe Think (rating 4.8 and 4.5), Smiling Mind (rating 4.5 and 3.75; >1 million downloads), and Thrive (rating 5.0 and 3.5; >50,000 downloads). Finally, 2 studies evaluated consumer apps that are no longer available: DeStressify [54] and Lantern [60]; therefore, their store statistics are not reported.

Of the 24 studies reporting on a newly developed app, 43% (9/21) are currently available, commercially or otherwise (eg, in mental health services), and 57% (12/21) are no longer available. Most respondents reported that it took several years to develop and test the app reported in their articles, ranging from 6 months to 6 years.

Markers of Successful Implementation

Adoption

Of the 20 apps sustained after development, several are available noncommercially and freely to users in local contexts, including in mental health services [37,42], for university students [36,59], and for corporate organizations [63]. Apps are also accessible to users via commercial channels, including Apple App Store (iPhone or Mac) and Google Play Store (Multimedia Appendix 6 [35-68]).

Coproduction

A total of 9 unique apps were reported as being coproduced with young people, 5 of which are either currently available or were previously available following the study but no longer available. The level of youth involvement in the coproduction of the apps varied across studies but involved activities such as a web-based survey, which was delivered to 150 young people (ie, the target end users) [61,69]; market research and beta testing [54]; design workshops with 15 key stakeholders, followed by a series of in-depth interviews [67,70]; focus groups with young people with lived experience that guided the development of app functionalities [42]; and study groups with teenagers and young people who were involved in all developmental phases of the app [43]. Coproduction data were not reported in 12 studies.

Acceptability and Appropriateness

Acceptability was generally well assessed with a variety of measures using both qualitative and quantitative methods (k=16). Most studies of implemented apps reported them as acceptable, with high user satisfaction and ease of use among young people and health care providers [43,45,61,69]. O'Dea et al [69] examined adolescents' attitudes toward the concept of a mobile phone app for relationship help and support and reported that, overall, 60.7% (91/150) were likely to use an app for relationship problems, and this was not associated with demographics or social support (P>.05). Notably, the likelihood of app use was found to be influenced by the perceived need for help, personal beliefs about app effectiveness, and whether the app was engaging and easy to use. Overall, adolescents found the proposed app content helpful, with an average of 99.3% (149/150) rating the strategies provided as somewhat to very helpful. More than 90% of respondents reported that the app was enjoyable, easy to use and understand, and that they would recommend it to a friend [61,69]. The barriers most commonly experienced were mismatched need, forgetfulness, and being time-poor [61,69]. Acceptability was also assessed for the "Personalized Real-time Intervention for Motivational Enhancement (PRIME)" app designed to improve motivation in young people with early-onset schizophrenia during an exit interview 12 weeks after the trial. Participants rated their satisfaction with specific features of the app, such as the ability to interact with peers and the different goal categories, on a scale from 1 (not at all) to 10 (very much) [67]. Schlosser [67] reported that participants rated their overall satisfaction with "PRIME" highly. Similarly, Broglia et al [38] explored the feasibility and acceptability of supplementing college counselling with the "Pacifica" app and whether this, in turn, had positive clinical outcomes. This blended approach to their intervention was shown to be acceptable and feasible and showed the potential to maintain clinical improvement in anxiety following the completion of a brief counselling intervention [38]. Egilsson et al [43] assessed acceptability with the Systematic Usability Scale, a widely used and relatively well-studied 10-item questionnaire on app usability, where scores range from 0 to 100, and a total score of >70 indicates satisfactory usability and user acceptance [71]. The mean total score on the Systematic Usability Scale was satisfactory (mean 78.09, SD 9.82), indicating adequate usability of the app they tested to improve the emotional and physical health of adolescents [43]. Acceptability data were unavailable for 53% (18/34) of studies, and appropriateness data were unavailable for 68% (23/34) of studies.

Feasibility

The extent to which interventions were feasible (ie, the actual fit or practicality) was reported in 35% (12/34) of studies that

XSL•FO
RenderX

used a broad range of metrics as indicators of utility and suitability, including log-in frequency, app activity, average number of sessions, recruitment duration, treatment preference, the percentage of participants who completed follow-up assessments, and randomization acceptability [38,42,43,56,67]. To evaluate the feasibility of "PRIME," the authors examined the log-in frequency, challenges completed, spontaneous and goal achievement moments, peer and coach interactions, and active use rate [67]. Participants in the "Stop, Breathe and Think" trial, which evaluated a publicly available mindfulness app, provided high satisfaction ratings and reported regular use of the app, particularly in the first 2 weeks. However, the rate of recruitment was slow over the course of an academic year, indicating potential feasibility and long-term sustainability concerns [56]. Of the 34 studies, feasibility data were not available for 22 (65%) studies.

Fidelity

Fidelity, the extent to which the interventions were delivered as intended, is a less-prominent implementation determinant for apps, given the content control inherent in the structure of the delivery mechanism. However, it is relevant, for example, in one of the included studies, where fidelity outcomes were reported in transcripts from counselling audio recordings where the app was provided as an adjunct to face-to-face counselling [38]. Transcripts were scored to assess the following criteria: (1) number of times the app was discussed, (2) duration of app discussion, (3) whether therapist reviewed client app use, (4) number of app features therapist suggested, and (5) missed opportunities to discuss client app use [38].

Engagement

Engagement-related factors, including user enrollment, attendance, session participation, homework completion, adherence, study retention, and dropout, were widely reported outcomes across studies. Engagement data were reported in all but 1 study [37]. A consistent and noteworthy finding across several studies was that engagement decreased over time [43,54,67]. Unsurprisingly, young people identified that accessibility and engagement issues, including user experience, influenced their likelihood of using the intervention [61,69]. For example, engagement with 1 app dropped after the first month of the trial before leveling out over the second and third months [67]. Egilsson et al [43] noted a decrease in average exercises performed between the first week of the intervention and subsequent intervention weeks, with a significant 76% decrease in the total number of in-app health exercises from week 1 to week 2.

A total of 10 articles evaluated a consumer (ie, available on the App Store or Google Play) smartphone app at the time of their publication. Data collection methods and the degree of detail for app use differed across studies. Most studies (6/10, 60%) collected information regarding app use from participants' self-reports [38,44,47,54,59,68], followed by passive activity tracking (2/10, 20%) provided by the official app teams [46,49], combined self-report and passive tracking (1/10, 10% [56]), or not specified (1/10, 10% [60]). Self-report measures were generally 1-item data points, varying from 10-point scales between "did not use at all" and "used as often as requested" [54] to number of times per week [56,59] and yes or no daily measures [47]. In total, 2 studies requested an in-app summary screenshot of completed minutes as a record [44,68], and Yang et al [68] further provided a paper calendar for record tracking. On the other hand, passive activity tracking provided greater detail, including the date, time, and duration of an in-app exercise [46,49,56]. Newman et al [60] did not specify app use data collection methods but recorded more details of use than other studies, such as total visits on the app, number of sessions on the app, and minutes on the app. It is noteworthy that Yang et al [68] reported that individuals in the experimental group who tried Headspace at least once had an average of 12 days of use over the course of their 30-day, laissez-faire (ie, use this app as you would normally) intervention without additional prompts or ongoing accountability. In addition, 74% of those who used the app during the 30-day intervention period continued to use it for an additional 30 days. These findings imply that motivating individuals to use an app just once may be an important step in retention.

Although app use measures provide straightforward indicators of attendance, session participation, homework completion, and adherence are unclear owing to variations in their definitions. Session participation may be synonymous with attendance for some researchers, whereas others operationalize it as meaningful, back-and-forth interactions between a user and agent (eg, app, coach, or counselor). A total of 20% (2/10) of studies reporting on consumer apps included an agent (therapists [38] and coaches [60]) as part of their intervention, but only 1 analyzed whether user-agent communication (ie, session participation) affected efficacy. Newman et al [60] reported "treatment usage" (a composite variable including user-agent interaction), which did not affect symptom outcomes. In addition, homework completion may be defined as completing the prescribed exercises in a study, whereas another study may define it as tasks in addition to the primary intervention. None of the included studies mentioned "session participation" or "homework." In terms of adherence, 50% (5/10) of studies explicitly measured adherence (eg, app use data, including time spent and sessions using the app) [46,47,49,54,59]. An additional study [56] did not operationalize "adherence," although they collected characteristic adherence data through self-report, asking users how many days in the past 2 weeks they had used the app. However, given the lack of conceptual clarity and consensus on adherence, it is difficult to assess whether adherence affects clinical health outcomes in the context of consumer app evaluation.

Implementation Cost

Data on the costs associated with implementing the apps, including cost-effectiveness and cost-benefit, are not readily available in the public domain. However, results from the survey indicate that, unsurprisingly, apps developed from scratch were the costliest to develop and test. Notably, several apps included in this review were developed from existing, ready-made platforms, which were modified for different interventions rather than building the app from the ground up. For example, the GGtude platform was developed in 2016 and hosts several apps designed for different populations and presenting problems [35,40,41,66]. This approach has been successful with several

XSL•FO
RenderX

studies reporting that daily use of apps from the GGtude platform during a period of 2 weeks (3 minutes a day) is associated with significant beneficial effects on mental health in nonclinical and subclinical samples [40,41,66]. Similarly, Levin et al [55] have focused their efforts on developing app prototypes rapidly using easy-to-use website platforms to customize intervention content and develop generalizable knowledge about principles and processes that work in mobile apps rather than building and developing new apps from scratch [55,57]. However, the authors note that their platform was not set up for commercialization or broad deployment because the app is delivered within the LifeData system and lacks key features needed for a public launch, including budget and support for ongoing technical maintenance and monitoring user inquiries and data. It was also noted that ongoing updates are likely needed to remain relevant and competitive with other market products, which involves undertaking regular market scans and content refinement to ensure the product remains well positioned and effective in an increasingly saturated market [61].

Another approach taken by the authors of the included studies was to evaluate existing, publicly available apps for a specific local context and also as a potential adjunct to existing in-person therapies (eg, university students and college counselling centers) [38,54,56,59,60]. For example, Broglia et al [38] contacted the developers of a mood monitoring app to test its use in conjunction with usual care counselling sessions. Similarly, Levin et al [56] conducted a pilot RCT to evaluate the feasibility and acceptability of a publicly available mindfulness app for university students. Using existing, commercially available apps or building prototypes using adaptable web platforms provides a lower cost and quicker alternative to developing and evaluating new apps from the ground up. However, implementing apps in distinct local contexts requires well-thought-out and tailored implementation strategies, with consideration given to common barriers, especially acceptability and feasibility.

Sustainability and Penetration

Sustainability can be measured in several ways, including apps being fully integrated into service settings and a steady budget for app advertising, maintenance, and updates [21]. The number of app downloads and interactions over time also provides an indication of sustained uptake over time. However, none of the included studies reported data on the sustainability of the interventions evaluated in their study. As such, the literature is sparse regarding the long-term integration and penetration of mobile interventions within mental health and other support service settings.

Study Quality

Methodological quality varied across the included studies (Multimedia Appendix 7 [35-68]). Most studies (32/34, 94%) were judged as having possible limitations in at least 1 criterion. Most studies (29/34, 85%) clearly described the randomization of the study participants or the process for recruiting a representative sample. Most randomized trials reported complete outcome data (30/32, 94%) and described samples that were comparable at baseline (27/32, 84%). However, few studies described the process used to blind the outcome assessor to the intervention group (10/32, 31%). Only half of the studies (16/32, 50%) reported acceptable adherence rates to the intervention. Regarding the 2 nonrandomized trials, both studies used appropriate outcome measures. Of the 2 studies, 1 (50%) was judged as not clearly accounting for confounders in the study design and analysis. Both studies were judged as adhering to the intervention protocol [38,53].

Discussion

Principal Findings

The primary aim of this systematic review was to determine the proportion of evidence-based mental health and well-being apps that have been successfully adopted and sustained in "real-world" settings. In total, 29% (10/34) of studies were identified that evaluated the effectiveness of 8 existing, commercially available mental health apps. The remaining 71% (24/34) of included studies evaluated 21 newly developed apps, of which 43% (9/21) are currently available, commercially or otherwise (eg, in mental health services), and 57% (12/21) were no longer available at the time of enquiry. Therefore, these results not only indicate a 43% implementation success rate of new apps but also provide some information on how existing efficacious commercial apps are promoted and sustained in the field.

Broadly synthesized using 10 dimensions of implementation, our review suggests that measures of adoption, acceptability, appropriateness, and feasibility are more frequently reported than indicators of cost, fidelity, sustainability, and penetration. Implementation outcomes were unavailable for many of the studies, precluding direct comparisons between those apps that were implemented and those that were not across constructs. To partially address our second research question, most of the apps that had been implemented confirmed a degree of adoption, acceptability, appropriateness, feasibility, and engagement. These determinants are relevant to a range of interventions and would benefit from broad systematic incorporation into the development of a smartphone app. Other important factors identified included the coproduction of interventions with young people and the need to embed apps within local settings, such as schools, universities, and mental health services, rather than relying on commercial strategies. Although assessing implementation outcomes using an integrated framework allows for a more systematic assessment of outcomes, the review highlights a lack of measurement precision around implementation constructs in that an array of overlapping terms, such as acceptability and usability, are often used interchangeably, highlighting the need for greater consensus on how to measure and report implementation determinants. There is also a need to elucidate the relationship between different implementation constructs (eg, acceptability and engagement) and intervention effectiveness (eg, clinical change) [22]. Modeling these relationships was not possible in the current review because of data quality issues. Further empirical research is needed to model the interrelationships between implementation variables to better understand the nature of their connections and their impact on implementation success.

XSL•FO
RenderX

Confirming acceptability and engagement in the early phases of intervention development, before implementation commences, is an important step in ensuring that the intervention has potential longevity and may be important for saving time and money. As evidence suggests that engagement is the highest in the first days or weeks of downloading an app, it might be more appropriate for apps to be developed with this anticipated behavior in mind [72].

Research has identified that intervention- and person-specific factors that influence engagement with mobile mental health interventions should be considered [25]. For example, getting individuals to use an app just once is an important step in retention. It is likely that an initial hook is needed to catch the attention of young people to encourage them to identify and download the app. Thereafter, rewards and gamification are reported by young people as motivating factors for ongoing engagement and retention [25]. Similarly, usability has been reported as important for promoting engagement. Acceptability and engagement are critical elements of mental health apps that support users [73]. However, it is recognized that operationalizing meaningful app engagement is not straightforward, as many downloaded apps are never used and more work is needed to define sustained engagement, what leads to it, and how to create products that achieve it [73,74]. Further research may explore characteristics facilitating initial use, continued use, and the implications of changing the treatment structure during an intervention (eg, Do individuals find it motivating or overwhelming?).

Barriers to Successful Implementation

The importance of identifying mental health interventions that are efficacious needs to progress in concert with addressing implementation requirements from the outset of app development. This can be a complicated focus for researchers and funding bodies, as in the absence of efficacy, it might seem premature to address dissemination issues, but given the high redundancy of scientific studies—less than half of the efficacious apps were actually identified as being in sustained use—these questions must be paramount at the outset of the research cycle.

In addition to the markers of successful implementation, there are several extant barriers, including high cost, time, funding constraints, and lengthy research processes. The process of applying for funding, app development, data collection, data analysis, and app release is lengthy and expensive, taking up to 6 years [39]. The sustainability and penetration of apps is contingent on projects being funded or grants being successful, and even when funding is secured, it is often time limited [42,67,75]. The way in which funding streams are set up means that it is harder to secure funding for the follow-on implementation research than for the initial evaluation project.

Other challenges include the continuous, rapidly evolving development of technology, which results in the need for continued code updates and upgrades to allow for compatibility with smartphone operating systems [42,61]. Additional noted barriers include the rapidly changing nature of many commercial providers supplying the app technology, sometimes with significant staff movement, changes to the focus of work depending on commercial drivers, and short life spans of some tech companies [60]. Given that only a small proportion of existing, commercially based apps are being well sustained, rather than developing de novo apps, it might be that evaluating more successful apps and adapting them to specific contexts can propel implementation in the field. However, it is recognized that the commercial app marketplace is highly competitive, and research has identified that, based on our estimates of monthly active users, mindfulness and meditation apps appear to be the most popular: Headspace and Calm account for 13.4 million users, Replika and Wysa account for 1.5 million users, and Reflectly and Daylio account for 840,000 users [76]. In total, these 6 apps monopolize the marketplace, many of which have been made freely available to young people and extensively marketed; hence, they penetrate the market and account for 83% of the monthly active users of mental health and wellness apps [76]. As noted by others, currently, the mental health and well-being apps that have been rigorously evaluated struggle to attract users, and apps with many users are rarely evaluated [76,77].

There are several possible explanations for the success of these commercial apps. Fish and Saul [44] reported that Headspace, one of the most popular consumer apps in the market, applies gamification techniques, such as accomplishment, empowerment, social influence, and ownership to improve engagement and motivate individuals to enjoy meditations and find them rewarding. Several engagement design features in Headspace, such as user tracking, reminder function, and push notifications, can improve engagement, these features were not actively part of any of the evaluated studies and were left for use at the discretion of the individuals [46]. However, Huberty et al [49] report that these features may be critical for engagement and acceptability.

Groups at Risk of Exclusion

An interesting finding of this review is that there are certain groups that are less likely to access mental health and well-being apps. It is noteworthy that most studies included in this review recruited university students (24/34, 71%). There was a marked absence of youth samples from underserved or marginalized populations, including but not limited to migrants, asylum seekers and refugees, those experiencing homelessness, and those from socioeconomically deprived backgrounds. These potentially high-risk groups are typically underrepresented in research, face access and engagement barriers when navigating health care systems, and experience digital exclusion [78]. The studies included in this systematic review had not targeted these populations, and research exploring the acceptability, appropriateness, and feasibility of app-based interventions is lacking.

Limitations

Although this review was rigorous, carefully executed, and used a robust methodological approach, it was not without limitations. Foremost, although the review team attempted to identify and include as many articles as possible, some articles may have been missed because of the inconsistencies in how implementation outcomes are recorded and reported. It was also difficult to ensure that all apps for this age group were identified

XSL•FO
RenderX

because those aged between 15 and 25 years are harder to differentiate in adolescent and adult studies, meaning we might have missed some relevant studies. Although the focus of this review was young people aged between 15 and 25 years, most of the included studies recruited university students, and most of the study samples had a mean age of >20 years. This finding suggests that there may be a lack of effective mental health apps targeted at mid to late teenagers. This may be a result of the ethical constraints of recruiting young people aged <18 years (16 years in the United Kingdom) to research studies, as it is easier to recruit those who are deemed able to independently consent to trial participation. Additional research is needed to determine whether mental health apps have been developed for this group; however, upon evaluation, they were deemed ineffective. Given the considerable heterogeneity in the social, emotional, and cognitive development and maturity of young people in this age range, research is needed to understand the potential differential impacts and utility of apps throughout adolescence and young adulthood.

In addition, the review included several apps that were delivered as an adjunct to in-person therapies. It is recognized that the reported successful implementation of these apps may in fact be a by-product of the success of the primary therapeutic intervention rather than the app itself. Research comparing face-to-face therapies with and without additional app support provides evidence to support the superiority of adjunctive interventions compared with standard intervention-only conditions [79]. However, additional research is needed to enable us to determine the role of apps in the successful implementation and sustainability of blended intervention programs.

Unpublished data were not included in the search, which may have affected the results of this review. Nevertheless, this approach was also seen as a further strength by ensuring that only peer-reviewed interventions were included. Given that the focus of our review was on apps that had been found to be efficacious, the publication bias effect was minimal. However, it is possible that companies that have developed the apps possess valuable data pertaining to some aspects of implementation. Commercial companies are likely to have collected a wealth of relevant and informative data on user engagement and app use over several years. However, owing to competing commercial interests and a lack of regulation in the mobile mental health app arena, these data are not publicly available or made openly accessible to researchers [80]. Nonetheless, an attempt was made to obtain implementation data when it was not included in the articles by contacting the study authors, most of whom responded.

Finally, more than two-thirds of the apps were not independently evaluated and, therefore, an important consideration is that "developer bias" may have impacted the analysis and subsequent findings of those trials. More independent and robust evaluation of apps is needed, and findings must be shared in a manner that is accessible to the scientific community as well as the users of the app. This will require time and resources but needs to become integral to the development process to mitigate potential bias in evaluations moving forward [81], be that within purely commercial or research contexts or for commercial-research collaborations.

Recommendations for Research and Practice

Coproduction With Young People

Coproducing interventions with young people may help improve the acceptability and feasibility of the end product, which, in turn, can improve intervention effectiveness [25,26]. Coproduction actively involves relevant stakeholder groups in the design process to help ensure that the technology developed meets their needs and is usable. In coproduction, intended users work with designers, developers, and researchers during the innovation and development process [82]. A range of methods have been used to enact young people's involvement in health research, often under the umbrella of "Young People's Advisory Groups" (YPAGs) [83]. This includes, but is not limited to, market research and beta testing, design workshops, in-depth interviews, and focus groups. Consistent reporting on the methods of involvement and outputs of YPAGs in publications will help develop a better understanding of the influence of YPAGs in youth mental health research, enabling better systems for meaningful youth involvement in research [83,84].

Cost-effective Prototyping

To address the lengthy and costly research and evaluation processes, adaptable and modifiable interventions can be developed from existing platforms, which can then be delivered quickly in a personalized way to meet a range of users' needs. As noted by others, this is complicated by the need to balance scientific rigor with the fast pace with which technology advances to achieve the adoption of evidence-based practice [85]. To address some of the time, funding, and financial burdens, it may be that researchers and clinicians work collaboratively with industry partners to capitalize on existing interventions and platforms, modifying and evaluating them for their local context. However, it is acknowledged that working with developers involves its own set of complex challenges. Often, researchers invest time and resources in the initial stages in building prototypes with developers only to find out that the intervention is ineffective, not feasible, or not economically viable. As an alternative, researchers can take advantage of free or low-cost systems for rapid prototyping, such as Qualtrics, or an Ecological Momentary Assessment platform, such as LifeData or MetricWire.

Assess and Report Implementation Outcomes

Intervention evaluation should include key implementation determinants, such as acceptability, feasibility, appropriateness, and cost. Currently, implementation outcomes are poorly reported, and publications reporting the results of RCTs focus on clinical outcomes, often neglecting sustainability, cost, and the process of embedding the intervention into "real-world" clinical practice. It has been reported that psychological intervention articles report, at most, 64% of the information needed to implement interventions [86]. It is important that researchers use consistent implementation measurement and reporting to allow meaningful and accurate comparisons across studies and share the information required by implementers. Assessing implementation outcomes using a conceptually

XSL•FO
RenderX

grounded framework allows for a more systematic assessment of outcomes while supporting the rigor and reproducibility of implementation research and providing the building blocks for implementation evaluation.

Hybrid Intervention Models

There has been increased attention toward "blended" approaches, where web-based support is provided as an adjunct to, rather than a replacement for, face-to-face treatment. The importance of human support in web-based therapies and the perceived value of blending mobile health interventions with traditional face-to-face treatment is well described in the literature [87-89]. A blended approach has the potential to reduce the relative load of costly face-to-face contact while boosting engagement, enhancing outcomes, and increasing treatment acceptability [88].

Early Consideration of Economic Sustainability

To support the implementation of evidence-based and efficacious interventions, early consideration of funding and costs are crucial. Funding bodies should endeavor to develop new, responsive funding streams, including implementation-specific grants, to focus on the implementation of existing evidence-based interventions, as the translation of evidence from feasibility to adoption is poorly realized, bringing considerable redundancy to the field of intervention research. Currently, it is harder for researchers to secure funding for follow-on work than for the initial evaluation project, and it is not possible to secure funding ahead of time for this in the initial app, as they are usually feasibility studies, and so effectiveness cannot be assumed. However, services are unable to adopt evidence-based practices until implementation drivers are tested and addressed. Furthermore, it is important that a clear business model is planned and in place, with consideration given to the potential market and how the implementation of the product will take place and any anticipated revenue generation [85].

Involvement of Vulnerable Groups

As in other areas of mental health research, young people from marginalized and underserved groups (eg, from low-income backgrounds, refugees, or asylum seekers; those not in education, employment, or training; members of ethnic and sexual minorities; and those under state care) were underrepresented in these studies, which typically focused on university students. Few attempts have been made in the literature to create new interventions or to adapt existing ones to meet the complex and heterogeneous needs of these young people [90]. Research exploring the acceptability, appropriateness, and feasibility of mental health app–based interventions for this group is lacking [89]. Work to assess the acceptability and feasibility of mental health apps for underserved young people is needed to ensure that they are not further excluded from research and to advance toward mental health provision that meets their support needs.

Conclusions

Despite the significant amount of funding that has been directed toward the development of mobile mental health interventions, few have published evidence-based data to support their use in real-world settings, and even fewer have been successfully transitioned into sustainable mental health interventions. Although it had been thought that smartphone apps held the potential to address many of the current issues facing service provision in youth mental health by improving the scalability and affordability of evidence-based mental health interventions for young people and addressing health disparities by providing wider access to underserved populations, more work is needed to improve key implementation drivers, such as uptake and adoption. Innovative and targeted funding mechanisms that are quick, responsive, and encouraging of broad stakeholder and industry partnerships, where data are openly shared, are essential to ensure mental health and well-being app development, evaluation, implementation, and sustainability proceeds in a direction that will enable evidence-based interventions to be made available quickly to young people who may benefit from them.

Acknowledgments

This project received funding from the European Union's Horizon 2020 research and innovation program (grant 754657).

The authors are grateful to those authors who took the time to complete the survey in this review, including, but not limited to, Professor Michael E Levin, Jennifer Krafft, Professor Guy Doron, Dinur Aboody, Dr Bridianne O'Dea, Dr Erlendur Egilsson, Dr Emma Broglia, Dr Tayla McCloud, Dr Silvia Cerea, Professor Maria Roncero, Dr Martina Di Simplicio, Professor Sandra Bucci, Dr Sung-Won Choi, Dr Ji-Won Hur, Dr Somaieh Borjalilu, Professor Michelle G Newman, Dr Kristin Thomas, Dr Baland Jalal, Dr David Plans, Professor Danielle Schlosser, Dr Mary Elizabeth Jung, and Dr Lynnette Nathalie Lyzwinski. The authors would also like to thank Verity Westgate and Professor Andrea Cipriani for their helpful and insightful feedback on an earlier version of this manuscript.

Data Availability

The data extracted to support the findings of this review can be obtained from the corresponding author on reasonable request.

XSL•FO
RenderX

Conflicts of Interest

SL is a consultant at Paradym Ltd, an organization that developed Paradym, an app for supporting well-being (not included in the current review).

Multimedia Appendix 1
Systematic search strategy.
[DOCX File , 21 KB - jmir_v24i11e40347_app1.docx]

Multimedia Appendix 2
Author survey questions.
[DOCX File , 16 KB - jmir_v24i11e40347_app2.docx]

Multimedia Appendix 3
Study characteristics.
[DOCX File , 24 KB - jmir_v24i11e40347_app3.docx]

Multimedia Appendix 4
Intervention characteristics.
[DOCX File , 21 KB - jmir_v24i11e40347_app4.docx]

Multimedia Appendix 5
Implementation and sustainability.
[DOCX File , 19 KB - jmir_v24i11e40347_app5.docx]

Multimedia Appendix 6
Implementation variables.
[DOCX File , 21 KB - jmir_v24i11e40347_app6.docx]

Multimedia Appendix 7
Mixed methods appraisal tool.
[XLSX File (Microsoft Excel File), 81 KB - jmir_v24i11e40347_app7.xlsx]

References

1. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 01;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
2. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]
3. Muñoz RF. The efficiency model of support and the creation of digital apothecaries. Clin Psychol Sci Pract 2017 Mar;24(1):46-49. [doi: 10.1111/cpsp.12174]
4. Clayborne ZM, Varin M, Colman I. Systematic review and meta-analysis: adolescent depression and long-term psychosocial outcomes. J Am Acad Child Adolesc Psychiatry 2019 Jan;58(1):72-79. [doi: 10.1016/j.jaac.2018.07.896] [Medline: 30577941]
5. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
6. Aschbrenner KA, Naslund JA, Tomlinson EF, Kinney A, Pratt SI, Brunette MF. Adolescents' use of digital technologies and preferences for mobile health coaching in public mental health settings. Front Public Health 2019 Jul 2;7:178 [FREE Full text] [doi: 10.3389/fpubh.2019.00178] [Medline: 31312629]
7. Leech T, Dorstyn D, Taylor A, Li W. Mental health apps for adolescents and young adults: a systematic review of randomised controlled trials. Child Youth Serv Rev 2021 Aug;127:106073. [doi: 10.1016/j.childyouth.2021.106073]
8. Buttazzoni A, Brar K, Minaker L. Smartphone-based interventions and internalizing disorders in youth: systematic review and meta-analysis. J Med Internet Res 2021 Jan 11;23(1):e16490 [FREE Full text] [doi: 10.2196/16490] [Medline: 33427682]
9. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 08;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]

XSL•FO
RenderX

10. Lipschitz J, Hogan TP, Bauer MS, Mohr DC. Closing the research-to-practice gap in digital psychiatry: the need to integrate implementation science. J Clin Psychiatry 2019 May 14;80(3):18com12659 [FREE Full text] [doi: 10.4088/JCP.18com12659] [Medline: 31091029]
11. Hollis C. Youth mental health: risks and opportunities in the digital world. World Psychiatry 2022 Feb;21(1):81-82 [FREE Full text] [doi: 10.1002/wps.20929] [Medline: 35015352]
12. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
13. Torous J, Haim A. Dichotomies in the development and implementation of digital mental health tools. Psychiatr Serv 2018 Dec 01;69(12):1204-1206 [FREE Full text] [doi: 10.1176/appi.ps.201800193] [Medline: 30256182]
14. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: exploring adolescents' perspectives. Health Informatics J 2016 Jun;22(2):265-275 [FREE Full text] [doi: 10.1177/1460458214555041] [Medline: 25385165]
15. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
16. Temkin AB, Schild J, Falk A, Bennett SM. Mobile apps for youth anxiety disorders: a review of the evidence and forecast of future innovations. Prof Psychol Res Pract 2020 Aug;51(4):400-413. [doi: 10.1037/pro0000342]
17. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
18. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
19. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
20. Connolly SL, Hogan TP, Shimada SL, Miller CJ. Leveraging implementation science to understand factors influencing sustained use of mental health apps: a narrative review. J Technol Behav Sci 2021;6(2):184-196 [FREE Full text] [doi: 10.1007/s41347-020-00165-4] [Medline: 32923580]
21. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]
22. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. eMental healthcare technologies for anxiety and depression in childhood and adolescence: systematic review of studies reporting implementation outcomes. JMIR Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]
23. Ellis LA, Augustsson H, Grødahl AI, Pomare C, Churruca K, Long JC, et al. Implementation of e-mental health for depression and anxiety: a critical scoping review. J Community Psychol 2020 Apr;48(3):904-920. [doi: 10.1002/jcop.22309] [Medline: 31944324]
24. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
25. Liverpool S, Mota CP, Sales CM, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
26. Moltrecht B, Patalay P, Bear HA, Deighton J, Edbrooke-Childs J. A transdiagnostic, emotion regulation app (Eda) for children: design, development, and lessons learned. JMIR Form Res 2022 Jan 19;6(1):e28300 [FREE Full text] [doi: 10.2196/28300] [Medline: 35044312]
27. Kien C, Griebler U, Schultes M, Thaler KJ, Stamm T. Psychometric testing of the German versions of three implementation outcome measures. Glob Implement Res Appl 2021 Aug 23;1(3):183-194. [doi: 10.1007/s43477-021-00019-y]
28. Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013 Nov 27;2013(11):CD010013 [FREE Full text] [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
29. Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27;2015(7):MR000042 [FREE Full text] [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]
30. Falconer J. Removing duplicates from an EndNote library. Library, Archives & Open Research Services Blog. 2018 Dec 7. URL: http://blogs.lshtm.ac.uk/library/2018/12/07/removing-duplicates-from-an-endnote-library/ [accessed 2021-02-02]
31. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a Web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
32. 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]

XSL•FO
RenderX

33. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
34. Crabtree BE, Miller WF. A template approach to text analysis: developing and using codebooks. In: Crabtree BF, Miller WL, editors. Doing Qualitative Research. Thousand Oaks, CA, USA: Sage Publications; 1999:163-177.
35. Aboody D, Siev J, Doron G. Building resilience to body image triggers using brief cognitive training on a mobile application: a randomized controlled trial. Behav Res Ther 2020 Nov;134:103723. [doi: 10.1016/j.brat.2020.103723] [Medline: 32920164]
36. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208 [FREE Full text] [doi: 10.2196/17208] [Medline: 32196462]
37. Borjalilu S, Mazaheri MA, Talebpour A. Effectiveness of mindfulness-based stress management in the mental health of Iranian university students: a comparison of blended therapy, face-to-face sessions, and mHealth app (Aramgar). Iran J Psychiatry Behav Sci 2019 May 12;13(2):e84726. [doi: 10.5812/ijpbs.84726]
38. Broglia E, Millings A, Barkham M. Counseling with guided use of a mobile well-being app for students experiencing anxiety or depression: clinical outcomes of a feasibility trial embedded in a student counseling service. JMIR Mhealth Uhealth 2019 Aug 15;7(8):e14318 [FREE Full text] [doi: 10.2196/14318] [Medline: 31418424]
39. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]
40. Cerea S, Ghisi M, Bottesi G, Carraro E, Broggio D, Doron G. Reaching reliable change using short, daily, cognitive training exercises delivered on a mobile application: the case of Relationship Obsessive Compulsive Disorder (ROCD) symptoms and cognitions in a subclinical cohort. J Affect Disord 2020 Nov 01;276:775-787. [doi: 10.1016/j.jad.2020.07.043] [Medline: 32738662]
41. Cerea S, Ghisi M, Bottesi G, Manoli T, Carraro E, Doron G. Cognitive behavioral training using a mobile application reduces body image-related symptoms in high-risk female university students: a randomized controlled study. Behav Ther 2021 Jan;52(1):170-182. [doi: 10.1016/j.beth.2020.04.002] [Medline: 33483115]
42. Di Simplicio M, Appiah-Kusi E, Wilkinson P, Watson P, Meiser-Stedman C, Kavanagh DJ, et al. Imaginator: a proof-of-concept feasibility trial of a brief imagery-based psychological intervention for young people who self-harm. Suicide Life Threat Behav 2020 Jun;50(3):724-740 [FREE Full text] [doi: 10.1111/sltb.12620] [Medline: 32057131]
43. Egilsson E, Bjarnason R, Njardvik U. Usage and weekly attrition in a smartphone-based health behavior intervention for adolescents: pilot randomized controlled trial. JMIR Form Res 2021 Feb 17;5(2):e21432 [FREE Full text] [doi: 10.2196/21432] [Medline: 33481750]
44. Fish MT, Saul AD. The gamification of meditation: a randomized-controlled study of a prescribed mobile mindfulness meditation application in reducing college students' depression. Simul Gaming 2019 Jun 04;50(4):419-435. [doi: 10.1177/1046878119851821]
45. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
46. Flett JA, Conner TS, Riordan BC, Patterson T, Hayne H. App-based mindfulness meditation for psychological distress and adjustment to college in incoming university students: a pragmatic, randomised, waitlist-controlled trial. Psychol Health 2020 Sep;35(9):1049-1074. [doi: 10.1080/08870446.2019.1711089] [Medline: 32046499]
47. Flett JA, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness (N Y) 2019;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
48. Franklin JC, Fox KR, Franklin CR, Kleiman EM, Ribeiro JD, Jaroszewski AC, et al. A brief mobile app reduces nonsuicidal and suicidal self-injury: evidence from three randomized controlled trials. J Consult Clin Psychol 2016 Jun;84(6):544-557. [doi: 10.1037/ccp0000093] [Medline: 27018530]
49. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
50. Hur JW, Kim B, Park D, Choi SW. A scenario-based cognitive behavioral therapy mobile app to reduce dysfunctional beliefs in individuals with depression: a randomized controlled trial. Telemed J E Health 2018 Sep;24(9):710-716. [doi: 10.1089/tmj.2017.0214] [Medline: 29323626]
51. Jalal B, Brühl A, O'Callaghan C, Piercy T, Cardinal RN, Ramachandran VS, et al. Novel smartphone interventions improve cognitive flexibility and obsessive-compulsive disorder symptoms in individuals with contamination fears. Sci Rep 2018 Oct 23;8(1):14923 [FREE Full text] [doi: 10.1038/s41598-018-33142-2] [Medline: 30353111]
52. Kageyama K, Kato Y, Mesaki T, Uchida H, Takahashi K, Marume R, et al. Effects of video viewing smartphone application intervention involving positive word stimulation in people with subthreshold depression: a pilot randomized controlled trial. J Affect Disord 2021 Mar 01;282:74-81. [doi: 10.1016/j.jad.2020.12.104] [Medline: 33401126]

XSL•FO
RenderX

53. Kajitani K, Higashijima I, Kaneko K, Matsushita T, Fukumori H, Kim D. Short-term effect of a smartphone application on the mental health of university students: a pilot study using a user-centered design self-monitoring application for mental health. PLoS One 2020 Sep 25;15(9):e0239592 [FREE Full text] [doi: 10.1371/journal.pone.0239592] [Medline: 32976515]
54. Lee RA, Jung ME. Evaluation of an mHealth app (DeStressify) on university students' mental health: pilot trial. JMIR Ment Health 2018 Jan 23;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
55. Levin ME, Krafft J, Seifert S, Lillis J. Tracking valued and avoidant functions with health behaviors: a randomized controlled trial of the acceptance and commitment therapy matrix mobile app. Behav Modif 2022 Jan;46(1):63-89. [doi: 10.1177/0145445520913987] [Medline: 32844679]
56. Levin ME, Hicks ET, Krafft J. Pilot evaluation of the stop, breathe and think mindfulness app for student clients on a college counseling center waitlist. J Am Coll Health 2022 Jan;70(1):165-173. [doi: 10.1080/07448481.2020.1728281] [Medline: 32150519]
57. Levin ME, Haeger J, An W, Twohig MP. Comparing cognitive defusion and cognitive restructuring delivered through a mobile app for individuals high in self-criticism. Cogn Ther Res 2018 Jul 13;42(6):844-855. [doi: 10.1007/s10608-018-9944-3]
58. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. The mindfulness app trial for weight, weight-related behaviors, and stress in university students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Apr 10;7(4):e12210 [FREE Full text] [doi: 10.2196/12210] [Medline: 30969174]
59. McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E. Effectiveness of a mobile app intervention for anxiety and depression symptoms in university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e15418 [FREE Full text] [doi: 10.2196/15418] [Medline: 32735221]
60. Newman MG, Jacobson NC, Rackoff GN, Bell MJ, Taylor CB. A randomized controlled trial of a smartphone-based application for the treatment of anxiety. Psychother Res 2021 Apr;31(4):443-454 [FREE Full text] [doi: 10.1080/10503307.2020.1790688] [Medline: 32662323]
61. O'Dea B, Han J, Batterham PJ, Achilles MR, Calear AL, Werner-Seidler A, et al. A randomised controlled trial of a relationship-focussed mobile phone application for improving adolescents' mental health. J Child Psychol Psychiatry 2020 Aug;61(8):899-913 [FREE Full text] [doi: 10.1111/jcpp.13294] [Medline: 32683737]
62. Orosa-Duarte Á, Mediavilla R, Muñoz-Sanjose A, Palao Á, Garde J, López-Herrero V, et al. Mindfulness-based mobile app reduces anxiety and increases self-compassion in healthcare students: a randomised controlled trial. Med Teach 2021 Jun;43(6):686-693. [doi: 10.1080/0142159X.2021.1887835] [Medline: 33645416]
63. Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D. Efficacy of the digital therapeutic mobile app BioBase to reduce stress and improve mental well-being among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 06;8(4):e17767 [FREE Full text] [doi: 10.2196/17767] [Medline: 31926063]
64. Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. BMC Fam Pract 2011 Nov 29;12:131 [FREE Full text] [doi: 10.1186/1471-2296-12-131] [Medline: 22123031]
65. Rodgers RF, Donovan E, Cousineau T, Yates K, McGowan K, Cook E, et al. BodiMojo: efficacy of a mobile-based intervention in improving body image and self-compassion among adolescents. J Youth Adolesc 2018 Jul;47(7):1363-1372. [doi: 10.1007/s10964-017-0804-3] [Medline: 29349593]
66. Roncero M, Belloch A, Doron G. Can brief, daily training using a mobile app help change maladaptive beliefs? Crossover randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 13;7(2):e11443 [FREE Full text] [doi: 10.2196/11443] [Medline: 30758294]
67. Schlosser DA, Campellone TR, Truong B, Etter K, Vergani S, Komaiko K, et al. Efficacy of PRIME, a mobile app intervention designed to improve motivation in young people with schizophrenia. Schizophr Bull 2018 Aug 20;44(5):1010-1020 [FREE Full text] [doi: 10.1093/schbul/sby078] [Medline: 29939367]
68. Yang E, Schamber E, Meyer RM, Gold JI. Happier healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complement Med 2018 May;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
69. O'Dea B, Achilles MR, Werner-Seidler A, Batterham PJ, Calear AL, Perry Y, et al. Adolescents' perspectives on a mobile app for relationships: cross-sectional survey. JMIR Mhealth Uhealth 2018 Mar 08;6(3):e56 [FREE Full text] [doi: 10.2196/mhealth.8831] [Medline: 29519775]
70. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: a cognitive neuroscience-informed mobile app intervention to enhance motivated behavior and improve quality of life in recent onset schizophrenia. JMIR Res Protoc 2016 Apr 28;5(2):e77 [FREE Full text] [doi: 10.2196/resprot.5450] [Medline: 27125771]
71. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
72. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

XSL•FO
RenderX

73. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766 [FREE Full text] [doi: 10.2196/30766] [Medline: 35099398]
74. Lattie EG, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME, et al. Uptake and usage of IntelliCare: a publicly available suite of mental health and well-being apps. Internet Interv 2016 May;4(2):152-158 [FREE Full text] [doi: 10.1016/j.invent.2016.06.003] [Medline: 27398319]
75. Levin ME, Haeger J, Cruz RA. Tailoring acceptance and commitment therapy skill coaching in the moment through smartphones: results from a randomized controlled trial. Mindfulness (N Y) 2018 Jul 21;10(4):689-699. [doi: 10.1007/s12671-018-1004-2]
76. Wasil AR, Palermo EH, Lorenzo-Luaces L, DeRubeis RJ. Is there an app for that? A review of popular apps for depression, anxiety, and well-being. Cogn Behav Pract 2022 Nov;29(4):883-901. [doi: 10.1016/j.cbpra.2021.07.001]
77. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
78. Robards F, Kang M, Usherwood T, Sanci L. How marginalized young people access, engage with, and navigate health-care systems in the digital age: systematic review. J Adolesc Health 2018 Apr;62(4):365-381. [doi: 10.1016/j.jadohealth.2017.10.018] [Medline: 29429819]
79. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
80. Singh S, Sagar R. Time to Have Effective Regulation of the Mental Health Apps Market: Maximize Gains and Minimize Harms. Indian J Psychol Med 2022 Jul;44(4):399-404 [FREE Full text] [doi: 10.1177/02537176221082902] [Medline: 35949633]
81. Wolf R, Morrison J, Inns A, Slavin R, Risman K. Average Effect Sizes in Developer-Commissioned and Independent Evaluations. Journal of Research on Educational Effectiveness 2020 Apr 06;13(2):428-447. [doi: 10.1080/19345747.2020.1726537]
82. Hill C, Creswell C, Vigerland S, Nauta MH, March S, Donovan C, et al. Navigating the development and dissemination of internet cognitive behavioral therapy (iCBT) for anxiety disorders in children and young people: a consensus statement with recommendations from the #iCBTLorentz Workshop Group. Internet Interv 2018 Jun;12:1-10 [FREE Full text] [doi: 10.1016/j.invent.2018.02.002] [Medline: 30135763]
83. Sellars E, Pavarini G, Michelson D, Creswell C, Fazel M. Young people's advisory groups in health research: scoping review and mapping of practices. Arch Dis Child 2021 Jul;106(7):698-704. [doi: 10.1136/archdischild-2020-320452] [Medline: 33208398]
84. Bevan Jones R, Stallard P, Agha SS, Rice S, Werner-Seidler A, Stasiak K, et al. Practitioner review: co-design of digital mental health technologies with children and young people. J Child Psychol Psychiatry 2020 Aug;61(8):928-940 [FREE Full text] [doi: 10.1111/jcpp.13258] [Medline: 32572961]
85. Hill C, Martin JL, Thomson S, Scott-Ram N, Penfold H, Creswell C. Navigating the challenges of digital health innovation: considerations and solutions in developing online and smartphone-application-based interventions for mental health disorders. Br J Psychiatry 2017 Aug;211(2):65-69. [doi: 10.1192/bjp.bp.115.180372] [Medline: 28522435]
86. Premachandra B, Lewis Jr NA. Do we report the information that is necessary to give psychology away? A scoping review of the psychological intervention literature 2000-2018. Perspect Psychol Sci 2022 Jan;17(1):226-238. [doi: 10.1177/1745691620974774] [Medline: 33651952]
87. Valentine L, McEnery C, Bell I, O'Sullivan S, Pryor I, Gleeson J, et al. Blended digital and face-to-face care for first-episode psychosis treatment in young people: qualitative study. JMIR Ment Health 2020 Jul 28;7(7):e18990 [FREE Full text] [doi: 10.2196/18990] [Medline: 32720904]
88. Sanderson C, Kouzoupi N, Hall CL. Technology Matters: the human touch in a digital age - a blended approach in mental healthcare delivery with children and young people. Child Adolesc Ment Health 2020 May;25(2):120-122. [doi: 10.1111/camh.12385] [Medline: 32307846]
89. Ramos G, Ponting C, Labao JP, Sobowale K. Considerations of diversity, equity, and inclusion in mental health apps: a scoping review of evaluation frameworks. Behav Res Ther 2021 Dec;147:103990 [FREE Full text] [doi: 10.1016/j.brat.2021.103990] [Medline: 34715396]
90. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. Curr Treat Options Psych 2019 Jul 5;6(3):243-255. [doi: 10.1007/s40501-019-00181-z]

Abbreviations

PRIME: Personalized Real-time Intervention for Motivational Enhancement
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

XSL•FO
RenderX

YPAG: Young People's Advisory Groups

Edited by R Kukafka; submitted 16.06.22; peer-reviewed by JW Hur, J Long, J Marshall; comments to author 30.07.22; revised version received 12.08.22; accepted 28.09.22; published 09.11.22.

Please cite as:
Bear HA, Ayala Nunes L, DeJesus J, Liverpool S, Moltrecht B, Neelakantan L, Harriss E, Watkins E, Fazel M
Determination of Markers of Successful Implementation of Mental Health Apps for Young People: Systematic Review
J Med Internet Res 2022;24(11):e40347
URL: https://www.jmir.org/2022/11/e40347
doi:10.2196/40347
PMID:36350704

©Holly Alice Bear, Lara Ayala Nunes, John DeJesus, Shaun Liverpool, Bettina Moltrecht, Lakshmi Neelakantan, Elinor Harriss, Edward Watkins, Mina Fazel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Understanding the Technological Landscape of Home Health Aides: Scoping Literature Review and a Landscape Analysis of Existing mHealth Apps

Elizabeth Fong-Chy Kuo[1*], BSc; Jacklyn Cho[2*], BSc; Iredia Olaye[2], MSc, MHA, PhD; Diana Delgado[2], MLS; Nicola Dell[1], PhD; Madeline R Sterling[2], MD, MPH, MSc

[1]Department of Information Science, Jacobs Technion-Cornell Institute, Cornell Tech, New York, NY, United States

[2]Division of General Internal Medicine, Department of Medicine, Weill Cornell Medicine, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Madeline R Sterling, MD, MPH, MSc
Division of General Internal Medicine
Department of Medicine
Weill Cornell Medicine
420 East 70th Street, Room LH-357
New York, NY, 10021
United States
Phone: 1 646 962 5029
Fax: 1 212 746 8165
Email: mrs9012@med.cornell.edu

Abstract

Background: Home health aides (HHAs) provide necessary hands-on care to older adults and those with chronic conditions in their homes. Despite their integral role, HHAs experience numerous challenges in their work, including their ability to communicate with other health care professionals about patient care while caring for patients and access to educational resources. Although technological interventions have the potential to address these challenges, little is known about the technological landscape and existing technology-based interventions designed for and used by this workforce.

Objective: We conducted a scoping review of the scientific literature to identify existing studies that have described, designed, deployed, or tested technology-based tools and apps intended for use by HHAs to care for patients at home. To complement our literature review, we conducted a landscape analysis of existing mobile apps intended for HHAs providing in-home care.

Methods: We searched the following databases from their inception to October 2020: Ovid MEDLINE, Ovid Embase, Cochrane Library, and CINAHL (EBSCO). A total of 3 researchers screened the yield using prespecified inclusion and exclusion criteria. In addition, 4 researchers independently reviewed these articles, and a fifth researcher arbitrated when needed. Among studies that met the inclusion criteria, data were extracted and summarized narratively. An analysis of mobile health apps designed for HHAs was performed using a predefined set of terms to search Google Play and Apple App stores. Overall, 2 researchers independently screened the resulting apps, and those that met the inclusion criteria were categorized according to their intended purpose and functionality.

Results: Of the 8643 studies retrieved, 182 (2.11%) underwent full-text review, and 4.9% (9/182) met our inclusion criteria. Approximately half (4/9, 44%) of the studies were descriptive in nature, proposing technology-based systems (eg, web portals and dashboards) or prototypes without a technical or user-based evaluation of the technology. In most (7/9, 78%) papers, HHAs were just one of several users and not the sole or primary intended users of the technology. Our review of mobile apps yielded 166 Android and iOS apps, of which 48 (29%) met the inclusion criteria. These apps provided HHAs with one or more of the following functions: electronic visit verification (29/48, 60%), clocking in and out (23/48, 48%), documentation (22/48, 46%), task checklist (19/48, 40%), communication between HHA and agency (14/48, 29%), patient information (6/48, 13%), resources (5/48, 10%), and communication between HHA and patients (4/48, 8%). Of the 48 apps, 25 (52%) performed monitoring functions, 4 (8%) performed supporting functions, and 19 (40%) performed both.

XSL•FO
RenderX

Conclusions: A limited number of studies and mobile apps have been designed to support HHAs in their work. Further research and rigorous evaluation of technology-based tools are needed to assess their impact on the work HHAs provide in patient's homes.

(J Med Internet Res 2022;24(11):e39997) doi:10.2196/39997

KEYWORDS

home health aides; home care services; mobile health; mHealth; mobile apps; mobile phone apps; smartphones; educational technology; technology; mobile phone

Introduction

Background: Home Health Aides

By 2060, the number of Americans aged >65 years is projected to reach approximately 95 million, making up almost one-fourth of the population in the United States. Most older adults, including those with multiple chronic conditions, prefer to stay in their homes and communities for as long as they can and avoid nursing homes, a concept referred to as "aging in place." To do so, they require help at home from family caregivers and home health aides (HHAs). HHAs represent the sixth fastest growing occupation in the United States; at present, there are 2.3 million HHAs in the United States and are expected to grow by 1.5 million by 2030 [1]. Largely employed by home care agencies, HHAs are trained and certified health professionals who provide assistance with personal care (such as activities of daily living [eg, bathing and dressing] and instrumental activities of daily living [eg, preparing meals, cleaning, and shopping], emotional support, and medically oriented care [eg, taking vital signs and medication reminders]) to older adults and those with chronic conditions at home [2]. Unlike other health professionals such as physicians, nurses, and physical therapists, HHAs interact with patients on a daily or near-daily basis, which gives them a unique vantage point from which to observe, support, and advise patients [3]. Therefore, they are often referred to as the "eyes and ears" for patients and the medical team in the home.

Professional Challenges on the Job

However, despite being integral to patient care, HHAs are an overlooked and underutilized group of health care professionals who experience challenges in caring for patients at home. Most women and minorities of color who earn low wages work long hours, have erratic schedules, and have limited opportunities for career advancement [4]. Studies have found that HHAs increasingly care for medically complex patients with a high burden of chronic diseases [5,6]. Despite their observations and insights into patients' health, they are rarely considered a part of the patient's medical team. In addition, when they try to report information, convey their concerns, have questions, or need advice, they have difficulty reaching their supervisors via phone. Furthermore, formal technology systems are lacking in conveying the data they collect to other health care professionals. eHealth, which refers to health services that are delivered through the internet or other technologies [7], mobile health (mHealth), and telehealth, which are both subsets of eHealth, have the potential to fill this gap and benefit HHAs. Technology can connect multiple aspects of the health care system (physicians, patients, staff, and HHAs) and increase the efficiency of health care delivery. Although eHealth interventions can be challenging to implement because of their complexity, the integration of technology may ultimately make HHA jobs easier [8].

Finally, although HHAs receive training for certification and maintenance, many of their courses are general and not disease specific, which may not meet the clinical needs of the older adults [9]. Qualitative studies have shown that the workforce wants and benefits from technological systems to address these issues. For example, when HHAs monitor their patients' blood pressure, a technological system that allows for the measurement, recording, and transmission of these data to other providers would be useful, especially when the values are abnormal. Similarly, when caring for patients and questions arise about their conditions, many HHAs rely on past clinical experiences or "Google" their questions using their personal devices. Instead, HHAs might prefer to have access to disease-specific (eg, stroke) information on their personal devices that they can reference while at work.

Objective

To meet these needs and inform future technological innovations for this workforce, a better understanding of the technology landscape is needed [10]. Herein, we conducted the first scoping review of the scientific literature to identify existing studies that have described, designed, deployed, or tested technology-based tools and apps that are intended for use by HHAs as they care for older adults or those with chronic conditions. To complement our literature review and provide additional context, we also conducted a landscape analysis of existing mobile apps pertaining to HHAs and the care they provide in the home environment.

Methods: Scoping Review

This scoping review is reported in line with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidance [11].

Guiding Framework

We conducted a scoping review using the 5-stage framework developed by Arksey and O'Malley [12]. The 5 stages include (1) identification of the research question, (2) identification of relevant studies, (3) selection of relevant studies for the review, (4) charting information and data from the selected literature, and (5) summarizing and reporting the results of the review.

Search Strategy

A medical librarian (DD) performed a comprehensive literature search on October 28, 2020, of Ovid MEDLINE(R) ALL, from

XSL•FO
RenderX

1946 to October 27, 2020, Ovid Embase (from 1974 to October 27, 2020), Cochrane Library (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, and Cochrane Methodology Register), and CINAHL (EBSCO) from inception to October 2020. The first search was conducted using Ovid MEDLINE. Subject headings and keywords were adapted for other databases. No restrictions were applied on language, publication date, or article type. Additional records were identified by reviewing reference lists and using the "Cited by" and "View references" features in Scopus of the included studies. The full set of search terms for Ovid MEDLINE is presented in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

This review was limited to studies that focused on technology-based tools, innovations, or interventions intended to be used by home health care workers (including HHAs, attendants, and personal care aides). Studies can be descriptive in nature (eg, overview of technology design), quasi-experimental, or randomized controlled trials. Only peer-reviewed studies published in the English language were included. Qualitative studies that did not discuss or propose an intervention, reviews, editorials, or scientific meeting abstracts were excluded. Studies that focused on other people who provide care at home (eg, nurses or family caregivers) were excluded. Studies that were conducted in nursing homes, long-term care centers, and acute rehabilitation centers were also excluded.

Selection of Studies

All studies identified following the database search were uploaded to the web-based systematic review software package Covidence (Veritas Health Innovation). First, the title and abstract reviews of all studies were completed independently by 3 authors (JC, IO, and ND). Disagreements were discussed and resolved through consensus. A record was kept of all the studies excluded and the reason for exclusion in Covidence. All studies that met the inclusion criteria (189 studies) went through a full-text screening process by the 4 authors independently (JC, IO, EFK, and ND), and any disagreements on the eligibility of the studies were reviewed by a fifth author (MRS).

Data Extraction

Data from the included studies were extracted using the following categories: (1) author, (2) country, (3) year of publication, (4) title of the study, (5) journal, (6) contribution, (7) technology innovation, (8) intended users, (9) study objective and systems goals, and (10) evaluation and assessment of innovation.

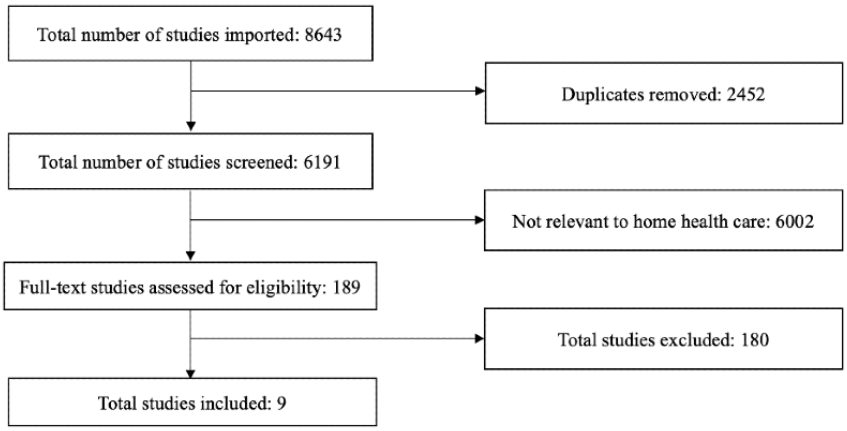
Results

Study Characteristics

In total, 8643 studies were imported from our search of the peer-reviewed literature. Among these 8643 studies, 2452 (28.36%) were excluded because they were duplicates. We screened 6191 abstracts and excluded 6002 (96.94%) studies because they were not relevant to home health care. A total of 189 full-text studies were assessed for eligibility, and 7 studies were included. A medical librarian (DD) identified 13 additional studies from the citation chasing process; among these, 2 studies met the inclusion criteria. Taken together, we identified 9 full-text studies that met the inclusion criteria (Figure 1).

The characteristics of the 9 included studies are presented in Table 1 (Multimedia Appendix 2 [13-21]). The studies were published between 2004 and 2018 in journals that focused on technology, computer science, home and long-term care, and gerontology or aging. Most studies were conducted in Europe (5/9, 56%), whereas one-third (3/9, 33%) were conducted in the United States and 11% (1/9) in Japan. More than half (6/9, 67%) of the studies were descriptive in nature, proposing technology-based systems (eg, web portal or dashboard) or prototypes without a technical or user-based evaluation. Of the 3 studies that included evaluations, all but one (1/3, 33%; Danilovich et al [21]) were design prototypes or feasibility pilot studies with limited data collection on the users (eg, nurses, family caregivers, or HHAs) of the technology or on patients. Of the 9 included studies, most (n=7, 78%) evaluated HHAs as just one of several caregivers or health professionals as intended end users, rather than as primary users. Most (6/9, 67%) of the technological interventions were web based, 22% (2/9) of studies described mobile apps, and 11% (1/9) of studies tested a DVD. We discuss these studies in detail in the following sections.

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



XSL•FO
RenderX

Table 1. Results from the scoping review of literature.

Author, year, and country[a]	Study title	Technology innovation	Intended users	HHA[b] role
Ogawa et al [13], 2004; Japan	A Java mobile phone-based "home helper" care report creation support system	Java mobile phone–based care report creation support system	Home helpers	Primary
Scandurra et al [14], 2006; Sweden	Visualisation and interaction design solutions to address specific demands in shared home care	Design prototype based on participatory design	GP[c], DN[d], and HHS[e] personnel	Peripheral; one of many
Paganelli et al [15], 2011; Italy	An ontology-based system for context-aware and configurable services to support home-based continuous care	Emilia Romagna Mobile Health Assistance Network (ERMHAN) service platform	Patients, family members, home care teams (clinicians, GPs, nurses, etc), and social community members (eg, social workers and volunteers)	Peripheral; one of many
Page et al [16], 2012; the United States	Improving care delivery using health information technology in the home care setting: development of the home continuation care dashboard	Web dashboard intended to bridge the gap between physicians and home care case managers	Physicians, home care case managers, patients, and caregivers	Peripheral; one of many
De Backere et al [17], 2016; Belgium	The OCareCloudS project: toward organizing care through trusted cloud services	OCarePlatform and cloud-based semantic system to offer information and knowledge-based services for older people and their informal and formal caregivers	Older patients residing at home and informal and formal caregivers	Secondary; 1 of 3 (patient is primary)
De Backere et al [18], 2017; Belgium	The OCarePlatform: a context-aware system to support independent living	Sensor-based in-home system	Multiple formal and informal caregivers involved in a patient's care	Peripheral; one of many
Danilovich et al [19], 2017; the United States	Design and development of a mobile exercise application for home care aides and older adult Medicaid home and community-based clients	Mobile exercise app. Content of the program itself seems static. Minimal data entry about patient (pain and mood)	Home HCA[f] and patients	Secondary; 1 of 2 (patient is primary)
Bourikas et al [20], 2017; the United Kingdom	Elderly support to inspired ageing (ESTIA)	Elderly Support to Inspired Ageing platform that enables medical and background information to be combined into a single server	Family, volunteers, older people, home care aides, hospitals	Primary
Danilovich et al [21], 2017; the United States	Translating Strong for Life into the Community Care Program: Lessons Learned	SFL[g]: Resistance Exercise Intervention: 35-minute DVD on warm-up and upper and lower extremity exercises for homebound older adult clients	Home HCA and patients	Secondary; 1 of 2 (patient is primary)

[a]Studies are listed in chronological order based on the year published.

[b]HHA: home health aide.

[c]GP: general practitioner.

[d]DN: district nurse.

[e]HHS: home help service.

[f]HCA: health care aides.

[g]SFL: strong for life.

Intended Users of Technology

HHAs were the intended users of the technology for 33% (3/9) proposed technology interventions. Of these 3 studies, only 1 (33%) study by Ogawa et al [13] exclusively focused on supporting HHAs in their work, whereas the remaining 2 (67%) studies included HHAs as peripheral users of the technology with the larger goal of supporting the patients. In a study of 21 home helpers (eg, HHAs) in Japan, Ogawa et al [13] described a Java mobile phone–based system intended for HHAs to use when caring for older adults at home. The system aimed to reduce the amount of time and technical challenges for HHAs reporting to their agencies and patients. In contrast, 2 studies by Danilovich et al [19,21] designed technology for HHAs to use; HHAs were one of several stakeholders involved in intervention development and deployment, and the primary focus of the technology was ultimately the patient. For example, in one of their 2017 studies, Danilovich et al [19] designed, developed, and piloted a mobile exercise app (app and videos) for frail older patients, and HHAs and physical therapists were trained on how to help these patients use the app. In a separate study, Danilovich et al [21] trained home care aides in an evidence-based resistance exercise program that consisted of workout DVDs. However, as in the previous study, the intended

XSL•FO
RenderX

users of the exercise DVD program were homebound and community-dwelling patients.

The remaining 67% (6/9) of studies examined technology tools that were not designed primarily for HHAs. HHAs were one of several types of users. Although these tools supported HHAs, they targeted patients and general members of the health care team, such as physicians, nurses, and physical therapists, as the main users. For example, Page et al [16] described the development of a web-portal home continuation care dashboard to facilitate communication among case managers, physicians, patients, and their caregivers at home.

Purpose of Technology

All (9/9, 100%) studies aimed to support HHAs in caring for patients at home. There was a wide range of distinct purposes of the technology proposed and tested. Overall, 78% (7/9) of studies proposed and described digital software platforms, ranging from web-based platforms to enhance communication in the home environment to a sensor-based in-home tool to alert caregivers of their patients' falls and physical injuries. Only the study by Ogawa et al [13] focused on designing and pilot-testing a Java mobile phone–based system to facilitate data entry and documentation exclusively for HHAs. The remaining platforms were designed to facilitate communication and coordination among patients and members of their health care team. These members included physicians, nurses, case managers, HHAs, and the patient's family members.

A total of 22% (2/9) of studies designed and tested a technological intervention intended to improve patients' physical mobility. Both studies by Danilovich et al [19,21] used technology (mobile apps and DVDs) to train HHAs on exercise regimens, which could benefit patients (mobility) and HHAs (job satisfaction). Specifically, Danilovich et al [19] developed and pilot-tested a mobile exercise app with HHA-patient dyads. The mobile app presented users (eg, older patients and their HHAs) with several exercise videos filmed by the researchers and HHAs. These videos are downloaded onto the app so that users can access the app without the internet. On completing the exercises, users are prompted to update their progress on the app. Although the researchers included HHAs as a key group of stakeholders who could assist patients with the app, the app was intended for patients and did not provide direct assistance to support HHAs' work.

Study Design and User Evaluation

Of the 9 studies, 6 (67%) presented descriptions of the proposed technology and system. None of these studies included evaluations of the technology or data on feedback from the intended users, HHAs. A study discussed the development and testing of a system prototype among home help service personnel, nurses, and general practitioners. However, no follow-up user evaluations or deployment data were evaluated.

Only 22% (2/9) of studies collected and reported quantitative and qualitative data on user evaluation and deployment efforts from the perspective of users, one of whom was HHAs. These evaluation efforts focused on overall program satisfaction. For example, a study by Danilovich et al [21] conducted a mixed methods randomized controlled trial and specifically focused on examining the effect of a technology-based exercise program on home care aides' perceptions of job satisfaction, achievement, and recognition. In this study, the researchers assigned 17 and 15 HHAs to the intervention and control groups, respectively. The intervention group received resistance bands and exercise programs in a DVD-based format for older adults to complete on days when the HHAs did not visit. The researchers compared preintervention and postintervention scores to assess the effectiveness of the resistance exercise intervention. Quantitative and qualitative assessments were conducted and evaluated using validated instruments (ie, Job in General and Work on Present Job scales) and blinded field observations by the researchers.

In another study by Danilovich et al [19], the authors developed and pilot-tested a novel mobile exercise app to engage older adults in physical activity. A total of 5 HHA-patient dyads were recruited to participate in the study. The participants provided quantitative and qualitative feedback via written questionnaires and semistructured interviews. The outcomes assessed were usability evaluations focusing on 3 domains: system usefulness, information quality, and interface quality. The participants provided further feedback on the functionality and aesthetics of the mobile app.

Methods: Landscape Analysis

On the basis of paucity of results from the scoping review, we conducted a landscape analysis of existing mobile apps that were designed for HHAs and could potentially assist them in their work caring for patients in the home.

Search Strategy

Two authors (EFK and JC) searched for existing mHealth apps created for HHAs on the Google Play store (for Android apps) and the Apple App store (for iOS apps) using a predefined set of terms (Multimedia Appendix 3).

Screening and Data Extraction

Our inclusion criteria for mobile apps included apps that (1) were available on the iOS Apple or Google Play stores, (2) were primarily designed for HHAs, and (3) supported HHAs with their work in patients' homes. For example, we included apps with features for documentation, communication, and training. These are resources that HHAs may use while working directly with patients.

Our initial search yielded 686 Android apps and 289 Apple apps that were screened for inclusion. We created a custom-built Python script to automatically save all the search results as a list, which facilitated further analysis. In our first pass, we removed apps that were clearly not relevant to HHAs (eg, patient self-tracking apps, radio stations, or self-help books) This yielded 175 apps: 148 Android and 27 Apple apps. For each app, we collected the app name, ID, year released, last year updated, number of downloads, app description, and number of reviews. We then removed an additional 8 apps that were duplicated in the data set (ie, had both Android and iOS versions). A total of 167 apps underwent independent review by 2 authors (EFK and JC), who examined each app's descriptions (found on the Google Play or the iOS Apple App stores) for their intended users (eg, HHA, medical professionals

XSL•FO
RenderX

including HHA, or nurses) and purposes (eg, connecting providers with patients, providing task checklists, and GPS tracking). Apps that did not make HHAs one of the primary users were excluded.

After doing so, a total of 67 mobile apps met our inclusion criteria for further review.

We verified the completeness of the resulting set of apps in 2 ways. (1) For each relevant app, we used search engine optimization software (Semrush) that given the name and URL of a relevant app, provided a list of "competitor" apps that would be likely serve the same purpose or provide similar functionality. We reviewed the suggested competitor apps for all HHA-relevant apps to confirm whether they were already present in the set of apps or assess whether they met the inclusion criteria. (2) In addition, we used the built-in "recommended app" features provided by both Google Play and Apple App stores. We entered the name of each HHA-relevant app and noted any alternative or similar apps that were recommended by each platform. We then checked these recommended apps to see if they were already in our data set or assessed whether they should be included. Neither of these processes yielded new apps that were not already present in our data set, which increased the confidence that our search process discovered all relevant apps.

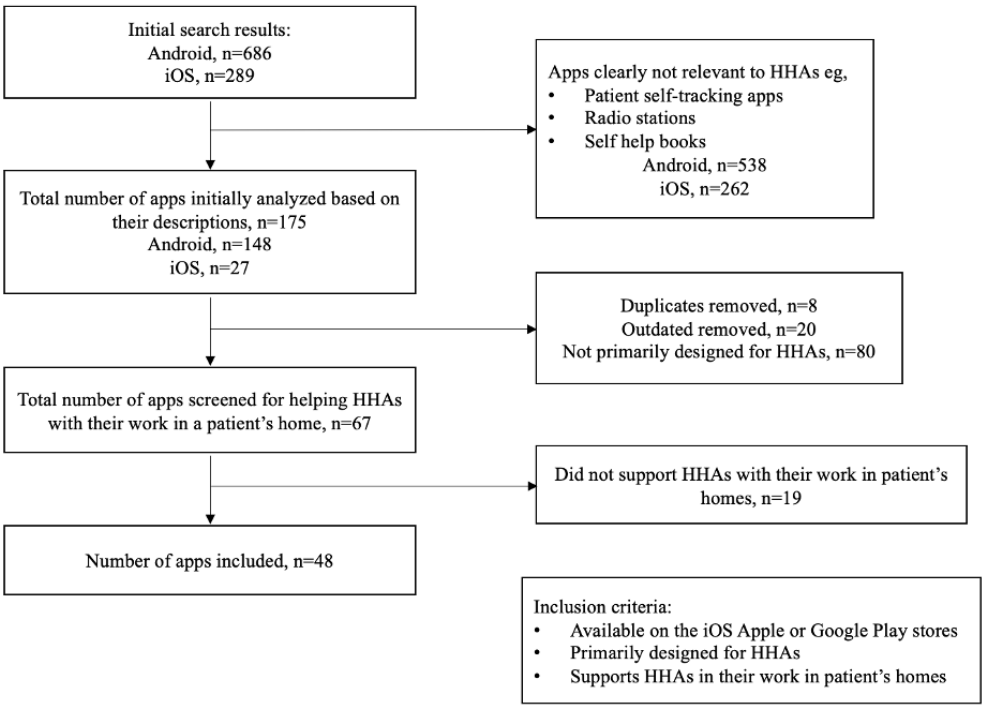
Results

Characteristics of the Included Mobile Apps

Since we sought to study apps that assisted HHAs with their work in a patient's home, apps that only served HHAs before or after their patient visit were eliminated. Through our previous categorizations based on each mobile app's description, we included apps that performed at least one of the following functions: (1) allowed HHAs to access their task checklist; (2) allowed HHAs to document their work of the day; (3) provided a place for HHAs to access their patient's information; (4) facilitated communication between an HHA and their agency; (5) facilitated communication between an HHA and their patient; (6) provided resources such as training courses, information, and so on for HHAs; (7) helped with electronic visit verification; and (8) assisted HHAs with clocking in and out. After applying these criteria to the 67 apps, 48 (72%) remained that performed one or more of the 8 core functions (Figure 2).

An overview of the characteristics of these 48 unique apps is presented in Table 2 (Multimedia Appendix 4). The majority of apps studied assisted with electronic visit verification (29/48, 60%); fewer apps provided means for communication between HHAs and patients (4/48, 8%) and resources for HHAs (5/48, 10%). Other notable functions that apps had were clocking in and out of shifts (23/48, 48%), documenting work performed by the HHA (22/48, 46%), allowing access to a task checklist for the day (19/48, 40%), facilitating communication between the HHA and the agency (14/48, 29%), and providing patient information (6/48, 13%). Each app may have more than one characteristic. We further categorized the 48 apps according to their primary purpose of monitoring HHAs, supporting HHAs, or both. We defined a monitoring app as one that helps an agency or employer keep track of HHAs' tasks or their location in the patient's home. We defined a supporting app as one that provides various resources or training tools for the HHA, gives the HHA information on their patient, or assists with communication between the HHA and their agency or patient.

Figure 2. Flow diagram for the landscape analysis of mobile health apps. HHA: home health aide.



XSL•FO
RenderX

Table 2. Results from the landscape analysis of mobile apps.

Name of the mobile app[a]	Year	Type	Primary users	Objective
Domiciliary Care Toolkit	2014	Android	Home care providers (including HCWs[b])	Supporting
HHAeXchange	2014	Android	HCW	Both
Verify Centre Home Health	2015	Android	HCW	Both
Alora Plus	2016	Android	HCW	Both
Connected Home Care	2016	Android	HCW	Both
Electronic Visit Verification	2016	Android	HCW	Monitoring
FreedomCare Plus	2016	Android	HCW	Monitoring
MedFlyt	2016	Android	Home care providers (including HCWs)	Both
PointClickCare Care at Home	2016	Android	HCW	Monitoring
CareConnect	2017	Android	HCW	Supporting
Axxess HomeCare	2017	Android	HCW	Both
Caretap EVV	2017	Android	Home care providers (including HCWs)	Monitoring
DCI Mobile EVV	2017	Android	HCW	Both
eRSP Mobile Connect	2017	Android	Home care providers (including HCWs)	Both
FormDox EVV for Aides	2017	Android	HCW	Both
Ally Home Care	2018	Android	HCW	Monitoring
August Systems Mobile for Caregivers	2018	Android	HCW	Both
AuthentiCare 2.0	2018	Android	HCW	Monitoring
ClearCareGo Caregiver	2018	Android	HCW	Monitoring
CliniqOS	2018	Android	Home care providers (including HCWs)	Both
CrescendoConnect	2018	Android	Home care providers (not specific to HCWs)	Both
Domiciliary Care Worker Gweithiwr Gofal Cartref	2018	Android	HCW	Supporting
Helpers Home Care	2018	Android	Home care providers (including HCWs)	Both
My EVV	2018	Android	HCW	Monitoring
MyEzcare—EVV	2018	Android	HCW	Monitoring
Mobile Caregiver+	2018	Android	HCW	Monitoring
Honor Care Pro	2018	Android	HCW	Both
UCP Caregiver Staffing	2018	iOS	HCW	Monitoring
BarbaraKares	2019	Android	HCW	Monitoring
CareTime	2019	Android	HCW	Monitoring
Cashe EVV	2019	Android	HCW	Both
KorEvv	2019	Android	HCW	Monitoring
MatrixCare for Home Care	2019	Android	HCW	Both
myHRresults—At Work	2019	Android	HCW	Monitoring
SwyftOps—Caregiver App	2019	Android	HCW	Monitoring
Vertex EVV	2019	Android	HCW	Monitoring
HomecareGPS Mobile	2019	iOS	HCW	Monitoring
ServTracker Mobile Home Care	2019	iOS	HCW	Monitoring
Moravia Shifts	2020	Android	HCW	Monitoring
Netsmart Homecare Mobile Phone	2020	Android	HCW	Both
BAYADA Home	2021	Android	HCW	Monitoring

XSL•FO
RenderX

Name of the mobile app[a]	Year	Type	Primary users	Objective
Careswitch	2021	Android	HCW	Both
Visit Wizard Mobile	__[c]	Android	HCW	Both
Best Care	—	Android	HCW	Monitoring
Caregiver App	—	Android	HCW	Monitoring
Caregiver Cloud Training	—	Android	Home care providers (including HCWs)	Supporting
Time4Care	—	Android	HCW	Monitoring
ViolaCare	—	Android	HCW	Monitoring

[a]Apps are listed in chronological order based on the year created or last updated.

[b]HCW: home care worker.

[c]Missing information.

Of the 48 apps, 25 (52%) apps focused on monitoring functions, 4 (8%) apps provided supporting functions, and 19 (40%) apps provided both. Of the 48 apps, 34 (71%) were developed and sponsored by software companies, 9 (19%) were developed and sponsored by individual home care agencies, 1 (2%) by government-partnered software companies, and 1 (2%) by government-partnered agencies. Specific developer information could not be found for 6% (3/48) of the apps. The apps that were developed between the years 2014 and 2021 were from 5 countries; most apps were from the United States (41/48, 85%), with the remainder from the United Kingdom, Ireland, Canada, and Ethiopia. The number of downloads ranged from 5 to >100,000, and user ratings ranged from 2 out of 5 stars to 5 out of 5 stars.

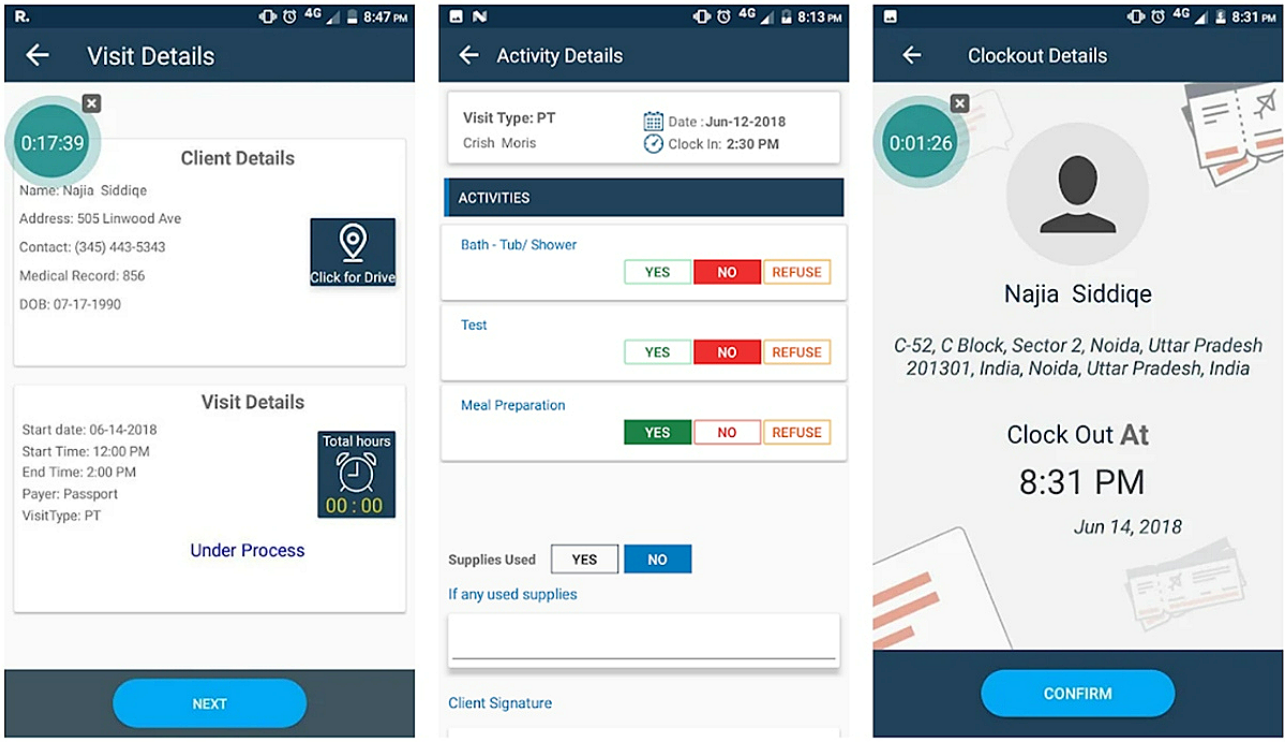
Apps That Only Monitor HHAs in the Home

The most common feature provided by the apps was monitoring of HHAs at home (25/48, 52%). This included monitoring whether HHAs arrived at the patient's home on time by logging HHAs' work hours (16/25, 64%), keeping track of HHAs' tasks (11/25, 44%), reporting their real-time GPS location in the patient's home via electronic visit verification (17/25, 68%), and documenting information about the patient (11/25, 44%).

For example, FormDox EVV (Figure 3), which was developed by FormDox Technology Solution, includes features such as electronic visit verification and documentation services. The app focuses on assisting agencies by tracking their HHAs and collecting data on HHA performance. It was rated 4.4 out of 5 stars by the users (Multimedia Appendix 4).

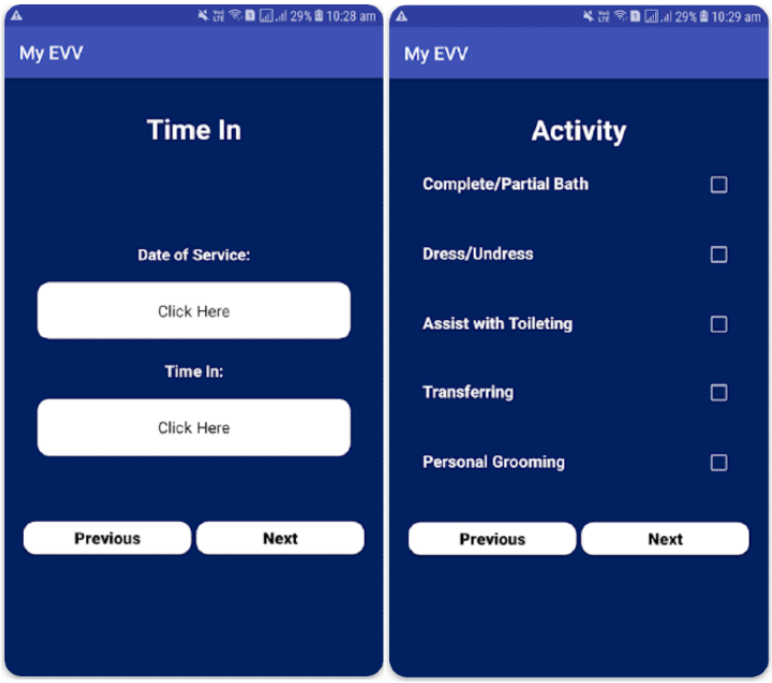
Another example is My EVV (Figure 4), which is an app that allows HHAs to clock in and out of shifts electronically and allows HHAs to document services provided to the patient. The app received 3.9 stars on the Google Play store and has been downloaded >10,000 times (Multimedia Appendix 4).

Figure 3. FormDox EVV interface.



XSL•FO
RenderX

Figure 4. My EVV interface.

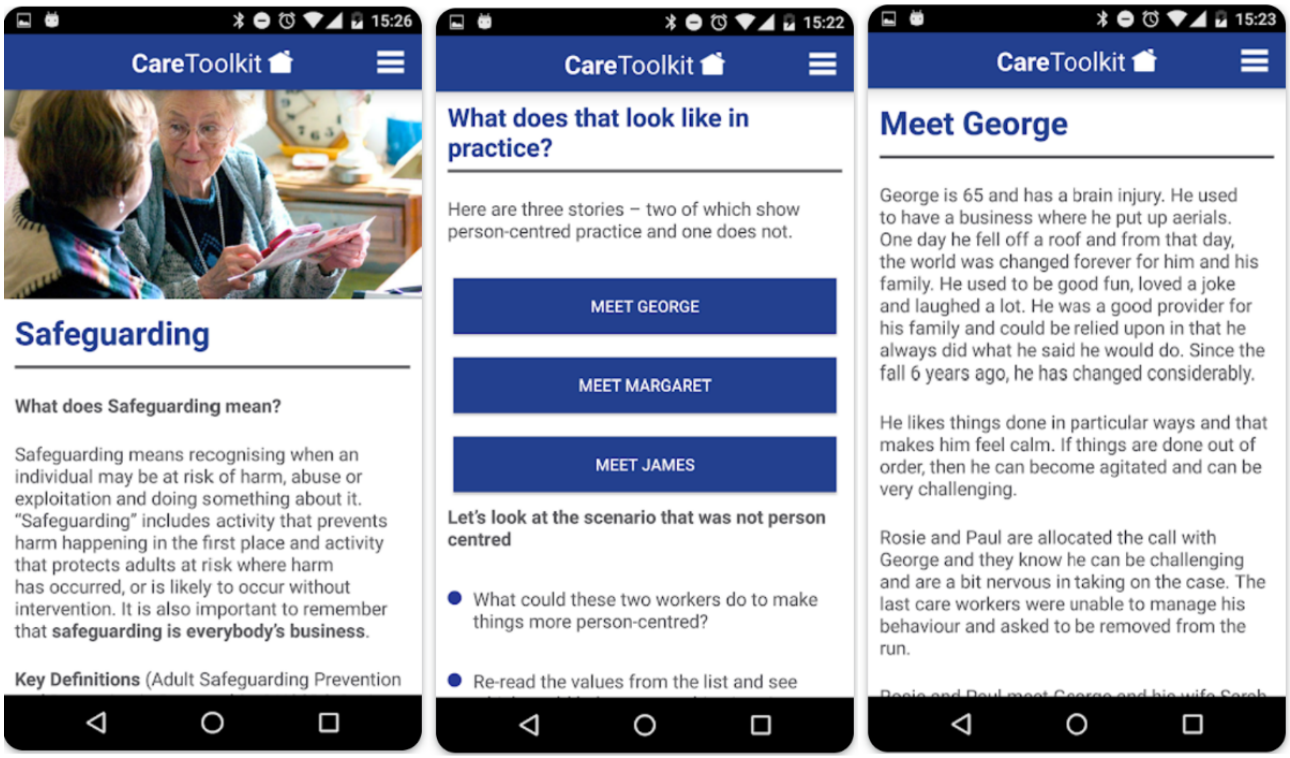


Apps That Only Support HHAs at Home as They Provide Care

Of the 48 apps, 4 (8%) focused solely on providing support for HHAs. Of the 4 apps, 3 (75%) them provided information or training resources for HHAs and 1 (25%) app assisted with communication between HHAs and their agencies. However, none of the apps included other supporting functions, such as assisting with communication between HHAs and their patients or providing information about the patient to the HHA.

An example of a supporting app is the Domiciliary Care Toolkit (Figure 5). This app was developed by the Northern Ireland Social Care Council in partnership with care providers. It provides guidelines for HHAs on how to conduct clinical tasks as well as resources on how to handle patients with dementia and delirium. The app is unique in that there are no functions that track HHAs, and its only purpose is to serve as a reference tool. It describes the issues an HHA may face and offers possible solutions through patient examples.

Figure 5. Domiciliary Toolkit interface.



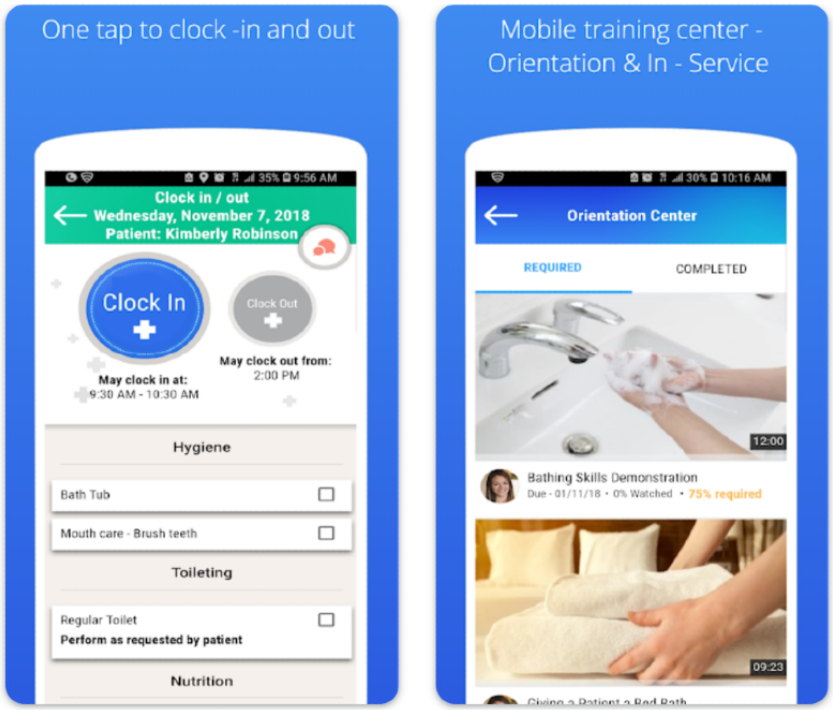
XSL•FO
RenderX

Apps That Both Monitor and Support HHAs

Of the 48 apps, 19 (40%) were apps that both supported and monitored HHAs. The 19 apps included features for keeping track of tasks (n=8, 42%), for HHAs to use for documentation (n=11, 58%), for electronic visit verification (n=12, 63%), for assisting with clocking in and out (n=7, 37%), for HHAs to find information about the patient (n=6, 32%), for facilitating communication between the HHA and patient (n=4, 21%), for facilitating communication between the HHA and agency (n=12, 63%), and for resources (n=1, 5%).

For example, the app MedFlyt (Figure 6) allows agencies to not only monitor HHAs through documentation features and reminders to clock in and out but also support HHAs with web-based training courses and communication (eg, instant messaging between HHA and the agency). With 2466 reviews and 4.7 stars, the app is one of the most highly rated apps of the 48 we reviewed (Table 2).

Figure 6. MedFlyt interface.



Discussion

Main Findings

Our findings illustrate the need for increased research on technological interventions for HHAs and the further development of mobile technologies that support HHAs with their work in patients' homes. The scoping review of the peer-reviewed literature yielded only 9 studies, and in most of them, HHAs were not the primary intended users of the technology. In addition, very few studies have assessed the feasibility and effectiveness of this technology among HHAs. The landscape analysis revealed only 4 existing apps that were solely focused on supporting HHAs, with most apps designed for agencies to monitor HHAs rather than assisting them with their work. The lack of studies from the scoping review that included data on user feedback and the dearth of research on how mobile technologies impact HHAs' work suggests an urgent need for research that rigorously evaluates technology-based tools to measure their effect on HHAs' caregiving in patients' homes.

Research in the field of human-computer interaction has long acknowledged the importance of actively engaging the eventual users of technologies in their design [22], such as via participatory or ethnographic methods that enable designers to deeply understand the context and problems being addressed [23-25]. Without taking the time to learn people's current practices and understand their perspectives, priorities, and values, designers run the risk of building technologies that are not appropriate or usable and fail to meet people's true needs. The stakes are particularly high in in-home health care, where poorly designed technologies might have a negative impact not only on HHAs but also on patient care.

Although prior work [26] suggested that HHAs are eager to play an active role in the design of technologies, none of the 9 papers in our scoping review discussed engaging with HHAs to deeply understand their needs and workflows before building an intervention. Moreover, none of the studies presented a long-term evaluation of the impact of the intervention on the HHAs' workflow or patient care. In addition, there is little evidence to suggest that any of the final 48 apps included in our landscape analysis have been rigorously evaluated for feasibility, clinical trials, or user deployments (eg, none of the apps in our search appeared in any papers in our scoping review). This suggests that to date, the perspectives of HHAs in both the design and evaluation of technology have been lacking.

Rather, our landscape analysis suggests that many of the existing apps have been primarily developed with home care agencies in mind, providing functionality that primarily serves the

XSL•FO
RenderX

administrative needs of the agency (eg, electronic visit verification) rather than providing on-the-job support for HHAs. Compounding these concerns, the COVID-19 pandemic has accelerated and amplified the use of technology in HHAs' work. For example, home care agencies have been rapidly transitioning to using digital tools for remote training, scheduling, and monitoring HHAs' work, but it remains unclear whether, or to what extent, these technology tools meet HHAs' needs or impact patient care [27]. However, some of the apps in our study provide functions that support HHAs, including the ability to obtain training, communicate with their agencies or patients, and search for patient information. Additional efforts are needed to understand how HHAs perceive these functions, including how they are used in real time and what their impact is on HHA employment attitudes (eg, job satisfaction) and self-efficacy in providing care.

Notably, several recent reviews of mHealth apps relate to and build on some of our main findings. A scoping review by Vaughan et al [28] analyzed existing studies of mHealth apps that support nurses in monitoring their patients' chronic wounds. They found that although several wound care apps are available for nurses to use, there is a lack of rigorous and standardized evaluations of these apps, few clinical trials, and a paucity of information about which apps nurses actually use in real time [28]. A scoping review by Dauletbaev et al [29] highlighted the increased the use of mHealth telemonitoring for patient care since the COVID-19 pandemic, a trend we also saw in-home health. Although both reviews highlight the importance of technology for the health care workforce and patient care, neither review included HHAs. Finally, a recent systematic review by Widdison et al [30] examined the effectiveness of randomized control trials that used mHealth apps to deliver pelvic floor muscle training exercises to patients with urinary incontinence. The rigor of the included studies signals a gap with the studies in our review and where the current research on HHAs needs to be conducted in the future.

Limitations

Our study has a few limitations. First, new studies and mobile apps may have been published or released after we collected data for our scoping review and landscape analysis, which may signal an underrepresentation of existing studies and apps. Second, mobile apps available at the time of data analysis may have been discontinued. Finally, our study examined mobile app descriptions on the Google Play or the Apple stores between 2019 and 2021, and our categorization of these apps depends on the accuracy of these descriptions. Future studies should verify publicly facing descriptions of these apps to confirm that they accurately represent the intended use of the products.

Conclusions

Our findings suggest that despite the integral role of HHAs in patient care and their exposure to and use of technology, few studies of technology-based interventions designed for this workforce exist and those that do lack rigorous evaluations. In addition, although many apps for the workforce are in use, most are designed from the perspective of the home care agency, not the HHA, and serve to monitor HHAs rather than support them in providing care to patients. Taken together, there is an urgent need for research that centers on the needs and perspectives of HHAs and using human-centered methods to engage HHAs in the design of technologies that truly support their essential caregiving work. Such approaches will also likely make HHAs feel more included and valued in the health care system, addressing the challenges identified in prior work. Therefore, more rigorous evaluations of both existing and new technologies, including clinical trials that effectively measure the impact of the technology on both HHAs and patients for whom they care, are warranted.

Acknowledgments

The authors would like to acknowledge Andrea Gallardo for her contributions to the landscape analysis. This work was supported by funding from the Robert Wood Johnson Foundation (76487), the National Science Foundation (2026577), and the National Heart, Lung, and Blood Institute (grant K23HL15060).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy terms for scoping review.
[DOCX File , 13 KB - jmir_v24i11e39997_app1.docx]

Multimedia Appendix 2
Full results from the scoping review of literature.
[DOCX File , 18 KB - jmir_v24i11e39997_app2.docx]

Multimedia Appendix 3
Search strategy terms for landscape analysis.
[DOCX File , 13 KB - jmir_v24i11e39997_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Full results from the landscape analysis of mobile apps.
[DOCX File , 21 KB - jmir_v24i11e39997_app4.docx]

References

1. U.S. Home Care Workers: Key Facts (2019). Paraprofessional Healthcare Institute. 2019 Sep 3. URL: https://phinational.org/resource/u-s-home-care-workers-key-facts-2019/ [accessed 2020-02-18]
2. Bercovitz A, Moss A, Sengupta M, Park-Lee EY, Jones A, Harris-Kojetin LD. An overview of home health aides: United States, 2007. Natl Health Stat Report 2011 May 19(34):1-31 [FREE Full text] [Medline: 21688727]
3. Sterling MR, Shaw AL, Leung PB, Safford MM, Jones CD, Tsui EK, et al. Home care workers in heart failure: a systematic review. J Multidiscip Healthc 2018 Sep 25;11:481-492 [FREE Full text] [doi: 10.2147/JMDH.S175512] [Medline: 30288046]
4. Home Health and Personal Care Aides. U.S. Bureau of Labor Statistics. URL: https://www.bls.gov/ooh/healthcare/home-health-aides-and-personal-care-aides.htm [accessed 2020-02-18]
5. Sterling MR, Silva AF, Leung PB, Shaw AL, Tsui EK, Jones CD, et al. "It's like they forget that the word 'health' is in 'home health aide'": understanding the perspectives of home care workers who care for adults with heart failure. J Am Heart Assoc 2018 Dec 04;7(23):e010134 [FREE Full text] [doi: 10.1161/JAHA.118.010134] [Medline: 30571599]
6. Sterling MR, Dell N, Piantella B, Cho J, Kaur H, Tseng E, et al. Understanding the workflow of home health care for patients with heart failure: challenges and opportunities. J Gen Intern Med 2020 Jun;35(6):1721-1729 [FREE Full text] [doi: 10.1007/s11606-020-05675-8] [Medline: 32026253]
7. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
8. Vanagas G, Engelbrecht R, Damaševičius R, Suomi R, Solanas A. eHealth solutions for the integrated healthcare. J Healthc Eng 2018 Jul 10;2018:3846892 [FREE Full text] [doi: 10.1155/2018/3846892] [Medline: 30123441]
9. Leung PB, Silva AF, Cho J, Kaur H, Lee A, Escamilla Y, et al. Eliciting the educational priorities of home care workers caring for adults with heart failure. Gerontol Geriatr Educ 2022;43(2):239-249. [doi: 10.1080/02701960.2020.1793760] [Medline: 32666906]
10. Sterling MR, Dell N, Tseng E, Okeke F, Cho J, Piantella B, et al. Home care workers caring for adults with heart failure need better access to training and technology: a role for implementation science. J Clin Transl Sci 2020 Apr 06;4(3):224-228 [FREE Full text] [doi: 10.1017/cts.2020.36] [Medline: 32695493]
11. 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]
12. 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]
13. Ogawa H, Yonezawa Y, Maki H, Sato H, Hahn AW, Caldwell WM. A Java mobile phone-based "Home Helper" care report creation support system. Biomed Sci Instrum 2004;40:76-79. [Medline: 15133938]
14. Scandurra I, Hägglund M, Koch S. Visualisation and interaction design solutions to address specific demands in shared home care. Stud Health Technol Inform 2006;124:71-76. [Medline: 17108506]
15. Paganelli F, Giuli D. An ontology-based system for context-aware and configurable services to support home-based continuous care. IEEE Trans Inf Technol Biomed 2011 Mar;15(2):324-333. [doi: 10.1109/TITB.2010.2091649] [Medline: 21075729]
16. Page TF, Brown EL, Ruggiano N, Roberts L, Hristidis V. Improving care delivery using health information technology in the home care setting: development of the home continuation care dashboard. Ann Long Term Care Clin Care Aging 2012 Dec;20(12):26-30 [FREE Full text]
17. De Backere F, Ongenae F, Vannieuwenborg F, Van Ooteghem J, Duysburgh P, Jansen A, et al. The OCareCloudS project: toward organizing care through trusted cloud services. Inform Health Soc Care 2016;41(2):159-176. [doi: 10.3109/17538157.2014.965306] [Medline: 25325572]
18. De Backere F, Bonte P, Verstichel S, Ongenae F, De Turck F. The OCarePlatform: a context-aware system to support independent living. Comput Methods Programs Biomed 2017 Mar;140:111-120. [doi: 10.1016/j.cmpb.2016.11.008] [Medline: 28254067]
19. Danilovich MK, Diaz L, Saberbein G, Healey WE, Huber G, Corcos DM. Design and development of a mobile exercise application for home care aides and older adult medicaid home and community-based clients. Home Health Care Serv Q 2017;36(3-4):196-210 [FREE Full text] [doi: 10.1080/01621424.2017.1381869] [Medline: 28937931]
20. Bourikas L, Allott N, Bahaj A, Connelly J, Gauthier S, James P. Elderly Support To Inspired Ageing (ESTIA). University of Southampton Institutional Repository. Edinburgh, UK: Passive Low Energy Architecture; 2017. URL: https://eprints.soton.ac.uk/419715/1/1199_v2.pdf [accessed 2022-10-21]
21. Danilovich MK, Hughes SL, Corcos DM, Marquez DX, Eisenstein AR. Translating strong for life into the community care program: lessons learned. J Appl Gerontol 2017 May;36(5):553-569 [FREE Full text] [doi: 10.1177/0733464815625833] [Medline: 26912729]

XSL•FO
RenderX

22. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA, USA: Sage Publications; 2004:445-456.
23. Muller MJ, Kuhn S. Participatory design. Commun ACM 1993 Jun;36(6):24-28. [doi: 10.1145/153571.255960]
24. Schuler D, Namioka A. Participatory Design: Principles and Practices. London, UK: Routledge; 1993.
25. Olson JS, Kellogg WA. Ways of Knowing in HCI. New York, NY, USA: Springer; 2014.
26. Tseng E, Okeke F, Sterling M, Dell N. "We can learn. Why not?": designing technologies to engender equity for home health aides. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Apr Presented at: CHI '20; April 25-30, 2020; Honolulu, HI, USA p. 1-14. [doi: 10.1145/3313831.3376633]
27. Sterling MR, Tseng E, Poon A, Cho J, Avgar AC, Kern LM, et al. Experiences of home health care workers in New York City during the coronavirus disease 2019 pandemic: a qualitative analysis. JAMA Intern Med 2020 Nov 01;180(11):1453-1459 [FREE Full text] [doi: 10.1001/jamainternmed.2020.3930] [Medline: 32749450]
28. Vaughan G, Prizeman G, Eustace-Cook J, Byrne G. Use of mHealth apps by nurses in the management of chronic wounds: a scoping review protocol. JBI Evid Synth 2021 Mar 02;19(10):2783-2789. [doi: 10.11124/JBIES-20-00401] [Medline: 33651753]
29. Dauletbaev N, Kuhn S, Holtz S, Waldmann S, Niekrenz L, Müller BS, et al. Implementation and use of mHealth home telemonitoring in adults with acute COVID-19 infection: a scoping review protocol. BMJ Open 2021 Sep 27;11(9):e053819 [FREE Full text] [doi: 10.1136/bmjopen-2021-053819] [Medline: 34580103]
30. Widdison R, Rashidi A, Whitehead L. Effectiveness of mobile apps to improve urinary incontinence: a systematic review of randomised controlled trials. BMC Nurs 2022 Jan 28;21(1):32 [FREE Full text] [doi: 10.1186/s12912-022-00812-6] [Medline: 35090464]

Abbreviations

HHA: home health aide
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by R Kukafka; submitted 06.06.22; peer-reviewed by M Attridge, R Damaševičius; comments to author 25.06.22; revised version received 08.07.22; accepted 28.09.22; published 11.11.22.

Please cite as:
Kuo EFC, Cho J, Olaye I, Delgado D, Dell N, Sterling MR
Understanding the Technological Landscape of Home Health Aides: Scoping Literature Review and a Landscape Analysis of Existing mHealth Apps
J Med Internet Res 2022;24(11):e39997
URL: https://www.jmir.org/2022/11/e39997
doi:10.2196/39997
PMID:36176033

©Elizabeth Fong-Chy Kuo, Jacklyn Cho, Iredia Olaye, Diana Delgado, Nicola Dell, Madeline R Sterling. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Consumer-Generated Discourse on Cannabis as a Medicine: Scoping Review of Techniques

Sedigheh Khademi Habibabadi[1*], PhD; Christine Hallinan[1,2*], PhD; Yvonne Bonomo[1], MD, PhD; Mike Conway[3*], PhD

[1]Department of General Practice, Melbourne Medical School, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia

[2]Health & Biomedical Research Information Technology Unit, The University of Melbourne, Melbourne, Australia

[3]School of Computing & Information Systems, The University of Melbourne, Melbourne, Australia

[*]these authors contributed equally

Corresponding Author:
Sedigheh Khademi Habibabadi, PhD
Department of General Practice
Melbourne Medical School
Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne
Grattan Street, Parkville, Victoria
Melbourne, 3010
Australia
Phone: 61 405761879
Email: sedigh.khademi@unimelb.edu.au

Abstract

Background: Medicinal cannabis is increasingly being used for a variety of physical and mental health conditions. Social media and web-based health platforms provide valuable, real-time, and cost-effective surveillance resources for gleaning insights regarding individuals who use cannabis for medicinal purposes. This is particularly important considering that the evidence for the optimal use of medicinal cannabis is still emerging. Despite the web-based marketing of medicinal cannabis to consumers, currently, there is no robust regulatory framework to measure clinical health benefits or individual experiences of adverse events. In a previous study, we conducted a systematic scoping review of studies that contained themes of the medicinal use of cannabis and used data from social media and search engine results. This study analyzed the methodological approaches and limitations of these studies.

Objective: We aimed to examine research approaches and study methodologies that use web-based user-generated text to study the use of cannabis as a medicine.

Methods: We searched MEDLINE, Scopus, Web of Science, and Embase databases for primary studies in the English language from January 1974 to April 2022. Studies were included if they aimed to understand web-based user-generated text related to health conditions where cannabis is used as a medicine or where health was mentioned in general cannabis-related conversations.

Results: We included 42 articles in this review. In these articles, Twitter was used 3 times more than other computer-generated sources, including Reddit, web-based forums, GoFundMe, YouTube, and Google Trends. Analytical methods included sentiment assessment, thematic analysis (manual and automatic), social network analysis, and geographic analysis.

Conclusions: This study is the first to review techniques used by research on consumer-generated text for understanding cannabis as a medicine. It is increasingly evident that consumer-generated data offer opportunities for a greater understanding of individual behavior and population health outcomes. However, research using these data has some limitations that include difficulties in establishing sample representativeness and a lack of methodological best practices. To address these limitations, deidentified annotated data sources should be made publicly available, researchers should determine the origins of posts (organizations, bots, power users, or ordinary individuals), and powerful analytical techniques should be used.

(J Med Internet Res 2022;24(11):e35974) doi:10.2196/35974

KEYWORDS

social media; data mining; internet and the web technology; consumer-generated data; medicinal cannabis; medical marijuana

XSL•FO
RenderX

Introduction

Medicinal Cannabis Pharmacovigilance

Cannabis has been widely used for a variety of purposes, including medicinal applications, throughout human history. Over the last century, its use has been prohibited in Europe, Northern America, and Australasia [1]. Since 2016, these jurisdictions have incrementally authorized the use of medicinal cannabis for certain conditions [2]. Given the substantial public interest in cannabis as medicine, there is a pressing need to better understand its safety and efficacy.

However, aside from clinical trials, there are scant data regarding the efficacy and side effects of medicinal cannabis [3-6]. One of the main methods for postmarketing safety surveillance of medications is the use of established pharmacovigilance reporting systems, which rely on reporting of adverse events by individuals [7-9]. Cannabis users are often unaware of these systems or the importance of reporting. They may find them too difficult to use or may not want to divulge personal details if these are required [10]. Users may not even think of reporting their side effects because they consider them an inherent experience of cannabis consumption, especially if they are not using an approved medical cannabis product.

Increasing the understanding of the efficacy and safety of cannabis as medicine is warranted because cannabis is a nonstandardized product, given the wide variety in growing conditions and production specifications [11]. This includes variations in climate, soil (or other growth media), water, light, and other factors that affect plant growth. Even if cannabis medicines in a country or state must adhere to mandatory standards (good manufacturing practice), some cannabis users prefer to grow or import their own cannabis [12]. These factors make the systematic assessment of the effectiveness of medical cannabis and its side effects difficult.

Social Media as a Pharmacovigilance Data Source

To gain additional insights into cannabis use and its effects, researchers are now turning to social media and web-based health forums. These platforms are a place for both patients and the general population to freely express and exchange their experiences and thus provide a valuable additional data source for monitoring public health [13]. Unlike other forms of highly curated data collection methods, such as surveys or interviews, social media provides an organic view of everyday thoughts, behaviors, and activities of people. Therefore, social media has the potential to provide insights beyond the boundaries of targeted investigations, including emergent events, observations of behavioral phenomena and subcultures, and insights for the social sciences [14].

The information contained in social media conversations is voluminous and not only potentially rich in content but also complex and varied. As an unstructured raw data source, credible information may be sparse and difficult to identify; there may be uncertainty about the origin of the data or the population they represent [15]. Furthermore, it is difficult to interpret the informal language and structure of social media posts, which are confounded by many competing sources, such as promotional posts, hashtags, and social media bots [16,17]. Social media bots automatically create content and interact with social media platform users [18]. A study found that between 9% and 15% of Twitter accounts are bots [19]. Notwithstanding these limitations, if these complexities can be successfully navigated, social media has the potential to be a great asset for increased understanding of cannabis as a medicine.

Our previous systematic scoping review [20] used PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [21] to understand the utility of web-based user-generated text in providing insight into the use of cannabis as a medicine. This paper examines the techniques, analyses, and limitations of these studies.

The objective of this research was to provide a review of studies that have used user-generated data in conjunction with computational methods to understand the medicinal use of cannabis in a population. We addressed the following research questions (RQs):

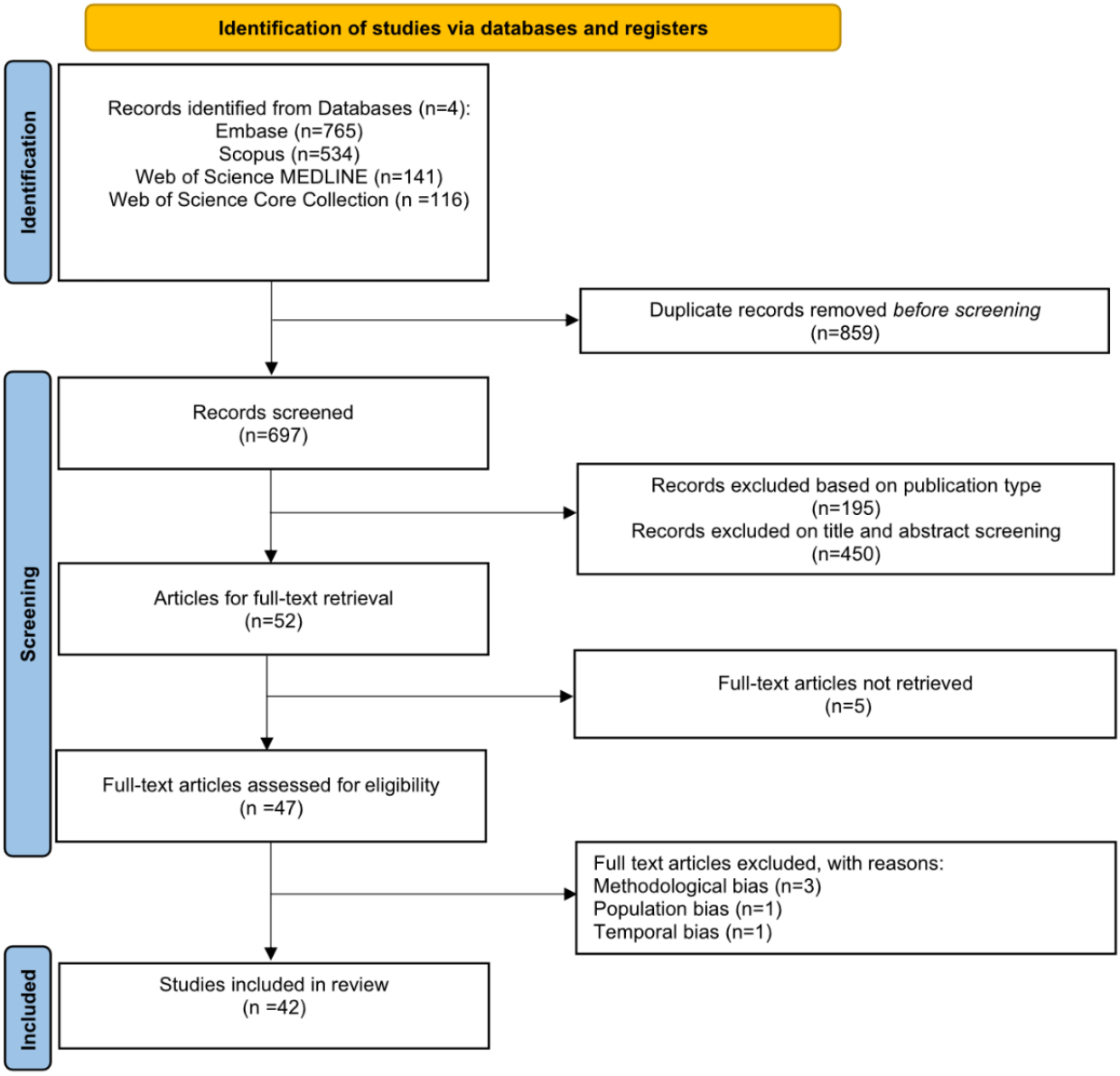
- RQ1: What consumer-generated data sources are used for studying cannabis?
- RQ2: What common techniques for collection and analysis of data are used?
- RQ3: What are the common limitations and challenges faced by the studies?

Methods

We searched for English-language studies that were indexed in MEDLINE, Embase, Web of Science, and Scopus databases and published between January 2010 and March 2022. Literature database queries were developed for these 4 databases. See Table S1 in Multimedia Appendix 1 [22-63] for the details of search terms used and Multimedia Appendix 1 Table S2 for the inclusion and exclusion criteria of the selected articles. A summary of the PRISMA flowchart is shown in Figure 1 [20].

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the study selection process [20].



Results

Overview

Table 1 provides a summary of each article that includes author names, publication year, data source, and duration of data collection, analysis, and number of items analyzed.

The year with the highest number of publications was 2020 (11/42, 26%), followed by 2017 and 2021 (6/40, 14%). Of the 42 studies, 6 (12%) were conducted in 2015 and 2019. The number of publications per year is shown in Table 2.

Regarding data sources, Twitter was used in 40% (17/42) of the reviewed studies, around 3 times the number of studies using either Reddit or web-based forums 14% (6/42). GoFundMe, YouTube, and Google Trends comprised 7% (3/42) of the total. Text was the focus of 83% (35/42) of the studies, whereas the others analyzed trends, videos, search logs, and images. Table 3 shows the distribution of the publications selected per data source.

XSL•FO
RenderX

Table 1. Articles included in the review.

Study	Source (duration)	Analysis	Number of items analyzed
McGregor et al [22], 2014	Web-based forums, Facebook, Twitter, and YouTube (not available)	• Thematic and content analysis of glaucoma-related posts on the following: • Analysis of the nature of the post (personal stories, information sharing or flagging, supportive comments, questions, answers, and general discussions) • Sentiment analysis (positive or negative)	3785 items
Cavazos-Rehg et al [23], 2015	Twitter (February to March 2014)	• Cannabis-related chatter by influential users on the following: • Sentiment analysis by using the Likert scale • Thematic analysis of tweets • Demographic analysis	7000 tweets
Daniulaityte et al [24], 2015	Twitter (October to December 2014)	• US dab-related tweets: • Counting and normalizing based on cannabis legalization policy	125,255 tweets (27,018 geolocated tweets)
Gonzalez-Estrada et al [25], 2015	YouTube (June 4-8, 2014)	• Content analysis of asthma-related videos on the following: • Source: professional society, media, asthma care provider, etc • Content: personal experience, medical professional, advertisement, patient education, alternative treatment, or to increase awareness • Quality scoring of misleading and useful info • Video characteristics or video statistics	200 most viewed videos
Krauss et al [26], 2015	YouTube (January 22, 2015)	• Analysis of dabbing-related videos on the following: • Characteristics of the people dabbing (age and skills) • Characteristics of the session • Messages included in the videos	116 videos
Thompson et al [27], 2015	Twitter (March 2012 to July 2013)	• Content analysis of cannabis-related tweets and retweets on the following: • Adolescence users (age, inferred from the user profile) • Sentiment (positive, negative, or unclear) • Subject (self, other, general, or subject unclear) • Use category (own use, use by others, or not mentioned) • Related behaviors (habitual use, social aspect, etc) • Positive aspects (better than other drugs and medical use)	36,939 original tweets and 10,000 retweets
Cavazos-Rehg et al [28], 2016	Twitter (January 2015)	• Dabbing-related tweets: • Thematic analysis of tweets to 7 themes • Subanalysis of 1 theme (extreme effects) into physiological or psychological effects • Geotagged tweets analysis for number per state • Demographic analysis	5000 tweets
Lamy et al [29], 2016	Twitter (May to July 2015)	• Content analysis of cannabis edible-related conversations: • Tweet sources (media, retail, or users) • Sentiment analysis (positive, negative, or neutral) • Word frequency analysis • Geotagging (policy impact on the volume of tweets)	3000 tweets
Mitchell et al [30], 2016	Web-based forums (October 2014)	• Thematic analysis of ADHD[a] and cannabis web-based forum posts on the following: • Impact of cannabis on ADHD symptoms (therapeutic, harmful, both, and none) • Other domains (mood, psychiatric conditions, and other [sleep]) • Comments about cannabis as medicinal (more effective than other ADHD medications, less effective, or not legal)	268 threads

XSL•FO
RenderX

Study	Source (duration)	Analysis	Number of items analyzed
Andersson et al [31], 2017	Web-based forums (April 18-19, 2016)	• Thematic analysis of conversations on headache-related posts	32 topics

Dai and Hao [32], 2017	Twitter (August 2015 to April 2016)	• Naive Bayes classifier on PTSD[b] and cannabis-related tweets:
		  • Sentiment analysis
		  • Analysis of prevalence of support of cannabis use for PTSD in association with state level legislation and socioeconomic factors
			66,000 cannabis-related and 31,184 geolocated tweets

Greiner et al [33], 2017	Web-based forums (November 2014 to March 2015)	• Content analysis of cannabis help forums on the following:
		  • Fields of interest (illness-related, social, financial, and legal issues)
		  • Self-help mechanisms (exchange of information, emotional support, group support)
		  • Analysis of sex and age when available
		  • Highly involved vs moderately involved users
			717 posts

Turner and Kantardzic [34], 2017	Twitter (August 2015 to April 2016)	• Supervised and unsupervised machine learning techniques of cannabis-related tweets:
		  • Binary classification to identify marijuana-related tweet
		  • Topic modeling
		  • User social network analysis
		  • Spatiotemporal analysis of conversations
			40,509 geolocated tweets

Westmaas et al [35], 2017	Web-based forums (January 2000 to December 2013)	• Topic modeling of Cancer Survivors Network:
		  • Analyze smoking or cessation-related content
		  • Analysis to determine the overall context in which these discussions occurred
			468,000 posts

Yom Tov and Lev Ran [36], 2017	Bing logs (November 2016 to April 2017)	• Statistical analysis of cannabis-related query logs	Not available

Cavazos-Rehg et al [37], 2018	YouTube (June 10-11, 2015)	• Cannabis review web-based videos:
		  • Sentiment analysis
		  • Physical or mental effects; is it promotional, encourage follow-up; depiction of consumption; video details and engagement statistics
		  • Current users survey (demographics, reason for use, and use of reviews)
			83 videos

Glowacki et al [38], 2018	Twitter (August to October 2016)	• Statistical analysis on opioid-related tweets:
		  • Clustering algorithm to find topics
		  • Analysis of trending hashtags, top influencers, and location of tweets
			73,235 tweets

Meacham et al [39], 2018	Reddit (January 2010 to December 2016)	• Analysis of modes of cannabis use mentions on Twitter on the following:
		  • Most frequent words
		  • Mentions of adverse effects
		  • Subjective highness
			400,000 posts

Leas et al [40], 2019	Google Trends (January 2004 to April 2019)	• Analysis on CBD[c] and cannabidiol terms to evaluate public interest	Not available

Meacham et al [41], 2019	Reddit (January 2017 to December 2019)	• Content analysis of dabbing-related questions on the following:
		  • Topics of questions
		  • After engagement and the types and sentiment of information
			193 questions

Nasralah et al [42], 2019	Twitter (January 2015 to February 2019)	• Analysis of opioid-dependent user's tweets:
		  • Thematic analysis of conversations
		  • Demographic analysis
			20,609 tweets

XSL•FO
RenderX

Study	Source (duration)	Analysis	Number of items analyzed
Pérez-Pérez et al [43], 2019	Twitter (February to August 2018)	• Lexicon- and rule-based analysis of bowel disease tweets on sentiments, network, gender, geolocation, symptoms, and food	24,634 tweets
Shi et al [44], 2019	Google Trends and Buzzsumo (January 2011 to July 2018)	• Google Trends analysis on cancer therapies to evaluate interest in cannabis vs other therapies	Not available
Allem et al [45], 2020	Twitter (May to December 2018)	• Topic analysis of cannabis-related tweets	60,861 nonbot and 8874 bot tweets

Janmohamed et al [46], 2020	Blogs, news, forums, and <1% other (August 2019 to April 2021)	• Topic modeling on vaping-related conversations:
		  • Analysis of word prevalence
		  • Analysis of change of topics over time
			4,027,172 documents or blogs

Jia et al [47], 2020	Google, Facebook, and YouTube (September 2019)	• Content analysis of glaucoma and CBD posts on the following:
		  • General discussion, information sharing, personal story, question, answer, and moderator comment
		  • Quality of information
		  • Source of information being professional or not and whether an opinion on glaucoma and medical cannabis use was expressed
		  • Analysis of professional accounts
			51 Google websites, 126 Facebook posts, and 37 YouTube videos

Leas et al [48], 2020	Reddit (January 2014 to August 2019)	• Content analysis of reasons for CBD use:
		  • Reasons for personal use (condition and wellness)
		  • Analysis based on categorized diagnosable conditions
			104,917 posts

Merten et al [49], 2020	Pinterest (July 31, August 18, and September 1, 2018)	• Content analysis of CBD and cannabidiol posts on the following:
		  • Mentions of mental and physical benefits
		  • Emotional appeal analysis
		  • Engagement statistics
			1280 pins

Mullins et al [50], 2020	Twitter (June to July 2017)	• Analysis of Ireland pain-related tweets on: • Topic analysis: sentiment analysis, analysis of most frequently occurring keywords, demographic analysis, and personal use analysis	941 tweets
Saposnik and Huber [51], 2020	Google Trends (January 2004 to December 2019)	• Google Trends analysis on autism and cannabis to analyze trends in search volume about the causes and treatments of Autism spectrum disorder over time	Not available

Song et al [52], 2020	GoFundMe (January 2012 to December 2019)	• Content analysis of alternative medicine and cancer campaigns on the following:
		  • Themes of patient narratives
		  • Types of alternative treatments used
		  • Demographics (gender, cancer type, cancer stage, insurance status, past treatment, future treatment, and alternative treatment)
			1474 campaigns

Tran and Kavuluru [53], 2020	Reddit and or FDA comments (January to April 2019)	• Content analysis on CBD posts for therapeutic effects and popular modes of consumption compared with FDA[d] comments	64,099 Reddit and 3832 FDA comments

van Draanen et al [54], 2020	Twitter (January 2017 to June 2019)	• Cannabis-related US and Canada posts:
		  • Topic modeling
		  • Sentiment analysis based on cannabis legalization policies
			1,200,127 tweets

Zenone et al [55], 2020	GoFundMe (January 2017 to March 2019)	• Thematic analysis of cancer and cannabis campaigns:
		  • Efficacy claims
		  • Treatment regimen classification
		  • CBD efficacy presentation
		  • Content analysis for Other: cancer stage, raised money, and number of donors
			155 campaigns

XSL•FO
RenderX

Study	Source (duration)	Analysis	Number of items analyzed
Pang et al [56], 2021	Twitter (December 2019 to December 2020)	• Thematic analysis of pregnancy- and cannabis-related tweets for safety during pregnancy, safety postpartum, and pregnancy-related symptoms	17,238 tweets

Rhidenour et al [57], 2021	Reddit (January 2008 to December 2018)	• Thematic analysis of veteran's cannabis posts on the following:
		  • Point of view, reasons for use, prescription drug use, or other substance use
		  • Test, legality, legal policy, and doctor-patient conversation
			974 posts

Smolev et al [58], 2021	Facebook (November 2018 to November 2019)	• Thematic analysis of traumatic brachial plexus injury posts on: antiopioid sentiment, preference for alternative options, and antigabapentin sentiment	7694 posts
Soleymanpour et al [59], 2021	Twitter (July 2019)	• Analysis of CBD marketing tweets and therapeutic claims	2,200,000 tweets
Zenone et al [60], 2021	GoFundMe (June 2017 to May 2019)	• Thematic analysis for informational pathways: self-directed research, recommendations from a trusted care provider, and insights shared by someone associated with or influencing the crowd funders personal network • Content analysis for intended outcome, social media shares, number of donors, total requested, and total received	164 campaigns
Turner et al [62] 2021	Twitter (October 2019 to January 2020)	• Analysis of personal and commercial CBD-related tweets; term and sentiment analysis	167,755 personal 143,322 commercial tweets
Allem et al [61], 2022	Twitter (January to September 2020)	• Analysis of cannabis-related conversation for health-related motivations or perceived adverse health effects	353,353 tweets
Meacham et al [63] 2022	Reddit (December 2015 to August 2019)	• Analysis of cannabis-related posts from an opioid use and an opioid recovery subreddit	908 posts from opioid recovery subreddits and 4224 posts from opioid use subreddits

[a]ADHD: attention-deficit hyperactivity disorder.

[b]PTSD: posttraumatic stress disorder.

[c]CBD: cannabidiol.

[d]FDA: Food and Drug Administration.

Table 2. Publications per year (n=42).

Year	Count, n (%)
2014	1 (2)
2015	5 (12)
2016	3 (7)
2017	6 (14)
2018	3 (7)
2019	5 (12)
2020	11 (26)
2021	6 (14)
2022	2 (5)

XSL•FO
RenderX

Table 3. Publications per data source (n=42).

Source	Count, n (%)
Twitter	17 (41)
Reddit	6 (14)
Web-based forums	6 (14)
GoFundMe	3 (7)
YouTube	3 (7)
Google Trends	3 (7)
Google, Facebook, and YouTube	1 (2)
Bing Search Engine	1 (2)
Facebook	1 (2)
Pinterest	1 (2)

Social Media Data Collection Strategies

Some studies obtained all their associated data from a specific subreddit [48,53,57] or a web-based forum [35] and subsequently sampled the data. Of 42 studies, 1 (2%) Twitter study collected tweets using a geolocation boundary box and then filtered the data for cannabis-related keywords [54].

Keyword-based filtering was used by many studies. Terms used for filtering were either common expressions for cannabis from dictionaries, such as Urban Dictionary, or were based on similar research in this domain. Of the 42 studies,1 (2%) study [36] used Urban Dictionary and web forums to create a comprehensive list of 123 terms related to cannabis consumption. Another study [57] first found all the terms related to marijuana by searching on Thesaurus.com and then used the word embedding likeness perusal software [64] to generate synonyms.

In a nonmedical cannabis-related study, word embeddings created from Twitter and Reddit data sets discovered synonyms and slang terms that could not be identified using other means. The study recommends this method of synonym discovery in advance for any data collection based on keyword filtering [65].

Of the 42 studies, 3 (7%) studies were user focused, with data derived from specific highly influential users [23], opioid-dependent users [42], or a US veteran-specific subreddit [57].

The largest data set manually annotated by the researchers was collected using cannabis-related keywords and consisted of 36,939 original tweets and 10,000 retweets [27]. Apart from that study, the average size of annotated data sets was approximately 1450 records. Of the 42 studies, 2 (5%) studies [23,28] used crowdsourcing services to annotate tweets, whereas the rest conducted in-house annotation. The duration of data collection ranged from 1 month to 6 years. Of the 42 studies, 2 (5%) of these studies made their annotated data available to other researchers [30,60].

Types of Analysis

Overview

The studies included in this review used a variety of analytical methods, including qualitative analysis, quantitative content analysis, machine learning, rule-based, and statistical analysis. The types of analysis include sentiment assessment, thematic analysis, content analysis, named entity recognition, social networks, and geographic analysis. Table S3 in Multimedia Appendix 1 summarizes the analyses.

Discovering Themes

Themes were identified in 62% (26/42) of the studies. Manual coding of the themes was performed by 69% (18/26) of the studies, either by using pre-existing categories or by observing a sample of the data and generating a codebook [22,23,25,26,28,30,31,37,41,47-49,52,55-58,60]. Of the 26 studies, 2 (7%) studies used the services of social media data analytics companies [42,50].

Of the 26 studies, 4 (15%) studies used topic modeling to infer themes or topics [34,35,46,54]. The algorithm of choice for this task is the latent Dirichlet allocation [66]. The choice of the number of topics was based on intrinsic evaluation metrics (eg, coherence and perplexity) and iterative qualitative analysis informed by prior experience with topic models. Of the 26 studies, 1 (4%) study used temporal topic modeling techniques to study changes in topics over time, with the goal of analyzing how web-based vaping narratives changed during the COVID-19 pandemic [46].

Of the 26 studies, 1 (4%) study identified themes by using rule-based methods. Frequency counts of the most common unigrams and bigrams were generated and formed the basis of the topics [45]. Another study used SAS Text Miner software, a text-topic node algorithm, to discover topics [38].

Demographic Analysis

Socioeconomic and demographic analyses of the study population were performed in 26% (11/42) of the studies. Of the 11 studies, 2 (27%) studies used the provided gender, age, and other user characteristics from user profiles or inferred from posts by users [33,52]. Of the 11 studies, 2 (27%) video-based

XSL•FO
RenderX

studies used the perceived age and gender of the subjects after observing the videos [25,26].

Of the 11 studies, 2 (18%) studies that used social media analytics providers obtained age and gender data by using the supplied analysis [42,50]. Of the 11 studies, 2 (18%) of the Twitter-based studies used a commercial tool called DemographicsPro, which uses proprietary algorithms to infer user demographic characteristics [23,28]. Other studies used existing census data [32], demographic information obtained from survey data [37], and a 2-step method based on a gender-name lexicon and a face recognition algorithm applied to users' profile information to identify the users' gender [43].

Geographic Analysis

Geolocation data analysis was performed in 40% (17/42) of the studies. User profiles or message metadata were used in 52% (9/17) of the studies [24,29,32,34,36,43,54,55,60]. Of the 17 studies, 2 (12%) studies used information provided by social media analytics companies [38,50]. The DemographicsPro tool was used in 5% (1/17) of studies [28]. Of the 17 studies, 3 (17%) studies used location information provided by Google Trends [40,44,51]. Another (1/17, 5%) study collected geographical information from survey data [37]. Of the 17 studies,1 (5%) video-based study used the geographic location of video channels [26].

Sentiment Assessment

An individual's perception of a topic can be characterized as having a positive, negative, or neutral sentiment. The analysis of these sentiments is often performed using automated language tools and is named "sentiment analysis" [67].

Out of the 12 studies that performed sentiment analysis, 5 (42%) used automated methods. Of the 12 studies, 1 (8%) study trained a binary Naive Bayes classifier on a sample of 1000 "marijuana" related tweets to classify posts into 2 opinion polarities, positive and negative or neutral [32]. Another study used sentiment analysis provided by a social media analytics company [50]. Of the 12 studies, 3 (25%) studies used Valence Aware Dictionary and Sentiment Reasoner (VADER) [68], a lexicon and rule-based sentiment analysis tool [43,54,62]. The VADER performance was compared with in-house machine learning classifiers trained on 3000 manually coded cannabis-related tweets, which showed a 30% performance improvement over VADER. Although VADER is widely used for general tweet sentiment analysis, its performance suffers in substance-use-related domains where negative words are often used to carry positive sentiments. For example, "I took CBD oil, that stuff was bad" [69]—in this sentence, "bad" actually means good.

User Analysis

For conducting user analysis, 57% (24/42) of the studies examined either the subject of the posts, as from individuals or others (ie, from self, retail, media, or professionals), or who the post was about (self, others, or general) [22,23,25-29,33,37,41-43,45,47-50,52,55,57,58,60-62].

When manual data labeling was performed, the determination of both the poster and subject of the post was part of the labeling process. Self-reporting and self-use were easily determined by observation of videos, as were most texts based on the structure of the language. For example, a study [27] first identified whether the subject of the tweet was about the self, other, or general and then identified whether the tweets were about actual cannabis use. This study included further categories of tone, related behavior, perceived impact, and social context. Automated labeling approaches look for phrases that indicate self-reporting. For example, a study on opioid addiction [42] looked for phrases such as "I am addicted" and "I have been addicted" in the context of opioid mentions. Classifiers were used in another study [59] to separate marketing tweets from nonmarketing tweets; however, their focus was on marketing tweets.

None of the studies used advanced natural language processing techniques to establish subjects and personal mentions. Social media bots are automated accounts that generate artificial activities on social media platforms [18]. Bot detection was used in only 4% (1/24) of studies, which used Twitter as a data source [45].

Other Analyses

Of the 42 studies, 2 (5%) studies examined the social networks of contributors to conversations. This allowed the identification of target communities and user interactions [34,43]. Of the 42 studies, 3 (7%) studies examined the impact of governmental cannabis legalization policies on the sentiments and opinions of people or on the volume of social media posts [24,28,54]. Term frequency and count analysis of words and phrases was performed in 12% (5/42) of studies [29,39,50,62,63].

Ethical Considerations

Institutional review boards (or their equivalents) ensured that research using human participants is conducted in an ethical manner [70]. Approval for and overseeing of a study by an institutional review board ensures that researchers adopt an ethically appropriate research protocol that respects the rights and interests of social media users; 62% (26/42) of the studies mentioned an ethics approval review being sought or the study being exempt from ethics requirements. There was no mention of ethics approval in 38% (16/42) of the studies.

External Validity

The use of standard reporting systems, such as the US Food and Drug Administration reports, helps to assess whether social media research findings can be generalized to real-world data. When a suitable ground-truth data set is not available, validating results against >1 social media platform improves the generalizability and validity of the results. Only a few studies used >1 social media data source or validated their findings against other data sources. Of the 42 studies, 2 (5%) studies used Food and Drug Administration data as an external ground-truth data source to validate their results [36,53]. Of the 42 studies, 1 (2%) study analyzed several web-based forums [31], and 2 (5%) other studies used several social media platforms as their data sources [22,47].

XSL•FO
RenderX

Discussion

In this study, we reviewed the technical aspects of peer-reviewed published works that used social media and other forms of user-generated data to understand the medicinal use of cannabis. All the studies concluded that these consumer-generated data sources are useful and provide a complementary resource for studying cannabis and medical conditions for which cannabis is used.

Principal Findings

The findings of this study are presented by answering the RQs.

RQ1: What Consumer-Generated Data Sources Are Used for Studying Cannabis?

Sources of consumer-generated data for cannabis research used by the reviewed studies include social media platforms, such as Twitter, Reddit, and YouTube; search queries, including Google Trend and Bing query logs; and web-based forums, crowdfunding platforms, blogs, and websites. Twitter was used in most of the studies. One of the studies concluded that, compared with unmoderated platforms, moderated sites focused more on evidence-based information and controlled misleading content [22].

RQ2: What Common Techniques for Collection and Analysis of Data Are Used?

Some studies have used social media analytics companies for some or all of their data collection and processing tasks. Other studies used application program interfaces to interact with Twitter and Reddit. Although Facebook allows researchers to access public posts from public pages through a dedicated platform [71], 2% (1/42) of studies [58] analyzed private Facebook posts—the method used to obtain data was not reported.

Approximately half of the studies used data sets of <8000 records and many of them used 1000 records. These studies either focused on understanding the characteristics and needs of users or the quality of information on the web, or they were directed by an RQ such as “Are individuals using CBD for diagnoseable conditions which have evidence-based therapies?” These analyses play a critical role in understanding the domain but are difficult to replicate and generalize.

More recent neural network–based natural language processing techniques have not been used in the studies in this review. These modern machine learning methods have the advantage that they require minimal data preparation and are characterized by the capacity to learn the nuance of language. However, to function effectively, they typically require high-quality annotated data—a scarce and expensive resource. Textual social media data are highly amenable to these techniques. Creating and sharing deidentified annotated data sets for this purpose should be encouraged within appropriate ethical, regulatory, and legal frameworks [72].

RQ3: What Are the Common Limitations and Challenges Faced by the Studies?

These limitations are mentioned in order of frequency.

Sample Representativeness

Most research on social media uses samples of available data. However, the extent to which the data samples are representative of the general population is often unclear. The limiting factors mentioned in these studies include sampling bias that is introduced as a result of the choice of keywords, data collection duration, and population biases.

Population biases often refer to the demographic composition of people using social media platforms being different from the general population and the difficulties in determining the demographic characteristics of users. Accessing accurate geographical locations has also been mentioned in previous studies. Obtaining these data is limited because even when users explicitly include demographic information (eg, with Facebook) or geographical information in their posts or profiles, these may be fabricated.

The choice of platform itself also imposes limitations. For instance, platform-specific features, such as sampling strategies, limit the amount of data that can be collected and the behavior and conversation of users depending on the platform or context. Of the 42 studies,1 (2%) study mentioned that the forums they investigated could be very procannabis and are likely inhabited by more experienced cannabis users [41]. Another study stated that individuals posting on YouTube about cannabis are likely to seek social networking opportunities [37].

Complications also arise because platform-specific algorithms spot and further promote popular themes and users to deliberately manage behavior and attract more platform engagement. This needs to be ameliorated by detecting and accounting for the algorithms and potentially by sampling from >1 platform.

Methodology Constraints

The use of small data sets by some of the studies impacts the generalizability of the results, and some of the researchers acknowledged this and indicated a plan to replicate their studies with more data and the use of automated methods. Consequently, we observed that although such studies may be sampling social media data for hypothesis generation, they do not leverage one of the most important features of social media data, which is the ability to observe the continuous generation of big data to create long-term data-centric insights [73].

Biases that could have been introduced by the choice of theme were also mentioned in the studies. Most researchers have attempted to mitigate this by creating annotation guidelines, having >1 person labeling data, and resolving disagreements.

Actual Use Detection

A limitation mentioned by several studies is that web-based search activities and social media posts containing cannabis-related keywords do not necessarily represent the actual use of cannabis by the poster. Depending on the context and goals of the research (for instance, if the research seeks to study a cannabis-consuming population), advanced text processing techniques are required to establish when personal cannabis use can be inferred. For such studies, establishing its use should be a crucial initial step. However, the detection of

XSL•FO
RenderX

personal use is challenging, especially in the informal, diverse, and specialized language used by niche communities.

Source Identification

Identifying the source of posts (ie, whether they were generated organically by individual users or by organizations or bots) was a commonly mentioned limitation. Content generated by health and commercial organizations, power users, and nonindividual accounts was understood to comprise a considerable amount of social media post volume on the web.

Limitations

This review used 4 literature databases in the search process to allow the maximum coverage of existing publications. However, we cannot be certain that we have covered all relevant publications. The choice of keywords for the literature search could also have impacted capturing all the relevant studies in this domain, for instance, *infodemiology* and *infoveillance* were not in the keywords. Articles included in this study were selected following a systematic approach and underwent a bias assessment for quality; however, biases could not be completely avoided. This study was also limited to English-language articles.

Conclusions

The number of studies in this field has steadily increased over the last few years. Social media conversations are wide ranging and offer opportunities for insights that cannot be obtained through formal information gathering. Researchers have realized the value of social media conversations as a place for users to freely express their experiences and concerns without risking judgment or penalty and that social media is the natural forum for many users of cannabis as medicine to share their insights into the benefits and issues they experience and perceive.

Manual qualitative analysis, statistical analysis, supervised and unsupervised machine learning, and rule-based methods are among methodologies used in these studies. Analyses of social media data that are limited to small data samples, although providing an effective means of hypothesis generation, are difficult to reliably reproduce and generalize. Where possible, the sharing of high-quality deidentified annotated data to allow the use of generalizable analytical techniques should be encouraged to advance this field.

To improve their validity and generalizability, studies could add additional social media data sources and check their results against established reporting systems. Studies could take advantage of emerging data analysis strategies that leverage big data, such as deep learning and transfer-learning-based approaches.

Acknowledgments

This review was supported by the Australian Centre for Cannabinoid Clinical and Research Excellence, funded by the National Health and Medical Research Council through the Centre of Research Excellence scheme (NHMRC CRE APP1135054).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supporting information (review keywords, inclusion and exclusion criteria, papers summary).
[OCX File , 45 KB - jmir_v24i11e35974_fig.ocx]

References

1. Li HL. An archaeological and historical account of cannabis in China. Econ Bot 1973 Oct;28(4):437-448 [FREE Full text] [doi: 10.1007/BF02862859]
2. Hallinan CM, Gunn JM, Bonomo YA. Implementation of medicinal cannabis in Australia: innovation or upheaval? Perspectives from physicians as key informants, a qualitative analysis. BMJ Open 2021 Oct 22;11(10):e054044 [FREE Full text] [doi: 10.1136/bmjopen-2021-054044] [Medline: 34686558]
3. Banerjee S, McCormack S. Medical Cannabis for the Treatment of Chronic Pain: A Review of Clinical Effectiveness and Guidelines. Ottawa, Canada: Canadian Agency for Drugs and Technologies in Health; 2019.
4. Kleeman-Forsthuber LT, Dennis DA, Jennings JM. Medicinal cannabis in orthopaedic practice. J Am Acad Orthop Surg 2020 May 01;28(7):268-277. [doi: 10.5435/JAAOS-D-19-00438] [Medline: 31800438]
5. Pawliuk C, Chau B, Rassekh SR, McKellar T, Siden HH. Efficacy and safety of paediatric medicinal cannabis use: a scoping review. Paediatr Child Health 2021 Jul;26(4):228-233 [FREE Full text] [doi: 10.1093/pch/pxaa031] [Medline: 34131459]
6. Pratt M, Stevens A, Thuku M, Butler C, Skidmore B, Wieland LS, et al. Benefits and harms of medical cannabis: a scoping review of systematic reviews. Syst Rev 2019 Dec 10;8(1):320 [FREE Full text] [doi: 10.1186/s13643-019-1243-x] [Medline: 31823819]
7. Martin JH, Lucas C. Reporting adverse drug events to the Therapeutic Goods Administration. Aust Prescr 2021 Mar;44(1):2-3 [FREE Full text] [doi: 10.18773/austprescr.2020.077] [Medline: 33664539]

XSL•FO
RenderX

8. EudraVigilance. European MedicinesAgency. URL: https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance [accessed 2022-05-24]
9. FDA Adverse Event Reporting System (FAERS) Public Dashboard. U.S. Food & Drug Administration. 2021. URL: https://tinyurl.com/yh22mc2c [accessed 2022-05-24]
10. Al Dweik R, Stacey D, Kohen D, Yaya S. Factors affecting patient reporting of adverse drug reactions: a systematic review. Br J Clin Pharmacol 2017 Apr;83(4):875-883 [FREE Full text] [doi: 10.1111/bcp.13159] [Medline: 27868226]
11. Chandra S, Lata H, ElSohly MA, Walker LA, Potter D. Cannabis cultivation: methodological issues for obtaining medical-grade product. Epilepsy Behav 2017 May;70(Pt B):302-312. [doi: 10.1016/j.yebeh.2016.11.029] [Medline: 28202406]
12. Hakkarainen P, Frank VA, Barratt MJ, Dahl HV, Decorte T, Karjalainen K, et al. Growing medicine: small-scale cannabis cultivation for medical purposes in six different countries. Int J Drug Policy 2015 Mar;26(3):250-256. [doi: 10.1016/j.drugpo.2014.07.005] [Medline: 25123694]
13. Paul MJ, Dredze M. Social monitoring for public health. In: Marchionini G, editor. Synthesis Lectures on Information Concepts, Retrieval, and Services. San Rafael, CA, USA: Morgan and Claypool Publishers; Aug 31, 2017:1-183.
14. Bode L, Davis-Kean P, Singh L, Berger-Wolf T, Budak C, Chi G, et al. Study designs for quantitative social science research using social media. PsyArXiv 2020:1-27. [doi: 10.31234/osf.io/zp8q2]
15. Mneimneh Z, Pasek J, Singh L, Best R, Bode L, Bruch E, et al. Data acquisition, sampling, and data preparation considerations for quantitative social science research using social media data. PsyArXiv 2021 Mar 15:1-45. [doi: 10.31234/osf.io/k6vyj]
16. Conway M, Hu M, Chapman WW. Recent advances in using natural language processing to address public health research questions using social media and consumergenerated data. Yearb Med Inform 2019 Aug;28(1):208-217 [FREE Full text] [doi: 10.1055/s-0039-1677918] [Medline: 31419834]
17. Allem JP, Ferrara E. Could social bots pose a threat to public health? Am J Public Health 2018 Aug;108(8):1005-1006. [doi: 10.2105/AJPH.2018.304512] [Medline: 29995482]
18. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun ACM 2016 Jun 24;59(7):96-104. [doi: 10.1145/2818717]
19. Varol O, Ferrara E, Davis CA, Menczer F, Flammini A. Online human-bot interactions: detection, estimation, and characterization. arXiv 2017.
20. Khademi Habibabadi S, Bonomo YA, Conway M, Hallinan CM. Social media discourse and internet search queries on cannabis as a medicine: A systematic scoping review. medRxiv 2022. [doi: 10.1101/2022.05.16.22275171]
21. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
22. McGregor F, Somner JE, Bourne RR, Munn-Giddings C, Shah P, Cross V. Social media use by patients with glaucoma: what can we learn? Ophthalmic Physiol Opt 2014 Jan;34(1):46-52. [doi: 10.1111/opo.12093] [Medline: 24325434]
23. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. J Adolesc Health 2015 Mar;56(2):139-145 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
24. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. "Time for dabs": analyzing Twitter data on marijuana concentrates across the U.S. Drug Alcohol Depend 2015 Oct 01;155:307-311 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.07.1199] [Medline: 26338481]
25. Gonzalez-Estrada A, Cuervo-Pardo L, Ghosh B, Smith M, Pazheri F, Zell K, et al. Popular on YouTube: a critical appraisal of the educational quality of information regarding asthma. Allergy Asthma Proc 2015;36(6):e121-e126. [doi: 10.2500/aap.2015.36.3890] [Medline: 26534743]
26. Krauss MJ, Sowles SJ, Mylvaganam S, Zewdie K, Bierut LJ, Cavazos-Rehg PA. Displays of dabbing marijuana extracts on YouTube. Drug Alcohol Depend 2015 Oct 01;155:45-51 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.08.020] [Medline: 26347408]
27. Thompson L, Rivara FP, Whitehill JM. Prevalence of marijuana-related traffic on Twitter, 2012-2013: a content analysis. Cyberpsychol Behav Soc Netw 2015 Jul;18(6):311-319 [FREE Full text] [doi: 10.1089/cyber.2014.0620] [Medline: 26075917]
28. Cavazos-Rehg PA, Sowles SJ, Krauss MJ, Agbonavbare V, Grucza R, Bierut L. A content analysis of tweets about high-potency marijuana. Drug Alcohol Depend 2016 Oct 01;166:100-108 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.06.034] [Medline: 27402550]
29. Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Martins SS, Boyer EW, et al. "Those edibles hit hard": exploration of Twitter data on cannabis edibles in the U.S. Drug Alcohol Depend 2016 Jul 01;164:64-70 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.04.029] [Medline: 27185160]
30. Mitchell JT, Sweitzer MM, Tunno AM, Kollins SH, McClernon FJ. "I Use Weed for My ADHD": a qualitative analysis of online forum discussions on cannabis use and ADHD. PLoS One 2016 May 26;11(5):e0156614 [FREE Full text] [doi: 10.1371/journal.pone.0156614] [Medline: 27227537]

XSL•FO
RenderX

31. Andersson M, Persson M, Kjellgren A. Psychoactive substances as a last resort-a qualitative study of self-treatment of migraine and cluster headaches. Harm Reduct J 2017 Sep 05;14(1):60 [FREE Full text] [doi: 10.1186/s12954-017-0186-6] [Medline: 28870224]
32. Dai H, Hao J. Mining social media data on marijuana use for Post Traumatic Stress Disorder. Comput Human Behav 2017 May;70(C):282-290. [doi: 10.1016/j.chb.2016.12.064]
33. Greiner C, Chatton A, Khazaal Y. Online self-help forums on cannabis: a content assessment. Patient Educ Couns 2017 Oct;100(10):1943-1950. [doi: 10.1016/j.pec.2017.06.001] [Medline: 28602568]
34. Turner J, Kantardzic M. Geo-social analytics based on spatio-temporal dynamics of marijuana-related tweets. In: Proceedings of the 2017 International Conference on Information System and Data Mining. 2017 Presented at: ICISDM '17; April 1-3, 2017; Charleston, SC, USA p. 28-38. [doi: 10.1145/3077584.3077588]
35. Westmaas JL, McDonald BR, Portier KM. Topic modeling of smoking- and cessation-related posts to the American Cancer Society's Cancer Survivor Network (CSN): implications for cessation treatment for cancer survivors who smoke. Nicotine Tob Res 2017 Aug 01;19(8):952-959. [doi: 10.1093/ntr/ntx064] [Medline: 28340059]
36. Yom-Tov E, Lev-Ran S. Adverse reactions associated with cannabis consumption as evident from search engine queries. JMIR Public Health Surveill 2017 Oct 26;3(4):e77 [FREE Full text] [doi: 10.2196/publichealth.8391] [Medline: 29074469]
37. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Murphy GM, Bierut LJ. Exposure to and content of marijuana product reviews. Prev Sci 2018 Feb;19(2):127-137 [FREE Full text] [doi: 10.1007/s11121-017-0818-9] [Medline: 28681195]
38. Glowacki EM, Glowacki JB, Wilcox GB. A text-mining analysis of the public's reactions to the opioid crisis. Subst Abus 2018;39(2):129-133. [doi: 10.1080/08897077.2017.1356795] [Medline: 28723265]
39. Meacham MC, Paul MJ, Ramo DE. Understanding emerging forms of cannabis use through an online cannabis community: an analysis of relative post volume and subjective highness ratings. Drug Alcohol Depend 2018 Jul 01;188:364-369 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.03.041] [Medline: 29883950]
40. Leas EC, Nobles AL, Caputi TL, Dredze M, Smith DM, Ayers JW. Trends in Internet searches for cannabidiol (CBD) in the United States. JAMA Netw Open 2019 Oct 02;2(10):e1913853 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.13853] [Medline: 31642924]
41. Meacham MC, Roh S, Chang JS, Ramo DE. Frequently asked questions about dabbing concentrates in online cannabis community discussion forums. Int J Drug Policy 2019 Dec;74:11-17 [FREE Full text] [doi: 10.1016/j.drugpo.2019.07.036] [Medline: 31400582]
42. Nasralah T, El-gayar O, Wang Y. What social media can tell us about opioid addicts: Twitter data case analysis. In: Proceedings of the 25th Americas' Conference on Information Systems. 2019 Presented at: AMCIS '19; August 15-17, 2019; Cancún, Mexico p. 15.
43. Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using Twitter to understand the human bowel disease community: exploratory analysis of key topics. J Med Internet Res 2019 Aug 15;21(8):e12610 [FREE Full text] [doi: 10.2196/12610] [Medline: 31411142]
44. Shi S, Brant A, Sabolch A, Pollom E. False news of a cannabis cancer cure. Cureus 2019 Jan 19;11(1):e3918 [FREE Full text] [doi: 10.7759/cureus.3918] [Medline: 30931189]
45. Allem JP, Escobedo P, Dharmapuri L. Cannabis surveillance with Twitter data: emerging topics and social bots. Am J Public Health 2020 Mar;110(3):357-362. [doi: 10.2105/AJPH.2019.305461] [Medline: 31855475]
46. 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]
47. Jia JS, Mehran N, Purgert R, Zhang QE, Lee D, Myers JS, et al. Marijuana and glaucoma: a social media content analysis. Ophthalmol Glaucoma 2021;4(4):400-404. [doi: 10.1016/j.ogla.2020.11.004] [Medline: 33242684]
48. Leas EC, Hendrickson EM, Nobles AL, Todd R, Smith DM, Dredze M, et al. Self-reported cannabidiol (CBD) use for conditions with proven therapies. JAMA Netw Open 2020 Oct 01;3(10):e2020977 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.20977] [Medline: 33057645]
49. Merten JW, Gordon BT, King JL, Pappas C. Cannabidiol (CBD): perspectives from pinterest. Subst Use Misuse 2020;55(13):2213-2220. [doi: 10.1080/10826084.2020.1797808] [Medline: 32715862]
50. Mullins CF, Ffrench-O'Carroll R, Lane J, O'Connor T. Sharing the pain: an observational analysis of Twitter and pain in Ireland. Reg Anesth Pain Med 2020 Aug;45(8):597-602. [doi: 10.1136/rapm-2020-101547] [Medline: 32503862]
51. Saposnik FE, Huber JF. Trends in Web searches about the causes and treatments of autism over the past 15 years: exploratory infodemiology study. JMIR Pediatr Parent 2020 Dec 07;3(2):e20913 [FREE Full text] [doi: 10.2196/20913] [Medline: 33284128]
52. Song S, Cohen AJ, Lui H, Mmonu NA, Brody H, Patino G, et al. Use of GoFundMe ® to crowdfund complementary and alternative medicine treatments for cancer. J Cancer Res Clin Oncol 2020 Jul;146(7):1857-1865. [doi: 10.1007/s00432-020-03191-0] [Medline: 32219517]
53. Tran T, Kavuluru R. Social media surveillance for perceived therapeutic effects of cannabidiol (CBD) products. Int J Drug Policy 2020 Mar;77:102688 [FREE Full text] [doi: 10.1016/j.drugpo.2020.102688] [Medline: 32092666]

XSL•FO
RenderX

54. van Draanen J, Tao H, Gupta S, Liu S. Geographic differences in cannabis conversations on Twitter: infodemiology study. JMIR Public Health Surveill 2020 Oct 05;6(4):e18540 [FREE Full text] [doi: 10.2196/18540] [Medline: 33016888]
55. Zenone M, Snyder J, Caulfield T. Crowdfunding cannabidiol (CBD) for cancer: hype and misinformation on GoFundMe. Am J Public Health 2020 Oct;110(S3):S294-S299. [doi: 10.2105/AJPH.2020.305768] [Medline: 33001729]
56. Pang RD, Dormanesh A, Hoang Y, Chu M, Allem JP. Twitter posts about cannabis use during pregnancy and postpartum:a content analysis. Subst Use Misuse 2021;56(7):1074-1077. [doi: 10.1080/10826084.2021.1906277] [Medline: 33821757]
57. 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]
58. Smolev ET, Rolf L, Zhu E, Buday SK, Brody M, Brogan DM, et al. "Pill pushers and CBD oil"-a thematic analysis of social media interactions about pain after traumatic brachial plexus injury. J Hand Surg Glob Online 2021 Jan;3(1):36-40 [FREE Full text] [doi: 10.1016/j.jhsg.2020.10.005] [Medline: 33537664]
59. Soleymanpour M, Saderholm S, Kavuluru R. Therapeutic claims in cannabidiol (CBD) marketing messages on Twitter. Proceedings (IEEE Int Conf Bioinformatics Biomed) 2021 Dec;2021:3083-3088 [FREE Full text] [doi: 10.1109/bibm52615.2021.9669404] [Medline: 35096472]
60. Zenone MA, Snyder J, Crooks VA. What are the informational pathways that shape people's use of cannabidiol for medical purposes? J Cannabis Res 2021 May 06;3(1):13 [FREE Full text] [doi: 10.1186/s42238-021-00069-x] [Medline: 33957993]
61. Allem JP, Majmundar A, Dormanesh A, Donaldson SI. Identifying health-related discussions of cannabis use on Twitter by using a medical dictionary: content analysis of tweets. JMIR Form Res 2022 Mar 25;6(2):e35027 [FREE Full text] [doi: 10.2196/35027] [Medline: 35212637]
62. 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]
63. Meacham MC, Nobles AL, Tompkins DA, Thrul J. "I got a bunch of weed to help me through the withdrawals": naturalistic cannabis use reported in online opioid and opioid recovery community discussion forums. PLoS One 2022 Feb 8;17(2):e0263583 [FREE Full text] [doi: 10.1371/journal.pone.0263583] [Medline: 35134074]
64. Boyd RL. WELP: Word Embedding Likeness Perusal (v1.03). Ryan Boyd. 2018. URL: https://github.com/ryanboyd/WELP [accessed 2022-10-30]
65. Adams N, Artigiani EE, Wish ED. Choosing your platform for social media drug research and improving your keyword filter list. J Drug Issues 2019 Mar 13;49(3):477-492. [doi: 10.1177/0022042619833911]
66. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text] [doi: 10.1162/jmlr.2003.3.4-5.993]
67. Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In: Aggarwal CC, Zhai C, editors. Mining Text Data. New York, NY, USA: Springer; 2012:415-463.
68. Hutto CJ, 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 Presented at: AAAI '14; June 1-4, 2014; Ann Arbor, MI, USA.
69. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. "When 'bad' is 'good'": identifying personal communication and sentiment in drug-related tweets. JMIR Public Health Surveill 2016 Oct 24;2(2):e162 [FREE Full text] [doi: 10.2196/publichealth.6327] [Medline: 27777215]
70. Grady C. Institutional review boards: purpose and challenges. Chest 2015 Dec;148(5):1148-1155 [FREE Full text] [doi: 10.1378/chest.15-0706] [Medline: 26042632]
71. CrowdTangle | Content Discovery and Social Monitoring Made Easy. URL: https://www.crowdtangle.com/ [accessed 2022-05-30]
72. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. Yearb Med Inform 2017 Aug;26(1):214-227 [FREE Full text] [doi: 10.15265/IY-2017-029] [Medline: 29063568]
73. Sarker A, DeRoos A, Perrone J. Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework. J Am Med Inform Assoc 2020 Feb 01;27(2):315-329 [FREE Full text] [doi: 10.1093/jamia/ocz162] [Medline: 31584645]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
VADER: Valence Aware Dictionary and Sentiment Reasoner

XSL•FO
RenderX

Edited by G Eysenbach; submitted 24.12.21; peer-reviewed by A Dormanesh, K O'Connor, J Thrul; comments to author 26.03.22; revised version received 16.06.22; accepted 27.07.22; published 16.11.22.

Please cite as:
Khademi Habibabadi S, Hallinan C, Bonomo Y, Conway M
Consumer-Generated Discourse on Cannabis as a Medicine: Scoping Review of Techniques
J Med Internet Res 2022;24(11):e35974
URL: https://www.jmir.org/2022/11/e35974
doi:10.2196/35974
PMID:36383417

©Sedigheh Khademi Habibabadi, Christine Hallinan, Yvonne Bonomo, Mike Conway. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Automated Digital Interventions and Smoking Cessation: Systematic Review and Meta-analysis Relating Efficiency to a Psychological Theory of Intervention Perspective

Leihao Sha[1], MD; Xia Yang[1], MM; Renhao Deng[1], MM; Wen Wang[2], PhD; YuJie Tao[1], MD; HaiLing Cao[1], MD; Qianshu Ma[1], MD; Hao Wang[1], MD; Yirou Nie[1], MM; Siqi Leng[1], MM; Qiuyue Lv[1], MD; Xiaojing Li[1], MD; Huiyao Wang[1], MD; Yajing Meng[1], MD; Jiajun Xu[1], MD; Andrew J Greenshaw[3], PhD; Tao Li[1,4], PhD; Wan-jun Guo[1], MD

[1]Mental Health Center and Sichuan Clinical Medical Research Center for Mental Disorders, West China Hospital of Sichuan University, Chengdu, China

[2]Chinese Evidence-Based Medicine Center, West China Hospital of Sichuan University, Chengdu, China

[3]Department of Psychiatry, University of Alberta, Edmonton, AB, Canada

[4]Mental Health Center and Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine, Hangzhou, China

Corresponding Author:
Wan-jun Guo, MD
Mental Health Center and Sichuan Clinical Medical Research Center for Mental Disorders
West China Hospital of Sichuan University
No. 28 Dianxin South St
Chengdu, 610041
China
Phone: 86 28 8542356
Email: guowjcn@163.com

Abstract

Background: Smoking remains a highly significant preventable global public health problem. In this context, digital interventions offer great advantages in terms of a lack of biological side effects, possibility of automatic delivery, and consequent human resource savings relative to traditional interventions. Such interventions have been studied in randomized controlled trials (RCTs) but have not been systematically reviewed with the inclusion of text-based and multiplatform-based interventions. In addition, this area has not been evaluated from the perspective of the psychological theoretical basis of intervention.

Objective: The aim of this paper is to assess the efficiency of digital interventions in RCT studies of smoking cessation and to evaluate the effectiveness of the strategies used for digital interventions.

Methods: An electronic search of RCTs was conducted using PubMed, Embase, and the Cochrane Library by June 30, 2021. Eligible studies had to compare automated digital intervention (ADI) to the use of a self-help guideline or no intervention. Participants were current smokers (aged 16 years or older). As the main outcome, abstinence after endpoint was extracted from the studies. Systematic review and meta-analysis were conducted to assess the efficiency of ADIs. Metaregressions were conducted to assess the relationship between intervention theory and effectiveness.

Results: A total of 19 trials (15,472 participants) were included in the analysis. The overall abstinence rate (95% CI) at the endpoint was 17.8% (17.0-18.7). The overall risk ratio of the intervention group compared to the controls at the endpoint was 17.8% (17.0-18.7). Cochrane risk-of-bias tool for randomized trials (ROB 2) suggested that most of the studies had a low risk of bias (56.3%). Psychological theory–related constructs or predictors, which refer to other theory-based concepts (rather than only behavioral theory) such as craving or anxiety, are associated with effectiveness.

Conclusions: This study found that ADI had a clear positive effect compared to self-help guidelines or to no intervention, and effectiveness was associated with theory-related constructs or predictors. ADIs should be promoted by policy makers and clinical practitioners to address the huge gap between the need for smoking cessation and availability of traditional treatment resources. Possible increases in ADI efficiency may be achieved by optimally integrating psychotherapeutic theories and techniques.

Trial Registration: PROSPERO CRD42021256593; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=256593

(J Med Internet Res 2022;24(11):e38206) doi:10.2196/38206

XSL•FO
RenderX

KEYWORDS

smoking cessation; automated; digital intervention; psychological theory; meta-analysis; systematic review; public health; side effects; interventions; randomized controlled trial; self-help

Introduction

Background

Smoking tobacco is the leading risk factor for noncommunicable diseases and the leading cause of substance-attributable mortality rates and disability-adjusted life years [1,2].

Most smokers use tobacco products constantly or relapse after quitting due to nicotine addiction. Therapies for smoking cessation include pharmacological agents and psychosocial approaches [3]. Pharmacotherapy is recommended for short-term use [4]. Food and Drug Administration–approved first-line medications include nicotine replacement therapy (NRT), bupropion, and varenicline [5], which generally result in higher quit rates than placebo [6]. Side effects of these medications, including nausea, vomiting, and neuropsychiatric symptoms, often limit use of such medication treatment for many general smokers [7,8]. Counseling and behavioral therapies are also effective in smoking cessation [9,10]. Compared to pharmacotherapy, counseling and behavioral therapies may enhance patients' motivation and provide education on general and specific strategies for smoking cessation and encouragement [5]. The shortage of trained counselors remains a barrier not only for general availability but also effectiveness of nonmedication interventions. Current inadequacies in availability of professional human resources, inequities in primary care, overwhelming serious cases in hospitals, unsatisfactory accessibility, poor cost-effectiveness issues, and lack of compliance restrict the efficiency of smoking cessation interventions and result in numerous untreated or relapsed smokers [3,11]. Such evidence indicates a pressing need for more cost-effective interventions.

With the progress of mobile and digital technologies, mobile health management and digital therapeutics present good prospects for managing chronic health conditions [12-14], especially in low- and middle-income countries (LMIC) [15]; this is because they have the advantages of having fewer side effects than pharmacotherapy, less need for trained human resources than counseling and behavioral therapies because of the possibility of automatic delivery, more individualized interventions, as well as great accessibility and portability, which may result in higher cost-effectiveness [11,12,16,17]. Such technologies are, therefore, potentially effective strategies for improving health delivery, especially in LMIC, where the lack of capacity in professional human and related resources is prominent [10]. Recent studies report positive indications that automated digital interventions (ADIs) are well accepted and may benefit smokers in LMICs [13,14], although the number of studies focusing on effectiveness was fewer than those investigating acceptability and feasibility of ADIs in LMICs [15,18,19]. For intervention content, earlier digital interventions provided text interventions (text-based), while multiplatform-based interventions, which provided diverse tools for interventions such as serious games and virtual reality, have been developed and studied more recently. Evidence indicates that mobile health management and digital interventions could be effective in many chronic conditions, including hypertension [20,21], diabetes [22,23], and mental illness [24,25]. Digital interventions may also be promising treatments for substance use disorders other than tobacco smoking, in relation to narrowing the enormous gap between the growing need and lack of professional human resources [26-28]. More than 30 clinical trials examining effectiveness of digital smoking cessation interventions have been conducted. A systemic review focused on text-based interventions has documented promising results [29] but did not provide analysis for ADIs. Neither systematic reviews nor meta-analyses are available that evaluate text-based and multiplatform-based interventions, although the latter are anticipated to be more acceptable to patients.

Theoretical Basis of Interventions

Effectiveness of digital interventions could be influenced by many aspects including psychological behavioral theories, which are often used to develop content of messages used for a respective intervention [30], environment [31], and strategies of implementation [32]. Many retrospective studies found that applying theories more appropriately could improve effectiveness through the implementation of theory-based interventions [33-35]. Interventions guided by different theories could result in different effectiveness. For example, regarding health-related behavior, it is reported that interventions based on the theory of planned behavior tended to have substantially greater effects than other theories such as the transtheoretical model or the elaboration likelihood model [36]. The theory coding scheme (TCS) is a tool for describing the theoretical basis of interventions [37] and is widely used in meta-analyses [31,36,38] for evaluating theory use. The TCS has different items and categories describing relevant theoretical constructs. The importance of the association between theoretical basis and intervention effectiveness for smoking cessation has not yet been assessed, although some reviews that focused on other disorders or conditions demonstrated diverse evidence for the relationship of intervention effectiveness to intervention theory with both positive [36,38] and null [30,39] associations.

Objective

To enhance smoking cessation interventions with ADIs, this systemic review plus meta-analysis aimed to assess the efficiency of automated digital interventions in randomized controlled trial (RCT) studies and to evaluate the association between intervention effectiveness and how the intervention strategy is based on theory.

Methods

Data Sources and Search

Through a broad search of databases, including PubMed, Embase, CNKI, and Cochrane Library, we identified relevant studies using the PRISMA (Preferred Reporting Items for

XSL•FO
RenderX

Systematic Reviews and Meta-Analyses) guidelines [40]. We also checked the reference lists of the included studies and relevant reviews for study selection. The search strategy combined terms related to digital intervention (ie, mobile health) and tobacco smoking cessation (ie, nicotine addiction). For the full search strategy, see Supporting Information 1 in Multimedia Appendix 1.

Inclusion and Exclusion Criteria and Quality Assessment

We included studies with the following criteria: (1) RCT; (2) study with participants aged >16 years and current tobacco smokers; (3) intervention automatedly delivered via a digital method and targeted smoking cessation; (4) abstinence assessment during the whole follow-up period of at least 3 months; (5) self-help guidelines or no intervention in control group; and (6) studies reported in English or Chinese. To focus on smoking cessation solely, study participants with other mental diseases are excluded from this analysis. Moreover, interventions that include financial incentives (which could limit the generalization of ADIs) and studies with rate of loss to follow-up over 60% are excluded from the analysis. No exclusion was made for duration of intervention, intervention frequency, study region study sample, or the content of the delivery or frequency of the messages.

Two authors independently reviewed the studies using the inclusion and exclusion criteria. The title and abstract of each study were screened initially, followed by full-text analysis if the title and abstract were consistent with the inclusion and exclusion criteria. Disagreement on evaluation for inclusion was resolved by discussion of the authors, and if necessary, a third reviewer was included in the discussion to reach consensus.

For quality assessment, we used the Cochrane risk of bias tool (Cochrane ROB 2) [41], which is a commonly used tool to access bias for clinical trials. The randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result were assessed to obtain an estimate of overall bias.

Data Extraction and Effect Size Calculation

We extracted follow-up data at the endpoint (abstinence during the whole follow-up period) as the main outcome for efficiency assessment and follow-up data at 3 months and 6 months (if available) as secondary outcomes. We also extracted author, publication year, study region, sample size, control type, intervention type, and intervention duration of all included articles as baseline information (Table 1). As key components of digital intervention, the timing was consistent across studies; all studies began their intervention around the quit date. Frequency was not reported in most of the studies as their intervention could be available whenever the participant wants or on a daily basis because of automated delivery. Therefore, we did not conduct further analyses in these aspects. We used the risk ratio (RR) as the efficiency measure, and RRs were calculated with follow-up data extracted from the included articles. When cases were reported as lost to follow-up, they were treated as relapse.

We used the TCS to assess the potential relationship between intervention theory and intervention effectiveness [42]. The TCS evaluates several aspects of a theory-based intervention, and these aspects will be analyzed, including “is theory mentioned?” “are the relevant theoretical constructs targeted?” “is theory used to select recipients or tailor interventions?” “are the relevant theoretical constructs measured?” “is theory tested?” and “is theory refined.” Two independent reviewers coded all included articles. In the case of differing opinions, consensus was achieved by discussion. The amended version of TCS was used in this analysis. Two items (“quality of measures” and “randomization of participants to condition”) were excluded from the amended vision of the TCS because these aspects were assessed in Cochrane ROB 2 previously [41]. The amended TCS has a total of 22 items, including all subitems. Each item was coded as 1 (present) or 0 (absent). All intervention theories mentioned in the study or in the reference list were recognized during the review process. The TCS also had 6 categories of theory use, and total scores of each category were calculated for further analysis.

Data Analysis

We conducted all analyses in R (version 4.0.4; R Foundation for Statistical Computing). Q statistic and I^2 were reported for study heterogeneity. If the I^2 was at least 40% and the Q statistic was significant ($P<.05$), the overall effect was considered heterogeneous [37]. We used a random-effects model to analyze the overall effect. We used intention-to-treat analysis to assess all data, which handles those cases lost to follow-up as relapse. To test the accuracy of the overall effects, sensitivity analysis was performed. We also conducted subgroup analysis for 3-month and endpoint abstinence to test the source of heterogeneity using baseline information. Funnel plot and Peters tests were employed to test the publication bias of this analysis because of the sample size [43]. Effect sizes of all studies with 95% CI and the weighted aggregate effects were represented using forest plots. The abstinence rates during follow-up (3 months, 6 months, and endpoint) were calculated and reported with a 95% CI for individual studies and overall.

To evaluate the association between intervention effectiveness and how the intervention strategy is based on theory, we conducted univariate and multivariate metaregressions for 3-month, 6-month, and endpoint abstinence with the TCS score (including each item, each category, and total score). If none of the studies or all of the studies met the standards of the item, no metaregression was conducted for this item. All items or categories that showed a significant association with effectiveness and were coded by more than one study were included in the multivariate regression analysis. The regression coefficient (B) represented the mean of the unstandardized effects that differentially included each TCS covariate. The regression coefficients were calculated and reported with 95% CI and *P* values.

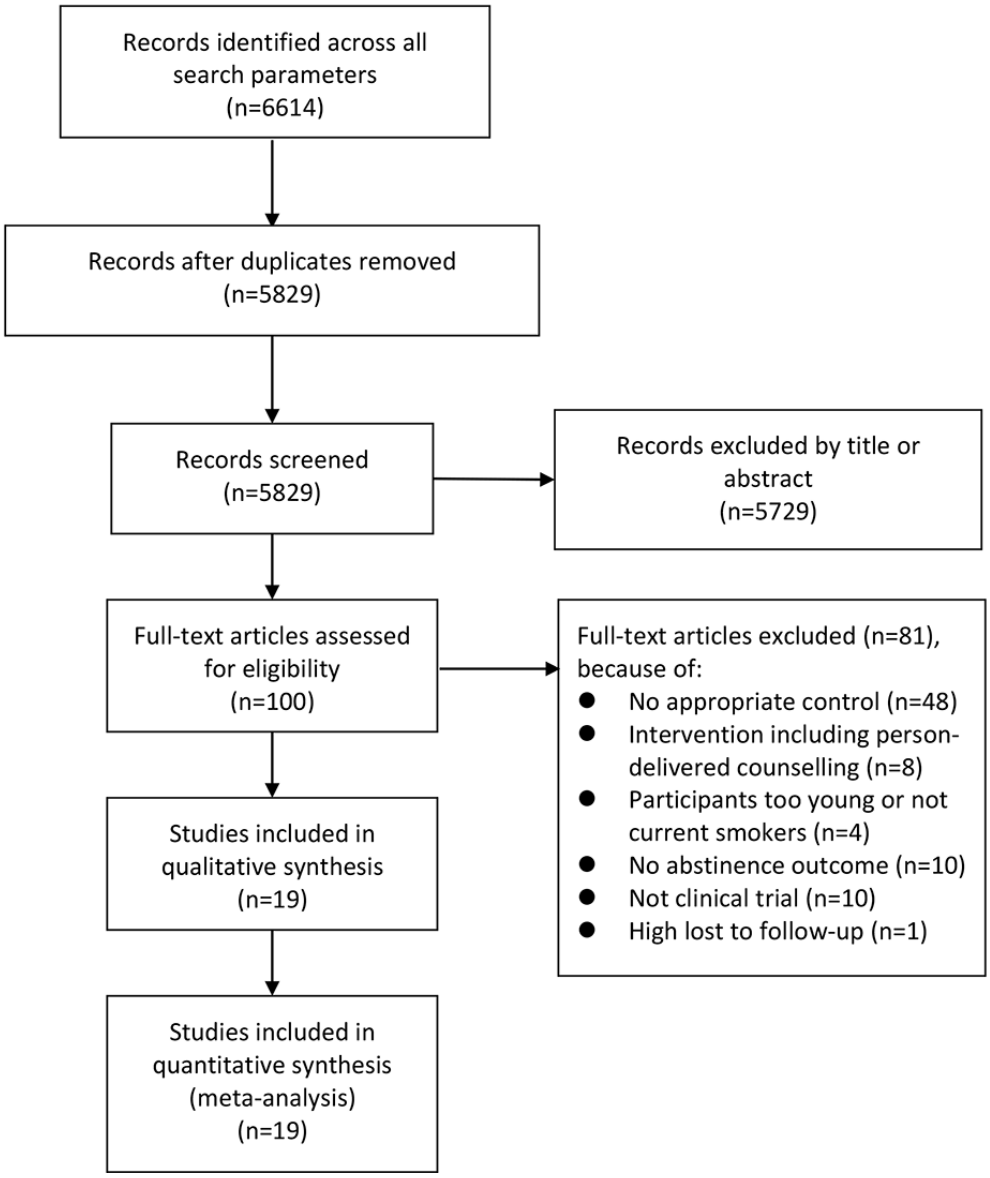
XSL•FO
RenderX

Results

Study Characteristics

A total of 6614 studies were identified after initial database search, and a total of 5829 studies were retained after removing duplicates. After reviewing the full texts, 19 studies were included in the analysis. For the details of study selection, see Figure 1. Table 1 demonstrates basic characteristics for each ADI clinical trial. Out of the 19 trials, 6 (32%) performed no intervention in the control group, and the interventions in the other 13 (68%) trials comprised use of self-help guidelines. Moreover, 8 (42%) trials tested effectiveness for text-based intervention, and the other 11 (58%) trials tested multiplatform-based interventions. Of all the reported theory uses, the transtheoretical model of behavior change (13/19, 68.42%) and cognitive behavioral therapy (12/19, 63.16%) were the most commonly used. The endpoints of included studies ranged from 3 months to 12 months. The sample size varied from 110 to 2478, with 9 trials over 500. Overall abstinence rates (3 months, 6 months, and endpoint) were demonstrated with 95% CIs in Table 1. The results from the Peters test and funnel plots suggested a low risk of publication bias with $P>.05$ (for details, see Supporting Information 2 in Multimedia Appendix 1). Total between-study heterogeneity was significant for 3-month abstinence, 6-month abstinence, and endpoint abstinence (3-month: $X^2=74.04$, $P<.05$ and $I^2=76\%$; 6-month: $X^2=63.40$, $P<.05$ and $I^2=81\%$; endpoint: $X^2=90.10$, $P<.05$ and $I^2=80\%$ endpoint). With respect to quality assessment, Cochrane ROB 2 suggested that, of the 19 included studies, 11 (58%) had a low risk of bias, 3 (16%) showed some concern of bias, and 5 (26%) showed a high risk of bias (Supporting Information 3 in Multimedia Appendix 1).

Figure 1. Flowchart.



XSL•FO
RenderX

Table 1. Descriptive information of included clinical trials.

Author[a]	Year	Study region	Intervention type	Control type	Endpoint	Sample size, n (%)	Abstinence, percentage[b] (95% CI)			
							3-month		Endpoint	
							Experiment	Control	Experiment	Control
Abroms [44]	2014	United States	Text-based	Self-help guideline	6 months	503 (3.25)	27.5 (22.2, 33.3)	16.2 (11.8, 21.5)	11.1 (7.5, 15.5)	5.0 (2.6, 8.5)
Baskerville [45]	2018	Canada	Multiplatform	Self-help guideline	6 months	1599 (10.33)	8.8 (6.9, 10.9)	9.1 (7.2, 11.4)	6.1 (4.6, 8.0)	1.5 (1.0, 2.7)
BinDhim [46]	2018	United States, etc[c]	Multiplatform	Self-help guideline	6 months	684 (4.42)	17.3 (13.4, 21.7)	7.9 (5.3, 11.3)	10.2 (7.2, 13.9)	4.7 (2.7, 7.5)
Brendryen [47]	2008	Norway	Multiplatform	Self-help guideline	12 months	290 (1.87)	30.0 (22.5, 38.0)	11.6 (6.9, 18.0)	20.1 (13.9, 27.6)	6.8 (3.3, 12.2)
Kraft [48]	2008	Norway	Multiplatform	Self-help guideline	12 months	396 (2.56)	44.7 (37.6, 51.9)	28.6 (22.5, 35.5)	37.6 (30.8, 44.7)	24.1 (18.4, 30.7)
Goldenhersch [49]	2020	Argentina	Multiplatform	Self-help guideline	3 months	120 (0.78)	33.3 (21.7, 46.7)	5.0 (1.0, 13.9)	33.3 (21.7, 46.7)	5.0 (1.0, 13.9)
Mavrot [50]	2017	Switzerland	Text-based	Self-help guideline	6 months	1120 (7.24)	20.2 (17.0, 23.8)	17.5 (14.4, 20.9)	17.0 (14.0, 20.3)	15.5 (12.6, 18.8)
Scholten [51]	2019	Netherlands	Multiplatform	Self-help guideline	3 months	144 (0.93)	29.2 (19.0, 41.1)	27.8 (17.9, 39.6)	29.2 (19.0, 41.1)	27.8 (17.9, 39.6)
Whittaker [52]	2011	New Zealand	Multiplatform	Self-help guideline	6 months	226 (1.46)	27.3 (19.2, 36.6)	21.6 (14.5, 30.1)	26.4 (18.4, 35.6)	27.6 (19.7, 36.7)
Rodgers [53]	2005	New Zealand	Text-based	No intervention	6 months	1705 (11.02)	29.0 (26.0, 32.2)	18.8 (16.2, 21.5)	25.4 (22.5, 28.4)	23.7 (20.9, 26.7)
Swartz [54]	2006	United States	Multiplatform	No intervention	3 months	351 (2.27)	12.3 (7.8, 18.2)	5 (2.3, 9.27)	12.3 (7.8, 18.2)	5 (2.3, 9.27)
Liao [55]	2018	China	Text-based	No intervention	6 months	1085 (7.01)	8.3 (6.3, 10.7)	2.2 (1.0, 4.1)	6.8 (5.0, 9.0)	1.9 (0.8, 3.8)
Nguyen [56]	2019	France	Text-based	Self-help guideline	12 months	2478 (16.02)	27.5 (25.1, 30.1)	23.5 (21.1, 25.9)	20.8 (18.6, 23.2)	20.6 (18.3, 22.9)
Bricker [57]	2020	United States	Text-based	Self-help guideline	12 months	2415 (15.61)	14.4 (12.5, 16.5)	7.8 (6.4, 9.5)	9.6 (8.0, 11.3)	5.4 (4.2, 6.8)
Mussenner [58]	2016	Sweden	Text-based	No intervention	3 months	1590 (10.28)	24.5 (21.6, 27.6)	13.8 (11.4, 16.4)	24.5 (21.6, 27.6)	13.8 (11.4, 16.4)
Michele [59]	2012	Turkey	Text-based	Self-help guideline	3 months	151 (0.98)	14.5 (7.5, 24.4)	6.7 (2.2, 14.9)	14.5 (7.5, 24.4)	6.7 (2.2, 14.9)
Mays [60]	2021	United States	Multiplatform	No intervention	6 months	232 (1.49)	38.1 (29.1, 47.7)	11.8 (6.6, 19.0)	51.3 (41.7, 60.8)	27.7 (19.9, 36.7)
Garcia-Pazo [61]	2021	Spain	Multiplatform	Self-help guideline	6 months	110 (0.71)	52.5 (39.3, 65.4)	34.7 (21.7, 49.6)	39.3 (27.1, 52.7)	32.7 (19.9, 47.5)
Chulasai [62]	2022	Thailand	Multiplatform	No intervention	3 months	273 (1.53)	58.4 (49.7, 66.7)	22.1 (15.4, 30.0)	58.4 (49.7, 66.7)	22.1 (15.4, 30.0)
Overall	N/A[d]	N/A	N/A	N/A	N/A	15,472	21.8 (20.9, 22.7)	14.5 (13.7, 15.3)	17.8 (17.0, 18.7)	13.5 (12.7, 14.2)

[a]First author, except for Kraft, who is the second author of the clinical trial. This exception was made because this clinical trial's first author was also Brendryen, though these 2 trials were completely different samples.

[b]Abstinence is calculated when treatment lost to follow up as relapse.

[c]United States, United Kingdom, Australia, and Singapore.

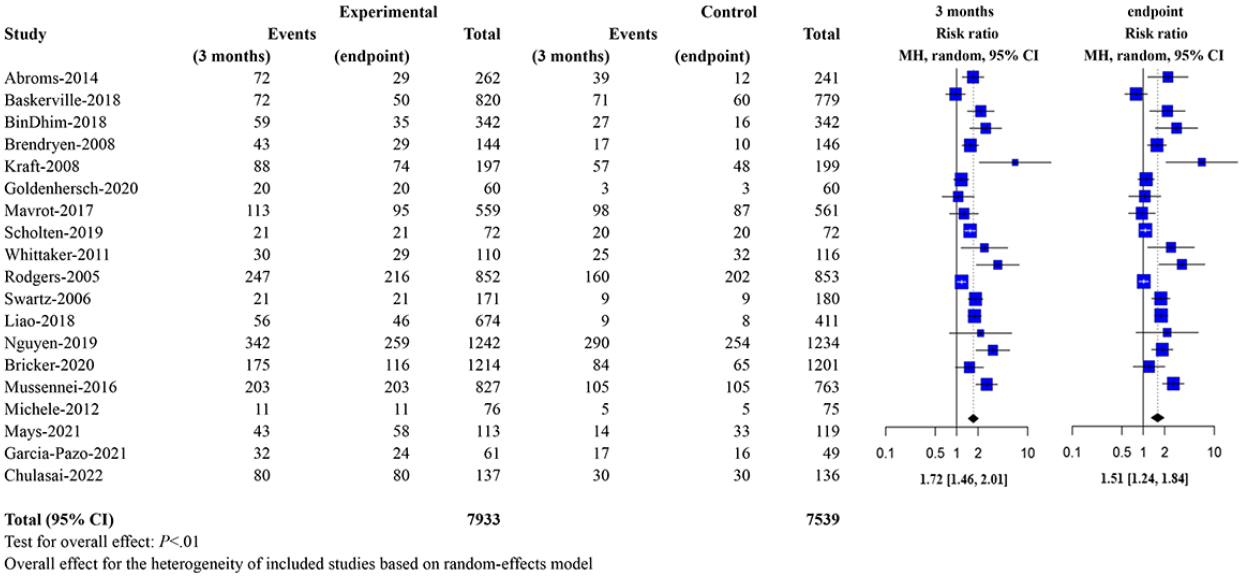
[d]N/A: not applicable.

XSL•FO
RenderX

Abstinence

Figure 2 [44-62] shows a forest plot of 3-month and final abstinence (for a forest plot of 6-month abstinence, see Supporting Information 4 in Multimedia Appendix 1). For final abstinence, the ADI had a moderate effect compared to controls (RR 1.58, 95% CI 1.31, 1.90). Similarly, for 3-month and 6-month abstinence, ADI also showed a moderate effect relative to controls (RR 1.72, 95% CI 1.46, 2.01) and (RR 1.43, 95% CI 1.17, 1.74), respectively. The sensitivity test demonstrated that the overall effect was strong. Omitting any trial would not change the overall effect significantly in the 3-month results, 6-month results, and endpoint results (for sensitivity test forest plots, see Multimedia Appendix 1: Supporting Information 5.1 for 3-month results; Supporting Information 5.2 for 6-month results; and Supporting Information 5.3 for endpoint results).

Figure 2. Forest plot of abstinence at 3 months and final follow-up. MH: Mantel-Haenszel method.



Subgroup Analysis

Subgroup analysis results for 3-month and endpoint abstinence are displayed in Table 2. We divided studies into subgroups by the most commonly used theories (transtheoretical model of behavior change and cognitive behavioral therapy), and there was no evidence that a particular theory or the combination of these theories had a significant effect on intervention efficiency. With respect to sample size, there was no significant difference between small-sample trials (sample size ≤500) and large-sample trials (sample size >500). Similarly, we did not find a significant difference between trials with 2 different controls or 2 different interventions. RR of studies that were compared with no intervention is significantly higher compared with that of self-help guideline. For intervention duration, we did not find any difference between interventions that lasted for more than 1 month and less than 1 month.

XSL•FO
RenderX

Table 2. Subgroup analysis of abstinence during the 3-month and endpoint abstinence.

Subgroup	Sample size	Risk ratio (95% CI) 3-month	Risk ratio (95% CI) Endpoint
Strategies used for digital interventions			
Both	7187	1.59 (1.32, 1.91)	1.43 (1.12, 1.83)
Cognitive behavioral therapy	1547	2.99 (1.62, 5.53)	2.30 (1.28, 4.12)
Transtheoretical model of behavior change	6321	1.48 (1.11, 1.97)	1.43 (1.12, 1.83)
Others	417	1.70 (0.69, 4.21)	1.65 (1.26, 2.15)
P value for subgroup difference	N/A[a]	.23	.50
Sample size			
>500	13,179	1.54 (1.28, 1.86)	1.41 (1.11, 1.78)
≤500	2293	1.98 (1.51, 2.60)	1.78 (1.34, 2.37)
P value for subgroup difference	N/A	.13	.21
Control			
Self-help guideline	10,126	1.51 (1.26, 1.80)	1.42 (1.14, 1.76)
No intervention	5236	2.22 (1.70, 2.91)	1.94 (1.35, 2.78)
P value for subgroup difference	N/A	.02	.14
Intervention type			
Text-based	11,047	1.59 (1.31, 1.92)	1.48 (1.16, 1.89)
Multiplatform-based	4425	1.85 (1.41, 2.44)	1.61 (1.24, 2.24)
P value for subgroup difference	N/A	.36	.54
Intervention duration			
>1 month	12,308	1.70 (1.41, 2.04)	1.59 (1.28, 1.97)
≤1 month	3164	1.89 (1.24, 2.90)	1.65 (1.00, 2.71)
P value for subgroup difference	N/A	.64	.89
Overall	15,472	1.72 (1.46, 2.01)	1.58 (1.31, 1.90)

[a]N/A: not applicable.

Intervention Theories

Tables 3 and 4 report the frequency of trials presenting the TCS items (both items and categories) and the results of univariate regression for abstinence at the endpoint. Most of the items and categories did not show any significant association with intervention efficiency except for I7, I10, and C2. However, only 1 study was coded present in I3 and I10, which was excluded from multivariate metaregression. To avoid collinearity, we did not run multivariate regression with I7 and C2 because I7 is an item inside category 2 (C2) Subsequently, I7 and C2, which pertain to theory-related constructs or predictors, were significantly and independently associated with intervention efficiency. For 3-month and 6-month abstinence, similar results were observed (Supporting Information 6 in Multimedia Appendix 1).

For detailed description of each item and category of the TCS, see Supporting Information 7 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 3. Univariate regression of the theory coding scheme (TCS) items, TCS categories, and total score at endpoint (results of metaregression).

Theory coding scheme item (item number)	Studies where item coded as present, n (%)	Univariate		
		B	95% CI	*P* value
Theory or model of behavior mentioned (I1)	19 (100)	N/A[a]	N/A	N/A
Targeted construct mentioned as predictor of behavior (I2)	15 (89)	0.1505	(–0.5630, 0.8640)	.68
Intervention based on single theory (I3)	2 (11)	0.4335	(–0.2318, 1.0989)	.20
Theory or predictors used to select recipients for the intervention (I4)	1 (5)	0.4489	(–0.6274, 1.5252)	.41
Theory or predictors used to select or develop intervention techniques (I5)	19 (100)	N/A	N/A	N/A
Theory or predictors used to tailor intervention techniques to recipients (I6)	1 (5)	0.4489	(–0.6274, 1.5252)	.41
All intervention techniques are explicitly linked to at least one theory-relevant construct or predictor (I7)	3 (16)	0.6459	(0.0530, 1.2387)	.03[b]
At least one, but not all, of the intervention techniques are explicitly linked to at least one theory-relevant construct or predictor (I8)	18 (95)	0.5413	(–0.3370, 1.4197)	.23
Group of techniques are linked to a group of constructs or predictors (I9)	3 (16)	0.2127	(–0.3329, 0.7583)	.45
All theory-relevant constructs or predictors are explicitly linked to at least one intervention technique (I10)	1 (5)	1.4679	(0.1052, 2.8306)	.03[b]
At least one, but not all, of the theory-relevant constructs or predictors are explicitly linked to at least one intervention technique (I11)	17 (89)	0.5188	(–0.1237, 1.1613)	.11
Theory-relevant constructs are measured: after intervention (I12a)	18 (95)	0.5413	(–0.3370, 1.4197)	.23
Theory-relevant constructs are measured: after and before intervention (I12b)	10 (53)	0.0811	(–0.3403, 0.5025)	.71
Changes in measured theory-relevant constructs or predictors (I13)	9 (47)	0.2884	(–0.1167, 0.6935)	.16
Mediator predicts the dependent variable (I14a)	2 (11)	0.2495	(–0.4229, 0.9218)	.47
Mediator predicts dependent variable, controlling for the independent variable (I14b)	2 (11)	0.2495	(–0.4229, 0.9218)	.47
Intervention does not predict the dependent variable when controlling the independent variable (I14c)	0 (0)	N/A	N/A	N/A
Mediated effect is statistically significant (I14d)	2 (11)	0.2495	(–0.4229, 0.9218)	.47
Results discussed in relation to theory (I15)	8 (42)	0.0104	(–0.4144, 0.4352)	.96
Appropriate support for theory (I16)	2 (11)	0.2495	(–0.4229, 0.9218)	.47
Results used to refine theory: adding or removing constructs to the theory (I17a)	0 (0)	N/A	N/A	N/A
Results used to refine theory: specifying that the interrelationships between the theoretical constructs should be changed (I17b)	0 (0)	N/A	N/A	N/A

[a]N/A: not applicable.

[b]$P<.05$.

XSL•FO
RenderX

Table 4. Univariate regression of the theory coding scheme (TCS) items, TCS categories, and included items.

Theory coding scheme categories (category number)	Items included	Univariate		
		B	95% CI	*P* value
Reference to underpinning theory (C1)	1, 2, 3	0.2683	(–0.1883, 0.7249)	.25
Targeting of relevant theoretical constructs (C2)	2, 5, 6, 7, 8, 9, 10, 11	0.2558	(0.0619, 0.4497)	.01[a]
Using theory to select recipients or tailor interventions (C3)	4, 6	0.2245	(–0.3137, 0.7626)	.41
Measurement of constructs (C4)	12a, 12b	0.1363	(–0.2128, 0.4854)	.44
Testing of theory: mediation effects (C5)	12a, 12b, 13, 14a, 14b, 14c, 14d, 15, 16	0.0524	(–0.0482, 0.1531)	.31
Refining theory (C6)	17a, 17b			
Total use of theory	All items	0.0524	(–0.0068, 0.1116)	.08

[a]*P*<.05.

Discussion

Principal Results

To our knowledge, this is the first systematic review and meta-analysis of trials of the effectiveness of ADIs, including both text-based and multiplatform-based trials, on smoking cessation. This meta-analysis provides the latest and strongest evidence on the overall effectiveness of ADIs for smoking cessation by finding that ADIs had a moderate effect (RR 1.58, 95% CI 1.31, 1.90) on smoking cessation, compared to self-help guidelines or to no intervention.

Comparison With Prior Work

Although plausible evidence from clinical trials that directly compared ADIs and traditional therapies (including pharmacotherapies and psychotherapies) is not available, this moderate effect of ADIs found in the present study is basically comparable with most traditional interventions documented in previous meta-analyses. We extracted the RR with 95% CI from previous meta-analyses, and the RR (95% CI) for endpoint abstinence compared to placebo or no treatment was as follows: RR 2.24, 95% CI (2.06, 2.43) [63]; RR 1.64, 95% CI (1.52, 1.77) for bupropion [64]; and RR 1.60, 95% CI (1.53, 1.68) for NRT [65], among which only varenicline showed a significantly larger RR than ADIs. Regarding the effectiveness of individualized face-to-face psychological interventions on smoking cessation, the most recently updated meta-analysis reported (RR 1.57, 95% CI 1.40, 1.77) as the overall RR (95% CI) of face-to-face psychological interventions compared to minimal interventions (self-help booklet and brief advice) for 6-month abstinence [10].

To compare rates of effectiveness directly, we also calculated the rates (95% CI) of abstinence during the endpoint for the aforementioned meta-analyses. Accordingly, the abstinence rate of the ADIs during the endpoint in this study, which achieved 16.4, 95% CI (15.5, 17.2), was lower than 25.6, 95% CI (24.5, 26.6) of varenicline, comparable to 19.7, 95% CI (8.8, 20.6) and 16.9, 95% CI (16.5, 17.3) for bupropion and NRT, and higher than 10.9, 95% CI (10.1, 11.8) of individualized face-to-face counselling [10,63,64,66]. When compared with pharmacotherapies, the following factors are noteworthy. First, a moderate level of side effects, especially nausea, is reported in all pharmacotherapies, and there is a notable increase in serious adverse effects, including infections and cardiovascular events, with the use of varenicline and NRT [63,65]. Second, an increase in dropouts due to side effects has been reported for bupropion [64]. In addition, participants engaging trials of pharmacotherapies may have more motivation than other interventions because they consented to participate, even though they were made aware in advance of possible side effects. For the comparison with face-to-face psychological interventions, the results of this study are consistent with the results of research on other substance use disorders, as a systematic review demonstrated that there might be no or little difference in effectiveness between digital interventions and face-to-face interventions on lowering alcohol consumption [67].

A systematic review and meta-analysis focused on text-based interventions for smoking cessation was conducted in 2019 [29]. The prior published meta-analysis regarding digital interventions on smoking cessation included both human-delivered interventions and automated-delivered interventions. With the advantages of less professional human resource costs comparing to human-delivered ones [11,12,16,17], the effectiveness of ADIs have not been investigated separately from human-delivered interventions in previous meta-analyses. It documented an RR of 1.54, 95% CI (1.19, 2.00) for final abstinence, which did not significantly differ from that for ADIs found by this study. It is also consistent with the finding of this study that the average RR for final abstinence did not significantly differ from between multiplatform-based (RR 1.48, 95% CI 1.16, 1.89) and text-based (RR 1.61, 95% CI 1.10, 2.35) interventions, but the RR of loss to follow-up of multiplatform-based interventions (RR 0.80, 95% CI 0.59, 1.09) was significantly lower than that of text-based interventions (RR 1.22, 95% CI 1.01, 1.48) at 3 months (*P*=.02), although those during the whole follow-up period (RR 0.87, 95% CI 0.56, 1.36 versus RR 1.18, 95% CI 0.95, 1.47) did not differ significantly (*P*=.06) (Supporting Information 7 in Multimedia Appendix 1). Whether multiplatform-based interventions, such as serious game–based [51] and virtual reality–based [49], may offer advantages in compliance and acceptability associated with the interest of participants needs to be further tested

XSL•FO
RenderX

directly, comparing studies with much larger samples. Mixed results were reported when digital interventions were applied as an adjuvant therapy to traditional interventions. An RCT found that abstinence for intervention group (traditional therapies plus digital interventions) was 2.15 times higher than control group (traditional interventions only) in 12-month follow-up (odds ratio=3.13, 95% CI 1.53, 6.71) [68]; however, 2 other trials documented null results [69,70].

Nonetheless, integrating available evidence with the findings of this study, we believe that ADIs could be an effective treatment for smokers. With our estimated effect size, ADIs could increase the abstinence rate by approximately 50% on average. Globally, if available for 20% of the 1.3 billion adult tobacco users who have tried to quit [71] (although most of them failed due to lack of available help), it would help many smokers achieve cessation.

To our knowledge, this study is also the first to assess the empirical evidence of a potential relationship between the effectiveness of digital interventions on smoking cessation and psychological theory. Even though there is a limited number of included studies (ie, 16), which could result in insensitive and underpowered metaregression [72], we found that TCS item I7, "All theory-relevant constructs or predictors are explicitly linked to at least one intervention technique," was significantly and independently associated with a higher rate of final abstinence. This item pertains to theory-related constructs or predictors, which refer to the theory that digital intervention is not exclusively based on the concept of addictive behaviors but also based on other concepts, such as craving, anxiety, and dependence, which may be more closely linked to the cognitive and affective mechanisms of addiction [3]. Similarly, in a previous meta-analysis on the effectiveness of digital interventions in reducing hazardous alcohol use, the TCS-based metaregression model also found that an item pertaining to theory-related constructs or predictors was significantly associated with better effectiveness [73]. Those findings highlight the importance of theory-related constructs or predictors, especially when developing or optimizing digital interventions for substance use disorders. Based on our findings and previous evidence, studying not only the outcome but also theory-related constructs or predictors offers significant promise for our attempts to unlock what is essentially a black box of mechanisms to understand the theoretical bases and apply that knowledge to improve overall and individualized effectiveness [74-77]. Accordingly, future studies focusing on optimally integrating psychotherapeutic theories and techniques would further increase the efficiency of digital interventions on smoking cessation at both the individualized and overall levels.

Limitations

We also conducted subgroup analysis to explain the moderate heterogeneity of the studies. However, the results suggested that the analyzed variates were not significantly connected to overall effectiveness. A high rate of loss to follow-up in some studies might contribute to the heterogeneity of this analysis [49]; although we conducted intention-to-treat analysis of the included studies, intention-to-treat analysis could underestimate the results, and complete case analysis could potentially benefit. However, insufficient data limit further analysis. Furthermore, the sensitivity analysis also suggested that any single study could not significantly alter the results. Future studies should enlarge the sample size, which could achieve more satisfactory heterogeneity and provide more accurate results for effect size analysis. Other limitations of this study are also noteworthy. First, the quality of evidence could be a concern because not all the original studies included in this meta-analysis used blind interventions for participants due to feasibility; the risk of bias regarding randomization and intervention delivery might be lowered by the automated procedures of interventions. Second, the self-report abstinence of some studies could bring performance bias to this study, although it would affect both the experimental group and the control group and would unlikely change the effect size measured by RR. Third, most of the included studies were conducted in Western countries, and this could limit the generalization of the results to non-Western countries. Studies conducted in Eastern countries are limited, and we believe that Eastern countries could learn from the experience of previous studies to address the problem of tobacco smoking, as the limited number of studies conducted in Eastern countries are also majorly reported helpful results. In addition, socioeconomic status and age could be potential confounding factors for effectiveness as certain studies were conducted on special populations (such as college students). However, insufficient data made it unable to perform subgroup analysis. For intervention theories, the number of included studies limited the power for detecting some associations, including when assessing differences in how theory is applied.

Conclusions

Those limitations notwithstanding, this study indicated that ADIs on the transtheoretical model of behavior change and cognitive behavioral therapy had a clear effect compared to self-help guidelines or to no intervention, and effectiveness is associated with some theory-related constructs or predictors. Accordingly, ADIs should be promoted by policy makers and clinical practitioners to fill the huge gap between the need for smoking cessation and treatment resources and should be studied further to increase efficiency by optimally integrating psychotherapeutic theories and techniques.

Acknowledgments

This study is supported by grants from the National Natural Science Foundation (82171487), the Department of Science of Sichuan provincial government (2019YFS0153), and the 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (2019HXFH026). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supporting information.
[PDF File (Adobe PDF File), 607 KB - jmir_v24i11e38206_app1.pdf]

References

1. Peacock A, Leung J, Larney S, Colledge S, Hickman M, Rehm J, et al. Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. Addiction 2018 Oct;113(10):1905-1926. [doi: 10.1111/add.14234] [Medline: 29749059]
2. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1923-1994 [FREE Full text] [doi: 10.1016/S0140-6736(18)32225-6] [Medline: 30496105]
3. Kalin NH. Substance use disorders and addiction: Mechanisms, trends, and treatment implications. Am J Psychiatry 2020 Nov 01;177(11):1015-1018. [doi: 10.1176/appi.ajp.2020.20091382] [Medline: 33135466]
4. West R, Raw M, McNeill A, Stead L, Aveyard P, Bitton J, et al. Health-care interventions to promote and assist tobacco cessation: a review of efficacy, effectiveness and affordability for use in national guideline development. Addiction 2015 Sep;110(9):1388-1403 [FREE Full text] [doi: 10.1111/add.12998] [Medline: 26031929]
5. Prochaska JJ, Benowitz NL. The past, present, and future of nicotine addiction therapy. Annu Rev Med 2016;67:467-486 [FREE Full text] [doi: 10.1146/annurev-med-111314-033712] [Medline: 26332005]
6. Cahill K, Stevens S, Perera R, Lancaster T. Pharmacological interventions for smoking cessation: an overview and network meta-analysis. Cochrane Database Syst Rev 2013 May 31(5):CD009329 [FREE Full text] [doi: 10.1002/14651858.CD009329.pub2] [Medline: 23728690]
7. Fiore MC, Jaén CR. A clinical blueprint to accelerate the elimination of tobacco use. JAMA 2008 May 07;299(17):2083-2085. [doi: 10.1001/jama.299.17.2083] [Medline: 18460668]
8. Quik M, Zhang D, Perez XA, Bordia T. Role for the nicotinic cholinergic system in movement disorders; therapeutic implications. Pharmacol Ther 2014 Oct;144(1):50-59 [FREE Full text] [doi: 10.1016/j.pharmthera.2014.05.004] [Medline: 24836728]
9. Stead LF, Buitrago D, Preciado N, Sanchez G, Hartmann-Boyce J, Lancaster T. Physician advice for smoking cessation. Cochrane Database Syst Rev 2013 May 31(5):CD000165 [FREE Full text] [doi: 10.1002/14651858.CD000165.pub4] [Medline: 23728631]
10. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3:CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
11. Grant BF, Saha TD, Ruan WJ, Goldstein RB, Chou SP, Jung J, et al. Epidemiology of DSM-5 drug use disorder: Results from the national epidemiologic survey on alcohol and related conditions-III. JAMA Psychiatry 2016 Jan;73(1):39-47 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.2132] [Medline: 26580136]
12. Tighe SA, Ball K, Kensing F, Kayser L, Rawstorn JC, Maddison R. Toward a digital platform for the self-management of noncommunicable disease: Systematic review of platform-like interventions. J Med Internet Res 2020 Oct 28;22(10):e16774 [FREE Full text] [doi: 10.2196/16774] [Medline: 33112239]
13. Choudhry NK, Krumme AA, Ercole PM, Girdish C, Tong AY, Khan NF, et al. Effect of reminder devices on medication adherence: The REMIND randomized clinical trial. JAMA Intern Med 2017 May 01;177(5):624-631 [FREE Full text] [doi: 10.1001/jamainternmed.2016.9627] [Medline: 28241271]
14. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using Mobile Apps for Health Management: A New Health Care Mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
15. Fu Z, Burger H, Arjadi R, Bockting CLH. Effectiveness of digital psychological interventions for mental health problems in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Psychiatry 2020 Oct;7(10):851-864 [FREE Full text] [doi: 10.1016/S2215-0366(20)30256-X] [Medline: 32866459]
16. Carson KV, Verbiest MEA, Crone MR, Brinn MP, Esterman AJ, Assendelft WJJ, et al. Training health professionals in smoking cessation. Cochrane Database Syst Rev 2012 May 16(5):CD000214. [doi: 10.1002/14651858.CD000214.pub2] [Medline: 22592671]
17. Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, et al. DSM-5 criteria for substance use disorders: recommendations and rationale. Am J Psychiatry 2013 Aug;170(8):834-851 [FREE Full text] [doi: 10.1176/appi.ajp.2013.12060782] [Medline: 23903334]
18. Byrne S, Kotze B, Ramos F, Casties A, Harris A. Using a mobile health device to manage severe mental illness in the community: What is the potential and what are the challenges? Aust N Z J Psychiatry 2020 Oct;54(10):964-969. [doi: 10.1177/0004867420945782] [Medline: 32772708]

XSL•FO
RenderX

19. Källander K, Tibenderana JK, Akpogheneta OJ, Strachan DL, Hill Z, ten Asbroek AHA, et al. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review. J Med Internet Res 2013 Jan 25;15(1):e17 [FREE Full text] [doi: 10.2196/jmir.2130] [Medline: 23353680]
20. Mishra SR, Lygidakis C, Neupane D, Gyawali B, Uwizihiwe JP, Virani SS, et al. Combating non-communicable diseases: potentials and challenges for community health workers in a digital age, a narrative review of the literature. Health Policy Plan 2019 Feb 01;34(1):55-66. [doi: 10.1093/heapol/czy099] [Medline: 30668690]
21. Schwalm J, McCready T, Lopez-Jaramillo P, Yusoff K, Attaran A, Lamelas P, et al. A community-based comprehensive intervention to reduce cardiovascular risk in hypertension (HOPE 4): a cluster-randomised controlled trial. Lancet 2019 Oct 05;394(10205):1231-1242. [doi: 10.1016/S0140-6736(19)31949-X] [Medline: 31488369]
22. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
23. Young HM, Miyamoto S, Dharmar M, Tang-Feldman Y. Nurse Coaching and Mobile Health Compared With Usual Care to Improve Diabetes Self-Efficacy for Persons With Type 2 Diabetes: Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e16665 [FREE Full text] [doi: 10.2196/16665] [Medline: 32130184]
24. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. Int J Bipolar Disord 2018 Jan 23;6(1):6. [doi: 10.1186/s40345-017-0110-8] [Medline: 29359252]
25. Xu DR, Xiao S, He H, Caine ED, Gloyd S, Simoni J, et al. Lay health supporters aided by mobile text messaging to improve adherence, symptoms, and functioning among people with schizophrenia in a resource-poor community in rural China (LEAN): A randomized controlled trial. PLoS Med 2019 Apr;16(4):e1002785 [FREE Full text] [doi: 10.1371/journal.pmed.1002785] [Medline: 31013275]
26. Litvin EB, Abrantes AM, Brown RA. Computer and mobile technology-based interventions for substance use disorders: an organizing framework. Addict Behav 2013 Mar;38(3):1747-1756. [doi: 10.1016/j.addbeh.2012.09.003] [Medline: 23254225]
27. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
28. Bennett KM, Clary KL, Smith DC, Lee CA. Usability and Acceptability of a Mobile App to Help Emerging Adults Address their Friends' Substance Use (Harbor): Quantitative Study. J Med Internet Res 2020 Nov 05;22(11):e16632 [FREE Full text] [doi: 10.2196/16632] [Medline: 33151158]
29. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
30. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474. [doi: 10.1037/a0032853] [Medline: 23730717]
31. Jia Y, Fu H, Gao J, Dai J, Zheng P. The roles of health culture and physical environment in workplace health promotion: a two-year prospective intervention study in China. BMC Public Health 2018 Apr 05;18(1):457 [FREE Full text] [doi: 10.1186/s12889-018-5361-5] [Medline: 29621986]
32. Chu AHY, Ng SHX, Tan CS, Win AM, Koh D, Müller-Riemenschneider F. A systematic review and meta-analysis of workplace intervention strategies to reduce sedentary time in white-collar workers. Obes Rev 2016 May;17(5):467-481. [doi: 10.1111/obr.12388] [Medline: 26990220]
33. Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. Soc Sci Med 2010 May;70(10):1618-1625. [doi: 10.1016/j.socscimed.2010.01.039] [Medline: 20207464]
34. Painter JE, Borba CPC, Hynes M, Mays D, Glanz K. The use of theory in health behavior research from 2000 to 2005: a systematic review. Ann Behav Med 2008 Jun;35(3):358-362. [doi: 10.1007/s12160-008-9042-y] [Medline: 18633685]
35. 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]
36. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
37. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
38. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
39. Gardner B, Wardle J, Poston L, Croker H. Changing diet and physical activity to reduce gestational weight gain: a meta-analysis. Obes Rev 2011 Jul;12(7):e602-e620. [doi: 10.1111/j.1467-789X.2011.00884.x] [Medline: 21521451]

XSL•FO
RenderX

40. 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]
41. Higgins JPT, 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:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
42. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
43. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med 2006 Oct 30;25(20):3443-3457. [doi: 10.1002/sim.2380] [Medline: 16345038]
44. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
45. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a Mobile Phone Intervention on Quitting Smoking in a Young Adult Population of Smokers: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893 [FREE Full text] [doi: 10.2196/10893] [Medline: 30355563]
46. BinDhim NF, McGeechan K, Trevena L. Smartphone Smoking Cessation Application (SSC App) trial: a multicountry double-blind automated randomised controlled trial of a smoking cessation decision-aid 'app'. BMJ Open 2018 Jan 21;8(1):e017105 [FREE Full text] [doi: 10.1136/bmjopen-2017-017105] [Medline: 29358418]
47. Brendryen H, Drozd F, Kraft P. A digital smoking cessation program delivered through internet and cell phone without nicotine replacement (happy ending): randomized controlled trial. J Med Internet Res 2008 Nov 28;10(5):e51 [FREE Full text] [doi: 10.2196/jmir.1005] [Medline: 19087949]
48. Brendryen H, Kraft P. Happy ending: a randomized controlled trial of a digital multi-media smoking cessation intervention. Addiction 2008 Mar;103(3):478-84; discussion 485. [doi: 10.1111/j.1360-0443.2007.02119.x] [Medline: 18269367]
49. Goldenhersch E, Thrul J, Ungaretti J, Rosencovich N, Waitman C, Ceberio MR. Virtual Reality Smartphone-Based Intervention for Smoking Cessation: Pilot Randomized Controlled Trial on Initial Clinical Efficacy and Adherence. J Med Internet Res 2020 Jul 29;22(7):e17571 [FREE Full text] [doi: 10.2196/17571] [Medline: 32723722]
50. Mavrot C, Stucki I, Sager F, Etter J. Efficacy of an Internet-based, individually tailored smoking cessation program: A randomized-controlled trial. J Telemed Telecare 2017 Jun;23(5):521-528. [doi: 10.1177/1357633X16655476] [Medline: 27365318]
51. Scholten H, Luijten M, Granic I. A randomized controlled trial to test the effectiveness of a peer-based social mobile game intervention to reduce smoking in youth. Dev Psychopathol 2019 Dec;31(5):1923-1943. [doi: 10.1017/S0954579419001378] [Medline: 31607279]
52. Whittaker R, Dorey E, Bramley D, Bullen C, Denny S, Elley CR, et al. A theory-based video messaging mobile phone intervention for smoking cessation: randomized controlled trial. J Med Internet Res 2011 Jan 21;13(1):e10 [FREE Full text] [doi: 10.2196/jmir.1553] [Medline: 21371991]
53. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin R, et al. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
54. Swartz LHG, Noell JW, Schroeder SW, Ary DV. A randomised control study of a fully automated internet based smoking cessation programme. Tob Control 2006 Feb;15(1):7-12 [FREE Full text] [doi: 10.1136/tc.2003.006189] [Medline: 16436397]
55. Liao Y, Wu Q, Kelly BC, Zhang F, Tang Y, Wang Q, et al. Effectiveness of a text-messaging-based smoking cessation intervention ("Happy Quit") for smoking cessation in China: A randomized controlled trial. PLoS Med 2018 Dec;15(12):e1002713 [FREE Full text] [doi: 10.1371/journal.pmed.1002713] [Medline: 30562352]
56. Nguyen Thanh V, Guignard R, Lancrenon S, Bertrand C, Delva C, Berlin I, et al. Effectiveness of a Fully Automated Internet-Based Smoking Cessation Program: A Randomized Controlled Trial (STAMP). Nicotine Tob Res 2019 Jan 04;21(2):163-172. [doi: 10.1093/ntr/nty016] [Medline: 29370407]
57. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of Smartphone Applications for Smoking Cessation: A Randomized Clinical Trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
58. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of Short Message Service Text-Based Smoking Cessation Intervention Among University Students: A Randomized Clinical Trial. JAMA Intern Med 2016 Mar;176(3):321-328 [FREE Full text] [doi: 10.1001/jamainternmed.2015.8260] [Medline: 26903176]
59. Ybarra M, Bağci Bosi AT, Korchmaros J, Emri S. A text messaging-based smoking cessation program for adult smokers: randomized controlled trial. J Med Internet Res 2012 Dec 27;14(6):e172 [FREE Full text] [doi: 10.2196/jmir.2231] [Medline: 23271159]
60. Mays D, Johnson AC, Phan L, Sanders C, Shoben A, Tercyak KP, et al. Tailored mobile messaging intervention for waterpipe tobacco cessation in young adults: A randomized trial. Am J Public Health 2021 Sep;111(9):1686-1695. [doi: 10.2105/AJPH.2021.306389] [Medline: 34436927]

XSL•FO
RenderX

61. García-Pazo P, Sesé A, Llabrés J, Fornés-Vives J. NoFumo+: A clinical trial of an mHealth for smoking cessation with hospitalized patients. Int J Environ Res Public Health 2021 Oct 06;18(19):10476 [FREE Full text] [doi: 10.3390/ijerph181910476] [Medline: 34639776]
62. Chulasai P, Chinwong D, Vientong P, Lertsinudom S, Kanjanarat P, Hall JJ, et al. Smartphone application for smoking cessation (Quit with US): A randomized controlled trial among young adult light smokers in Thailand. Int J Environ Res Public Health 2022 Jul 06;19(14):8265 [FREE Full text] [doi: 10.3390/ijerph19148265] [Medline: 35886120]
63. Cahill K, Lindson-Hawley N, Thomas KH, Fanshawe TR, Lancaster T. Nicotine receptor partial agonists for smoking cessation. Cochrane Database Syst Rev 2016 May 09(5):CD006103 [FREE Full text] [doi: 10.1002/14651858.CD006103.pub7] [Medline: 27158893]
64. Howes S, Hartmann-Boyce J, Livingstone-Banks J, Hong B, Lindson N. Antidepressants for smoking cessation. Cochrane Database Syst Rev 2020 Apr 22;4:CD000031 [FREE Full text] [doi: 10.1002/14651858.CD000031.pub5] [Medline: 32319681]
65. Stead LF, Perera R, Bullen C, Mant D, Hartmann-Boyce J, Cahill K, et al. Nicotine replacement therapy for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD000146. [doi: 10.1002/14651858.CD000146.pub4] [Medline: 23152200]
66. Hartmann-Boyce J, Chepkin SC, Ye W, Bullen C, Lancaster T. Nicotine replacement therapy versus control for smoking cessation. Cochrane Database Syst Rev 2018 May 31;5:CD000146 [FREE Full text] [doi: 10.1002/14651858.CD000146.pub5] [Medline: 29852054]
67. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9:CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
68. Carrasco-Hernandez L, Jódar-Sánchez F, Núñez-Benjumea F, Moreno Conde J, Mesa González M, Civit-Balcells A, et al. A Mobile Health Solution Complementing Psychopharmacology-Supported Smoking Cessation: Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Apr 27;8(4):e17530 [FREE Full text] [doi: 10.2196/17530] [Medline: 32338624]
69. Vidrine DJ, Frank-Pearce SG, Vidrine JI, Tahay PD, Marani SK, Chen S, et al. Efficacy of Mobile Phone-Delivered Smoking Cessation Interventions for Socioeconomically Disadvantaged Individuals: A Randomized Clinical Trial. JAMA Intern Med 2019 Feb 01;179(2):167-174 [FREE Full text] [doi: 10.1001/jamainternmed.2018.5713] [Medline: 30556832]
70. Wetter DW, McClure JB, Cofta-Woerpel L, Costello TJ, Reitzel LR, Businelle MS, et al. A randomized clinical trial of a palmtop computer-delivered treatment for smoking relapse prevention among women. Psychol Addict Behav 2011 Jun;25(2):365-371 [FREE Full text] [doi: 10.1037/a0022797] [Medline: 21500879]
71. Asma S, Mackay J, Song SY, Zhao L, Morton J, Palipudi KM. The GATS Atlas. URL: http://gatsatlas.org/ [accessed 2022-11-09]
72. López-López JA, Marín-Martínez F, Sánchez-Meca J, Van den Noortgate W, Viechtbauer W. Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. Br J Math Stat Psychol 2014 Feb;67(1):30-48. [doi: 10.1111/bmsp.12002] [Medline: 23297709]
73. Garnett C, Crane D, Brown J, Kaner E, Beyer F, Muirhead C, et al. Reported Theory Use by Digital Interventions for Hazardous and Harmful Alcohol Consumption, and Association With Effectiveness: Meta-Regression. J Med Internet Res 2018 Feb 28;20(2):e69 [FREE Full text] [doi: 10.2196/jmir.8807] [Medline: 29490895]
74. Blalock DV, Calhoun PS, Crowley MJ, Dedert EA. Telehealth Treatment for Alcohol Misuse: Reviewing Telehealth Approaches to Increase Engagement and Reduce Risk of Alcohol-Related Hypertension. Curr Hypertens Rep 2019 Jun 17;21(8):59. [doi: 10.1007/s11906-019-0966-3] [Medline: 31209579]
75. Heinz A, Deserno L, Zimmermann US, Smolka MN, Beck A, Schlagenhauf F. Targeted intervention: Computational approaches to elucidate and predict relapse in alcoholism. Neuroimage 2017 May 01;151:33-44. [doi: 10.1016/j.neuroimage.2016.07.055] [Medline: 27480622]
76. Brown B, Gude WT, Blakeman T, van der Veer SN, Ivers N, Francis JJ, et al. Clinical Performance Feedback Intervention Theory (CP-FIT): a new theory for designing, implementing, and evaluating feedback in health care based on a systematic review and meta-synthesis of qualitative research. Implement Sci 2019 Apr 26;14(1):40 [FREE Full text] [doi: 10.1186/s13012-019-0883-5] [Medline: 31027495]
77. Hysong SJ, Kell HJ, Petersen LA, Campbell BA, Trautner BW. Theory-based and evidence-based design of audit and feedback programmes: examples from two clinical intervention studies. BMJ Qual Saf 2017 Apr;26(4):323-334. [doi: 10.1136/bmjqs-2015-004796] [Medline: 27288054]

Abbreviations

ADI: automated digital intervention
LMIC: low- and middle-income countries
NRT: nicotine replacement therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

XSL•FO
RenderX

RR: risk ratio
TCS: theory coding scheme

Edited by R Kukafka; submitted 23.03.22; peer-reviewed by S Wen, W Wei, J Thrasher; comments to author 23.07.22; revised version received 13.09.22; accepted 01.11.22; published 16.11.22.

Please cite as:
Sha L, Yang X, Deng R, Wang W, Tao Y, Cao H, Ma Q, Wang H, Nie Y, Leng S, Lv Q, Li X, Wang H, Meng Y, Xu J, Greenshaw AJ, Li T, Guo WJ
Automated Digital Interventions and Smoking Cessation: Systematic Review and Meta-analysis Relating Efficiency to a Psychological Theory of Intervention Perspective
J Med Internet Res 2022;24(11):e38206
URL: https://www.jmir.org/2022/11/e38206
doi:10.2196/38206
PMID:36383408

©Leihao Sha, Xia Yang, Renhao Deng, Wen Wang, YuJie Tao, HaiLing Cao, Qianshu Ma, Hao Wang, Yirou Nie, Siqi Leng, Qiuyue Lv, Xiaojing Li, Huiyao Wang, Yajing Meng, Jiajun Xu, Andrew J Greenshaw, Tao Li, Wan-jun Guo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Devices for Assessing Motor Functions in Mobility-Impaired and Healthy Populations: Systematic Literature Review

Christine C Guo[1*], PhD; Patrizia Andrea Chiesa[1*], PhD; Carl de Moor[1*], PhD; Mir Sohail Fazeli[2*], MD, PhD; Thomas Schofield[2*], PhD; Kimberly Hofer[2*], BSc; Shibeshih Belachew[1*], MD, PhD; Alf Scotland[3*], MSc

[1]Biogen Digital Health, Biogen Inc, Cambridge, MA, United States

[2]Evidinno Outcomes Research Inc., Vancouver, BC, Canada

[3]Biogen Digital Health International GmbH, Biogen Inc, Baar, Switzerland

[*]all authors contributed equally

Corresponding Author:
Alf Scotland, MSc
Biogen Digital Health International GmbH
Biogen Inc
Neuhofstrasse 30
Baar, 6340
Switzerland
Phone: 41 79 414 69 23
Email: alf.scotland@biogen.com

Abstract

Background: With the advent of smart sensing technology, mobile and wearable devices can provide continuous and objective monitoring and assessment of motor function outcomes.

Objective: We aimed to describe the existing scientific literature on wearable and mobile technologies that are being used or tested for assessing motor functions in mobility-impaired and healthy adults and to evaluate the degree to which these devices provide clinically valid measures of motor function in these populations.

Methods: A systematic literature review was conducted by searching Embase, MEDLINE, CENTRAL (January 1, 2015, to June 24, 2020), the United States and European Union clinical trial registries, and the United States Food and Drug Administration website using predefined study selection criteria. Study selection, data extraction, and quality assessment were performed by 2 independent reviewers.

Results: A total of 91 publications representing 87 unique studies were included. The most represented clinical conditions were Parkinson disease (n=51 studies), followed by stroke (n=5), Huntington disease (n=5), and multiple sclerosis (n=2). A total of 42 motion-detecting devices were identified, and the majority (n=27, 64%) were created for the purpose of health care–related data collection, although approximately 25% were personal electronic devices (eg, smartphones and watches) and 11% were entertainment consoles (eg, Microsoft Kinect or Xbox and Nintendo Wii). The primary motion outcomes were related to gait (n=30), gross motor movements (n=25), and fine motor movements (n=23). As a group, sensor-derived motion data showed a mean sensitivity of 0.83 (SD 7.27), a mean specificity of 0.84 (SD 15.40), a mean accuracy of 0.90 (SD 5.87) in discriminating between diseased individuals and healthy controls, and a mean Pearson *r* validity coefficient of 0.52 (SD 0.22) relative to clinical measures. We did not find significant differences in the degree of validity between in-laboratory and at-home sensor-based assessments nor between device class (ie, health care–related device, personal electronic devices, and entertainment consoles).

Conclusions: Sensor-derived motion data can be leveraged to classify and quantify disease status for a variety of neurological conditions. However, most of the recent research on digital clinical measures is derived from proof-of-concept studies with considerable variation in methodological approaches, and much of the reviewed literature has focused on clinical validation, with less than one-quarter of the studies performing analytical validation. Overall, future research is crucially needed to further consolidate that sensor-derived motion data may lead to the development of robust and transformative digital measurements intended to predict, diagnose, and quantify neurological disease state and its longitudinal change.

(J Med Internet Res 2022;24(11):e37683) doi:10.2196/37683

XSL•FO
RenderX

KEYWORDS

motor function; medical devices; computers; handheld; smartwatch; smartphone; mobility; wearable electronic devices; Parkinson disease; Parkinsonian disorders; gait; mobile phone

Introduction

Background

Patient care is changing with the dawn of smart sensing technology. Mobile and wearable devices can provide continuous as well as objective monitoring and assessment of many health outcomes [1]. Until recently, outcomes that represent various motor functions (ie, any movement of the entire body or part of the body that is controlled by motor neuron activity) have typically been measured by patient reports (eg, number of falls) or physician assessment (eg, gait abnormalities). Physician assessments are based on very brief observations in an office or clinic [2], whereas self-reported outcomes are subjective and often not as sensitive nor as supervised as in-clinic measures [3]. Finally, measurements may vary between assessors depending on the level of training, familiarity, and experience [4,5].

Wearable technologies have recently emerged as a potential supplemental source of data on motor function. Such technologies could increase the objectivity and ease of assessment for motor functions during clinical trials and care while also allowing for a richer dimension of data to be captured. Real-world and continuous monitoring of patient motor functions through wearable and mobile sensors is increasingly being investigated in areas such as disease progression through motor fluctuations in Parkinson disease [6], detection of amyotrophic lateral sclerosis [7], and tremor activity in essential tremor [8].

Data from digital measurement solutions can enhance the quality of clinical trials, as illustrated by the acceptance of wearable device–measured stride velocity (95th percentile) by the European Medicines Agency (EMA) as an end point in Duchenne muscular dystrophy [9]. Given the implications these new data courses could have on the field, the current regulatory environment for mobile technologies is in flux [10]. US and European regulatory bodies are responding to this emerging opportunity by adapting their regulatory processes to these technological advances [11].

Objectives

Previous reviews have described the characteristics of their patient samples and sensors involved in collecting motor function data [12-20]. However, they do not evaluate the degree of validity produced by such sensors. This review follows the terminology used in previous reviews [21,22] and differentiates between analytical validation (ie, the same motion behavior is measured by an independent source and compared with the sensor-derived motion behavior) and clinical validation (ie, a clinical characteristic or measure of interest is measured and compared with the sensor-derived motion behavior). Gaining insight into the current clinical validity and utility of the data captured by mobile and wearable sensing technologies is of utmost importance. So, the aim of this study was to describe the existing scientific literature on digital measurement solutions that are being used or tested for assessing motor functions in mobility-impaired and healthy adults and to evaluate the degree to which these tools provide clinically valid measures of motor function in these populations. Specifically, we aimed to answer the following research questions: (1) What types of digital devices exist that capture motor function in mobility-impaired and healthy populations? (2) In what types of studies and in what populations have these devices been evaluated? (3) What outcomes do these digital devices measure? (4) What types of technologies and algorithms are used to capture and store the data? (5) To what degree have these technologies and their output been validated using established and recognized criteria?

Methods

Literature Review

This review was conducted in accordance with the Cochrane Handbook for Systematic Reviews of Interventions [23], and reporting is based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [24]. We included clinical trials (randomized and nonrandomized) as well as observational studies (case-control, retrospective cohort, prospective cohort, and cross-sectional) that provided validity estimates from wearable or mobile technologies to assess motor functions in adults (aged ≥18 years). Studies published in English after 2015 were included to focus on the most advanced technologies that are being used to assess motor function.

Study eligibility criteria were defined using an adapted PICO (Population, Intervention, Comparator, Outcomes) framework. We applied criteria based on the technology instead of the intervention or comparator, as the research question focused on the validity of measurement and not treatment efficacy (Table S1 in Multimedia Appendix 1 [25-115]).

A systematic literature search was conducted (January 1, 2015, to June 24, 2020) in the MEDLINE, Embase, and CENTRAL databases. Searches of relevant conferences for the last 3 years (2018-2020) were conducted via Embase. Search strings are available in Tables S2-S6 in Multimedia Appendix 1. Gray literature searches were also conducted to capture studies from sources that were not included in the main literature databases, which included the US Food and Drug Administration website as well as the United States and European clinical trials registry databases for clinical trials which had reported results but were not published in peer-reviewed journals (for the years 2018-2020).

After duplicate removal, all titles and abstracts were screened for potential eligibility according to the prespecified PICO criteria, after which full-text articles were assessed using the same criteria. Study selection was performed by 2 independent reviewers, and disagreements were resolved through discussion.

XSL•FO
RenderX

If no consensus could be achieved, a third researcher was consulted for arbitration.

A total of 2 independent reviewers extracted all relevant data from the final list of included studies. A reconciliation phase was again deployed to resolve any discrepancies between the reviewers, and a third reviewer intervened to resolve any remaining conflicts. The following data were extracted where available: (1) authors, year of publication, country, study setting, and follow-up period; (2) study design; (3) participant characteristics; (4) outcomes; (5) technology characteristics; and (6) validity outcomes. Motor function outcomes were manually sorted into categories by reviewers to facilitate summary where necessary.

Study Quality

A total of 2 independent reviewers assessed the quality of the included studies using the ROBINS-E (Risk Of Bias In Nonrandomized Studies of Exposures) tool [116]. A third investigator intervened to reach consensus if there were any remaining unresolved discrepancies following reconciliation between the decisions of the 2 reviewers.

Statistical Analyses

Effect size estimates were extracted from each study where reported, including standardized mean differences (ie, Cohen *d*), correlation coefficients (eg, Pearson *r*), sensitivity, specificity, accuracy, and area under the curve (AUC). In cases where studies provided none of these aforementioned effect size classes, effect sizes were calculated based on the information available in the manuscript using standard formulas [117,118]. To facilitate comparison across the studies, extracted effect sizes were converted to Pearson *r*–based effect size estimates where possible. This extraction and conversion process allows for studies to be directly compared via *r*-based effect sizes, estimates of sensitivity and specificity, and estimates of accuracy. The average effect sizes were calculated across all studies as well as by specific study and sample characteristics of interest. As *r* is bound by −1 and +1, *r*s were transformed into *Zr* using the procedure described by Fisher for analyses [119,120] and then back-transformed for reporting. Differences across groups in the magnitude of obtained effect sizes were tested using restricted information maximum likelihood derived SEs [117] using the inverse variance weight [121]. A random effects approach was taken, which includes in the denominator an extra variance component representing true variation in the population from which the included studies can be considered a random sample. A significance threshold of .05 was used to determine if values significantly differed between groups.

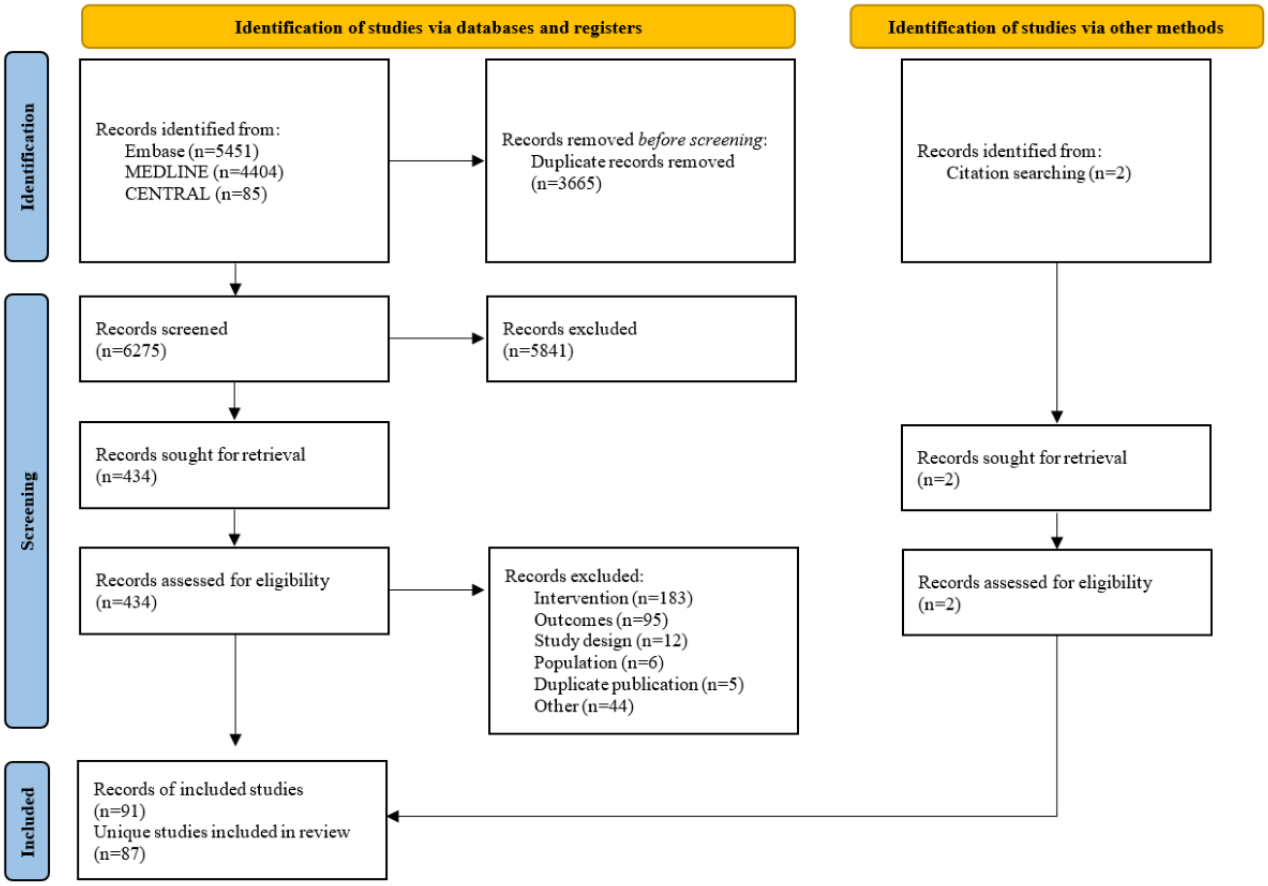
Results

Study Selection

A total of 9940 abstracts were identified from the electronic databases, and 2 articles [25,26] were included from handsearching of a systematic review identified in our searches [122]. After the removal of duplicates and exclusion based on title and abstract screening, 436 records remained for the full-text screening. A list of the records excluded during full-text screening and the reason for exclusion are provided in Table S7 in Multimedia Appendix 1. A total of 91 publications describing 87 primary studies fulfilled all inclusion criteria (Figure 1).

XSL•FO
RenderX

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



Study Characteristics and Data Collection

Across the 87 studies (n), the most common country settings reported were the United States (n=15) [27-41], United Kingdom (n=10) [42-53], Italy (n=5) [54-58], Spain (n=4) [59-62], South Korea (n=4) [63-66], Germany (n=3) [67-69], and Japan (n=3) [70-72]. At least 1 study was conducted in each of the following countries: Canada (n=2) [73,74], the Netherlands (n=2) [75,76], Portugal (n=2) [77,78], Sweden (n=1) [79,80], Taiwan (n=2) [81,82], Australia (n=1) [83], Brazil (n=1) [84], Demark (n=1) [85], France (n=1) [86], Israel (n=1) [87], Greece (n=1) [88,89], Lithuania (n=1) [90], Norway (n=1) [91], and United Arab Emirates (n=1) [92]. Of the remaining reporting studies, 6 were multinational [93-98]. Sample size ranged from 8 [33] to 1465 [94] (median 40.5 participants). A total of 7995 participants were enrolled in the included studies. Table S8 in Multimedia Appendix 1 presents the list of included publications as well as key study characteristics.

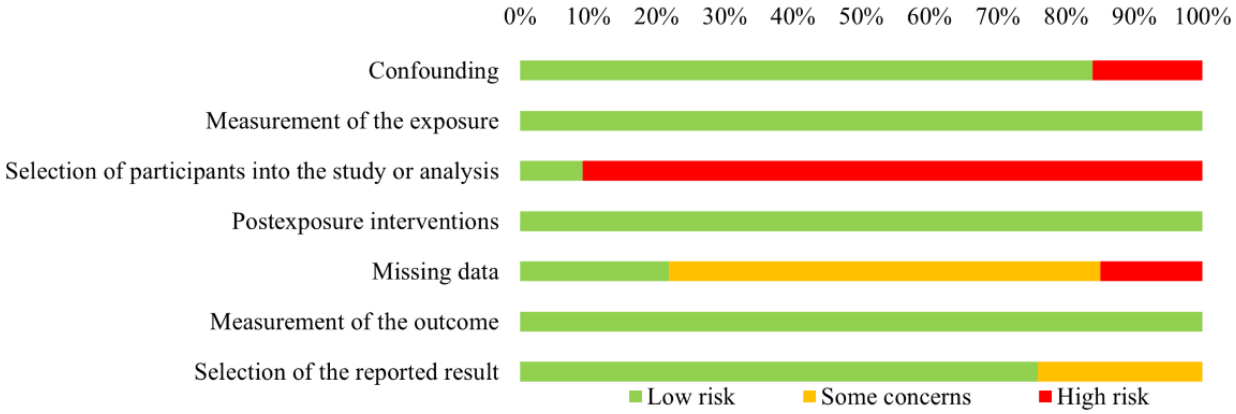
All 87 studies were observational in nature. Most studies (n=50) did not report whether the study was conducted in a single-center or multicenter setting. However, among those that did report, 20 and 17 studies were single center and multicenter, respectively. Approximately half of the included studies were conducted in a laboratory setting (n=42), 11 studies were home based, and 15 were a combination of a laboratory-based and home-based setting. The remaining 19 studies did not specify the study setting. The included studies were categorized into 2 follow-up types: cross-sectional (n=62) with a follow-up period of ≤1 week and longitudinal (n=25) in which participants were followed up for ≥1 week. Follow-up length of longitudinal studies ranged from 7 days [42,45,59,91,99] to 8 years [46]. A total of 30 studies reported the time allocated for data collection; in other words, the time needed to collect data in one session of data collection. In addition, 18 studies were able to capture their data in a session between 20 seconds [52,95] and 24 hours [71]. Moreover, 13 studies required their participant to use the device for multiple days for their collection period, which ranged from 2 [41,62] to 14 consecutive days [40]. This review follows the terminology used in previous reviews [21,22] for analytical validation (ie, the same motion behavior is measured by an independent source and compared with the sensor-derived motion behavior) and clinical validation (ie, a clinical characteristic or measure of interest is measured and compared with the sensor-derived motion behavior). Analytic validation was only performed in 21% (13/62) of cross-sectional studies and 4% (1/25) of longitudinal studies. Most of these studies performed clinical validation of sensor-based motion data. Studies applied a wide variety of technologies to capture motion outcomes. Motion data were captured by ≥30 different devices, including novel wearables (18/42, 43% devices), smartphone or smart watch (13/42, 31%), mass market digital technology (7/42, 17%), other digital technology (eg, PC; 3/42, 7%), and mass market wearables (1/42, 2%). Approximately 1 in 5 studies included a mass market device.

In terms of quality, studies were generally low to moderate risk of bias (Figure 2; Table S9 in Multimedia Appendix 1). Less than 20% (14/42) of studies did not show that groups were balanced in terms of key baseline characteristics and were

XSL•FO
RenderX

considered high risk for confounding. The risk of bias arising from measurement of the exposure was most often low because exposures were generally whether the patient had a disease or was healthy, and misclassifications were next to nonexistent. For the domain of selection of participants into the study, studies were often high risk of bias. Disease diagnosis (ie, the exposure) did not generally coincide with the start of follow-up, and the diseases being studied could fluctuate over time. Many of the studies relied on volunteers to participate in the study, and this may have led to participants entering the study if they were in a particularly good or bad disease state (eg, Parkinson disease has *on* and *off* states). Furthermore, no corrections that may have alleviated selection biases in the analysis were conducted. Studies were generally low risk with regard to the domain concerned with the risk of bias owing to postexposure interventions. By design, the included studies did not administer interventions to alleviate the effects of exposures, and therefore, bias was not a concern. Regarding missing data, this was not often accounted for, leading to high risk of bias in that domain. However, studies were generally low risk of bias for measurement of outcomes, as motor function outcomes were assessed objectively and similarly across groups. Finally, over half of the studies were rated low risk for selection of the reported result.

Figure 2. Distribution of study quality across included studies.



Concepts of Interest and Context of Use

Approximately half of the included studies compared the association between sensor-derived motion data and a standardized clinical assessment across diverse disease conditions (n=44). Other studies compared mobility-impaired diseased participants to a healthy control group of participants with no mobility impairment (n=43). The most represented disease condition was, by far, Parkinson disease (n=51); stroke (n=5); Huntington disease (n=5); and depression, cognitive impairment, cerebral palsy, and multiple sclerosis (n=2 for each). All other disease groups were only represented in a single study.

Among the 67 studies that reported the mean age of participants, values ranged from 23.6 years [92] to 77.2 years [95] for mobility-impaired participants and from 19.5 years [29] to 78.9 years [87] for healthy participants. Control groups were generally well-matched by participant age and sex. Among the 71 studies that reported the proportion of males or females in their sample, the average percentage of the sample that were male ranged from 22.8% [62] to 100% [72,84] in mobility-impaired participants and from 11% [41] to 100% [84] in healthy participants. Studies with the largest sex imbalances were those addressing the less frequently studied disease states (ie, represented in only 1 or 2 studies). In contrast, Parkinson disease, Huntington disease, and stroke reflected a more balanced representation of females and males.

The primary motion outcomes were gait (n=30), gross motor movements (n=25), fine motor movements (n=23), motor symptom severity (n=9), bradykinesia (n=7), motor fluctuations (n=6), dyskinesia (n=5), balance control (n=5), postural stability (n=4), voice or speech impairments (n=3), facial expression impairments (n=1), and nocturnal movements (n=1). A summary of commonly reported outcomes by disease that the outcome was measured in is provided in Table 1.

The most common motions that participants were required to enact for sensor data collection across these studies were based on diverse active motor tasks: multimovement tasks (16/87, 18%) including balancing and reaction time during tests such as the Timed Up and Go, the Cognitive Dual Task Timed Up and Go, and the Manual Dual Task Timed Up and Go, unscripted daily activities (17/87, 20%), walking (10/87, 11%), tapping (9/87, 10%), and scripted activities of daily living (7/87, 8%). Less commonly used motions (<5% of studies) included several real-world tasks such as reaching, sit-to-stand motion, seated tremors, wrist pronation-supination tracing or pointing, typing, seated conversation, standing, and sleeping movement. Together, these motions were used to extract ≥75 distinct motion outcomes across the included studies. Most of these outcomes only appeared in one study and were only measured at a single sensor location in each study (per our inclusion criteria). One exception was walking cadence, with different studies measuring it using sensors worn at wrists, ankles, lower back, and chest and in the pants pocket. Additional exceptions were tremor, dyskinesia, and bradykinesia (each measured using sensors placed on the wrists or ankles).

XSL•FO
RenderX

Table 1. Summary of commonly reported outcomes by disease in which the outcome was investigated.

Disease and motor function outcome category	Motor function outcome
Acquired brain injury	
Gross motor impairment or performance and upper body	• Peak upper limb velocity [35] • Upper limb velocity [35]
Alzheimer disease	
Fine motor impairment or performance and continuous motion	• Spiral tracing [82]
Depressive tendencies	
Fine motor impairment or performance and discrete motion	• Finger tap speed [92] • Flight time [92] • Hold time [92]
Healthy participants	
Bradykinesia	• Bradykinesia score [94,100]
Dyskinesia	• Dyskinesia score [100]
Fine motor impairment or performance and continuous motion	• Spiral tracing [82,90]
Fine motor impairment or performance and discrete motion	• Correct finger taps [25,83] • Finger tap accuracy [38,101] • Finger tap count [38,95,101] • Finger tap duration [38,101] • Finger tap interval [38,101] • Finger tap reaction time [38,42,58] • Finger tap rhythm [42,95] • Finger tapping test [102] • Flight time [83,88,103] • Hold time [88]
Gait	• Joint velocity [77] • Step cadence [69,75,81,99] • Step count [40,41,44,74,104] • Step length [44,46,81] • Stride duration [44] • Turning speed [26] • Walking speed [41,69,81]
Gross motor impairment or performance and lower body	• Lower limb velocity [105]
Gross motor impairment or performance and whole body	• Joint velocity [106]
Motor symptom severity	• Rest tremor [102]
Postural stability	• Trunk acceleration [50]
Huntington disease	
Cognitive impairment	• Stroop Color and Word Test [96]
Dyskinesia	• Chorea score [96,107]
Fine motor impairment or performance, discrete motion	• Finger tap speed [96]
Gait	• Step cadence [99]
Mild cognitive impairment	
Fine motor impairment or performance and continuous motion	• Spiral tracing [82]
Multiple sclerosis	

XSL•FO
RenderX

Disease and motor function outcome category	Motor function outcome
Fine motor impairment or performance and discrete motion	• Finger tap count [25]
Gait	• Turning speed [26]
Neurological disorders[a]	
Fine motor impairment or performance and continuous motion	• Spiral tracing [90]
Neuromuscular disorders[b]	
Gait	• Step count [104]
Parkinson disease	
Bradykinesia	• Bradykinesia score [34,48,53,94,97,100,108]
Cognitive impairment	• Stroop Color and Word Test [83]
Dyskinesia	• Dyskinesia score [53,100] • Finger tapping test [56]
Fine motor impairment or performance and discrete motion	• Correct finger taps [83,109] • Finger tap accuracy [38,101] • Finger tap count [38,95,101] • Finger tap duration [38,101] • Finger tap interval [38,101] • Finger tap reaction time [38,42,49] • Finger tap rhythm [42,95] • Finger tapping test [102] • Flight time [88,103,110] • Hold time [88]
Gait	• Freezing of gait [49,54,61,64,93,111,112] • Step cadence [75] • Step count [31,40] • Step length [44,46] • Stride duration [44] • Turning speed [97]
Gross motor impairment or performance and upper body	• Peak upper limb velocity [33]
Gross motor impairment or performance and whole body	• Joint velocity [106]
Motor fluctuations	• On or off state [34,60,62,68,98]
Motor symptom severity	• Rest tremor [49,102] • Tremor test [34,48,97]
Postural stability	• Trunk acceleration [50]
Rapid eye movement (REM) sleep behavior disorder	
Fine motor impairment or performance and discrete motion	• Finger tap reaction time [42] • Finger tap rhythm [42]
Stroke	
Gait	• Step cadence [81] • Step count [41,74] • Step length [81] • Walking speed [41,81]
Transthyretin familial amyloid polyneuropathy	

XSL•FO
RenderX

Disease and motor function outcome category	Motor function outcome
Gait	• Lower limb velocity [78] • Step length [78] • Stride duration [78] • Walking speed [78]
Gross motor impairment or performance and upper body	• Upper limb velocity [78]

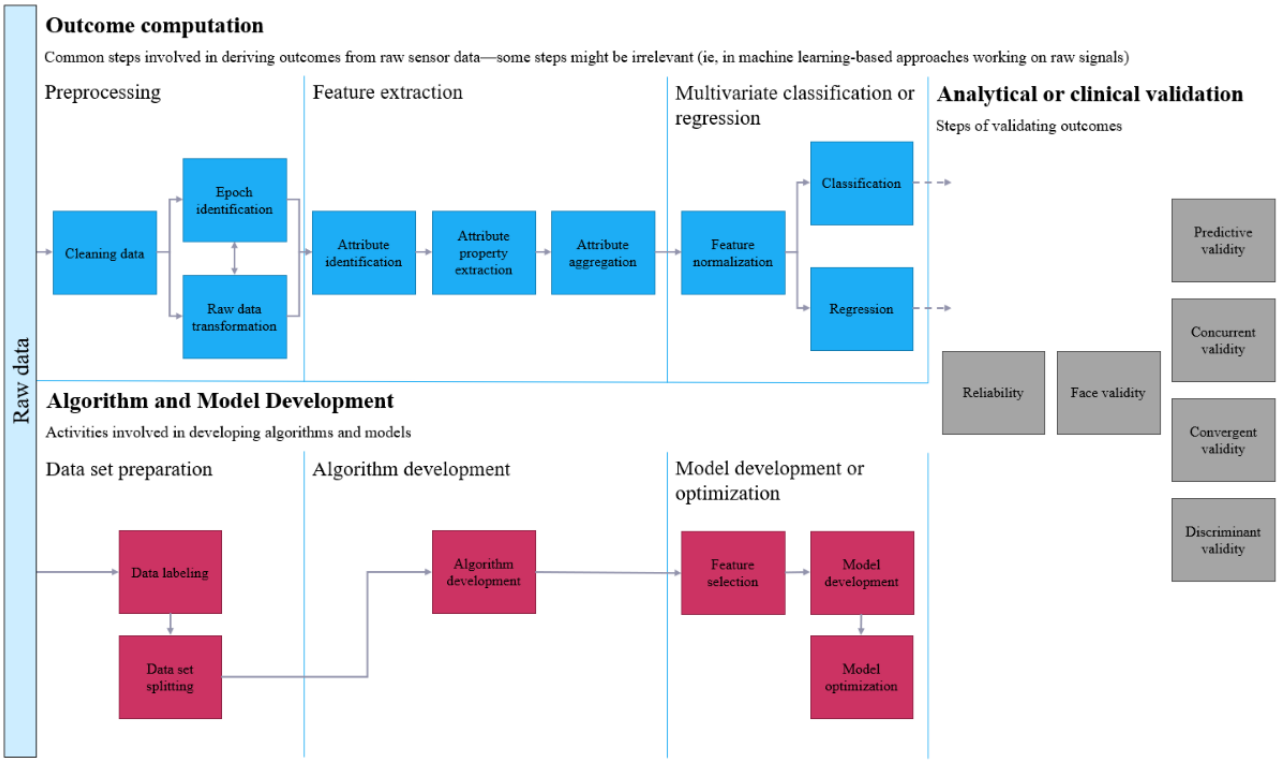
[a]Including Parkinson disease, Huntington disease, early dementia, cerebral palsy, and poststroke.

[b]Including Duchenne muscular dystrophy, limb-girdle muscular dystrophy, and spinal muscular atrophy.

Data Processing and Analysis

The process through which these researchers converted their raw data to validity coefficients is illustrated in Figure 3. On collection of the raw data, 2 parallel processes were typically seen: outcome computation and algorithm or model development. Following the completion of these 2 processes, the model was subjected to either analytical or clinical validation.

Figure 3. Flowchart of the process of converting raw data to validity coefficients.



Outcome Preparation

In ≥90% of the studies, the raw data were first preprocessed before feature processing engineering and analyses. One preprocessing step frequently seen among these studies was the splitting of raw data into temporal epochs or slices. This was done because training an algorithm to detect movement features across long periods greatly reduced the algorithm's validity. Data were trimmed by temporal position (eg, the beginning and ending of the motion recording) or based on extreme values (eg, outliers >4 SDs from the mean). Raw data were subjected to some form of standardization or transformation in ≥90% of the studies.

Although algorithm training (eg, feature selection and threshold determination) typically occurred using data across all participants, several studies took the approach of building the feature detection algorithm using data across all participants but then allowing each participant to vary in latter stages such as feature selection or determining thresholds [34,54,63,68]. Validity estimates from this smaller group of studies were similar in magnitude to those studies that applied the same features and thresholds to the classification of all participants.

Researchers have to decide which of the hundreds of identified candidate features to treat as a signal (by retaining them in the model) and which to dismiss as mostly noise (by excluding them from the model). Relatively few studies clearly described whether they moved all detected features to the next analytic stage (feature selection), but some studies compared prediction based on all extracted features to prediction based on top-performing features [42,49]. These studies reported that the inclusion of additional features did not guarantee a meaningful increase in algorithm performance or validity. One study using smartphones to assess Parkinson disease symptoms found AUC values >0.90 for 998 detected features, with a drop to 0.75 when

XSL•FO
RenderX

based on the top 30 features [49]. A second study of participants with Parkinson disease concluded, "Accuracies obtained using the 30 most salient features were broadly comparable with the corresponding sensitivity and specificity values obtained using all 998 features" [42].

Algorithm or Model Development

The included studies showed no clear preference regarding algorithms for feature selection or classification, but the 2 most frequently applied approaches were support vector machines (12/87, 14%) and random forests (4/87, 5%). Authors of these studies were sensitive to the complications of trying to train a classification model with groups of different sizes, as most of the comparative studies included in this review include approximately equal sizes of participants with a disease or disorder and healthy controls.

No consistent pattern emerged from within-study comparisons of feature selection algorithms. A wrist-based sensor was able to detect upper limb movement among participants with pre-Parkinson disease best when using random forests relative to support vector machines and naïve Bayes [55]. A smartphone app testing motor impairment found that both neural networks and boosting outperformed support vector machines and Fisher linear discriminant analysis [90]. Not all motions required feature selection across studies (several needed only to define logic rules to estimate movement angles using geometry), and some studies used proprietary algorithms that were not described in detail. One study that studied freezing of gait among participants with Parkinson disease using a smartphone app found neural networks performed better than other bagging algorithms, including random forest, multilayer perception, decision tree, support vector machine, and naïve Bayes [64]. Another study on motor symptoms among participants with Parkinson disease using ankle-worn sensors found that support vector machines performed better than logistic regression and decision trees [80]. Using smartphone motion data to predict motor impairment among participants with Parkinson disease, another study found that random forests based on Ridge regression outperformed those based on Lasso, or Gini impurity, and that linear support vector machines outperformed logistic regression and boosting [103]. The sole consistent pattern that emerged was that supervised machine learning techniques performed better than unsupervised techniques (eg, naïve Bayes).

Analytical and Clinical Validation

The most common validity criterion was clinical condition (37/87, 43%), which was used in many of these studies to establish known-group construct discriminant validity of sensor-derived motion data by comparing participants with a diseased condition to healthy controls (Table S10 in Multimedia Appendix 1). The second most common validity criterion was the clinical validity established by assessing the convergence or concurrence with traditional standardized clinical assessments (30/87, 34%; eg, Wolf Motor Function Test and Unified Parkinson Disease Rating Scale). Other criteria were clinician ratings (7/87, 8%), research device (9/87, 10%), treatment status (3/87, 3%), and patient-reported outcome (1/87, 1%). Longitudinal studies were more likely to use nonsupervised assessments, whereas cross-sectional studies were more likely to use clinician-supervised assessments.

Across studies, motion data from the sensors identified showed an average Pearson *r* clinical validity coefficient of 0.52 (Figure 4 [27,28,31,35-41,44,47,48,50-53,57,58,66-74,76,77,80-84,86,91,92,95-99,101,102,104,106,108-110,112,113,115]). Among the studies that did not provide sufficient information to calculate a Pearson *r*, the average validity was 0.83 (sensitivity), 0.84 (specificity), and 0.90 (accuracy). These values could be interpreted as very good [123]. The magnitude of validity coefficients did not vary (P=.10) between health care–related devices (mean r=.47), personal electronic devices (mean r=.44), and entertainment consoles (mean r=.63). Validity coefficients for motor function generated by healthy adults were higher than those generated by participants with a disease state or impairment (z score 3.19; P=.001). The only statistical decision that consistently predicted higher validity coefficients was the decision to trim observations during the preprocessing stage based on value (ie, outliers; z score 2.10; P=.04). There was no difference in validity coefficients across trimming observations based on temporal placement, transforming data, standardizing data, or which feature detection and validation analyses were used. The funnel plot from these studies was asymmetrical in a manner consistent with bias toward higher coefficients (Figure S1 in Multimedia Appendix 1). The magnitude of validity coefficients did not significantly vary across the different device types (Table 2).

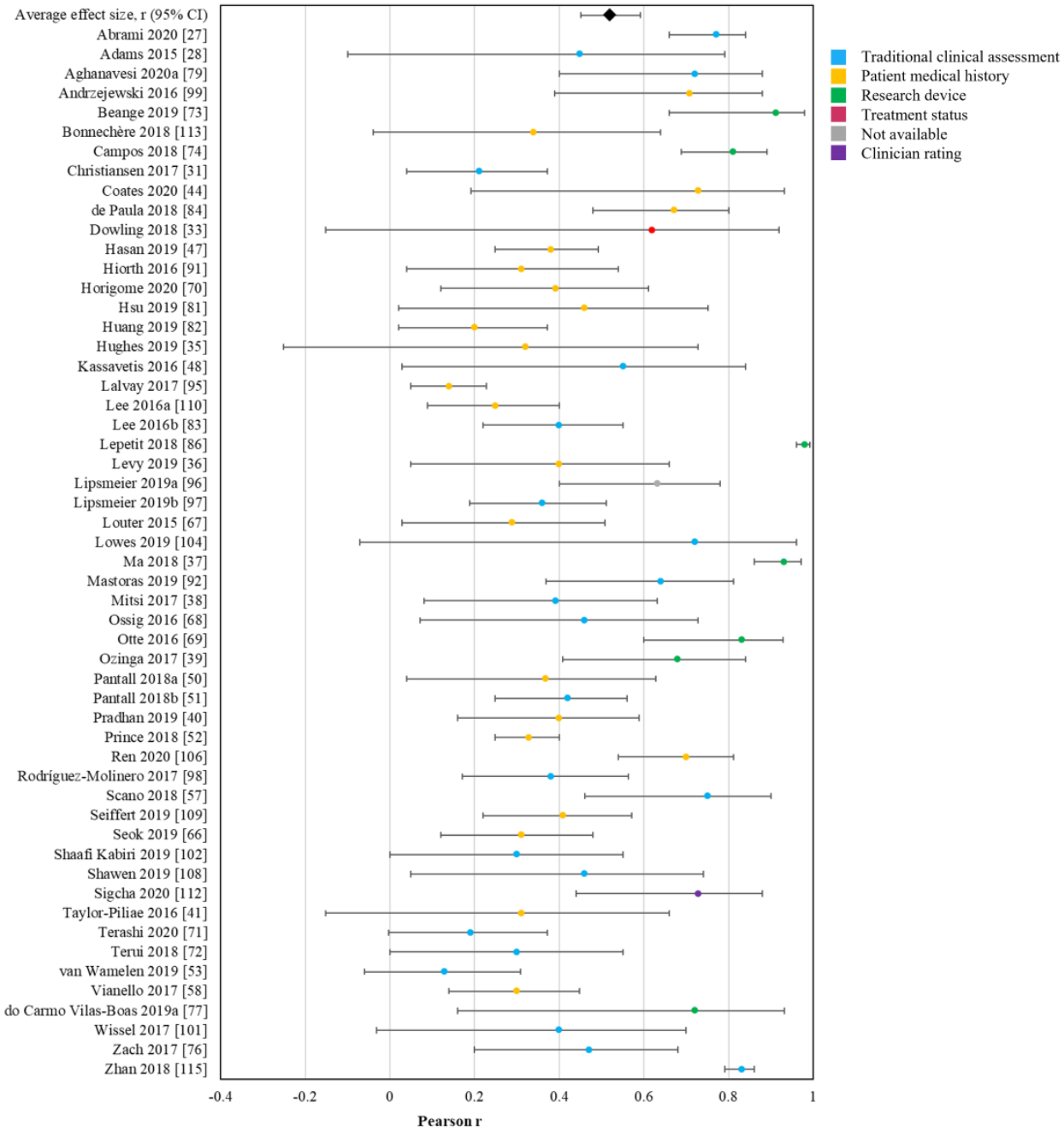
Taken as a whole, no consistent pattern emerged from within-study comparisons of the relative analytic validity of any specific motion signal. One study using Kinect found high Pearson *r* validity coefficients (r>0.50) for more than 40 distinct motion outcomes but very low validity coefficients for a handful including deflection range roll (measured in degrees), mean sway velocity roll (measured in degrees per second), and up-down deviation (measured in centimeters) [69]. A second study using Kinect found Pearson *r* validity coefficients above 0.50 for variables related to steps taken, distance, and speed but coefficients below 0.50 for variables related to angles (eg, trunk, hips, ankle, trunk, upper limb, and full body) [78]. A third study using a triaxial accelerometer worn on the waist found Pearson *r* validity coefficients above 0.50 for gait, arising from chair, body bradykinesia, hypokinesia, and overall posture and validity coefficients below 0.50 for rigidity of lower and upper extremities axial rigidity, postural stability, legs agility, and tremors in lower or upper extremities [98]. These numbers are in the same range as single items from widely established clinical tools [124-126]. As the validity coefficients for these single motions were moderate, it reinforces the need for future studies and clinical applications to include multiple validated motion signals for any screening or diagnostic tool to achieve adequate levels of composite test validity.

Regarding clinical validation, no clear within-study evidence emerged regarding the relative superiority or inferiority of motion data captured in laboratory settings versus data captured in home settings (Table 1). For example, 1 study comparing typing behavior of participants recently diagnosed with Parkinson disease to the typing behavior of healthy controls found AUC values of 0.76 (when administered at home) versus

XSL•FO
RenderX

0.83 (when administered in clinic) [59]. A second study comparing participants with Parkinson disease to healthy adults on motor function during an activities of daily living task found slightly higher accuracy, sensitivity, and specificity when the task was completed at home [87].

Figure 4. Forest plot of the validity of sensor-derived digital measurements of motor function. Middle points represent the point estimate effect size Pearson r, and the surrounding bars represent 95% CI. Colors indicate the type of validity criteria used.



XSL•FO
RenderX

Table 2. Summary table of the between-study and within-study findings on the differences in the validity of sensor-derived measurements of motor function across various groups.

Are there differences in the validity of sensor-derived measures of motor function as captured	Between-study (ie, meta-analytic) findings	Within-study findings
Using mass market devices vs medical sensors?	• No: digital technology vs mass market digital technologies (*P*=.22); mass market digital technology vs medical devices (*P*=.21); digital technology vs medical devices (*P*=.32)	Insufficient data to evaluate
At specific sensor locations?	• No: wrist vs ankle (*P*=.73); wrist vs chest (*P*=.73); wrist vs hand (*P*=.54); wrist vs thigh (*P*=.59); wrist vs back (*P*=.63); wrist vs pocket (*P*=.78); wrist vs nonwearable (0.31) • No: ankle vs chest (*P*=.46); ankle vs hand (*P*=.38); ankle vs thigh (*P*=.73); ankle vs waist (*P*=.60); ankle vs back (*P*=.49); ankle vs pocket (*P*=.65); ankle vs nonwearable (*P*=.58) • No: chest vs hand (*P*=.30); chest vs thigh (*P*=.39); chest vs waist (*P*=.70); chest vs back (*P*=.82); chest vs pocket (*P*=.50); chest vs nonwearable (*P*=.89) • No: hand vs thigh (*P*=.58); hand vs waist (*P*=.75); hand vs back (*P*=.78); hand vs pocket (*P*=.42); hand vs nonwearable (*P*=.53) • No: thigh vs waist (*P*=.86); thigh vs back (*P*=.73); thigh vs pocket (*P*=.54); thigh vs nonwearable (*P*=.40) • No: waist vs back (*P*=.87); waist vs pocket (*P*=.39); waist vs nonwearable (*P*=.24) • No: back vs pocket (*P*=.45); back vs nonwearable (*P*=.48); pocket vs nonwearable (*P*=.50)	Insufficient data to evaluate
home vs in the laboratory?	• No; *P*=.33	No; 1 study found AUC[a] values of 0.76 (when administered at home) vs 0.83 (when administered in clinic) [59]. A second study found slightly higher accuracy, sensitivity, and specificity when the task was completed at home [87].
In longitudinal vs cross-sectional studies?	• No; *P*=.29	No; One study found high Pearson *r* validity coefficients (*r*>0.50) for over 40 distinct motion outcomes but very low validity coefficients for a handful, including deflection rage roll (measured in degrees), mean sway velocity roll (measured in degrees per second), and up-down deviation (measured in centimeters) [69]. A second study found Pearson *r* validity coefficients above 0.50 for variables related to steps taken, distance, and speed, but coefficients below 0.50 for variables related to angles (eg, trunk, hips, ankle, trunk, upper limb, and full body) [78]. A third study found Pearson *r* validity coefficients above 0.50 for gait, arising from chair, body bradykinesia, hypokinesia, and overall posture and validity coefficients below 0.50 for rigidity of lower and upper extremities axial rigidity, postural stability, legs agility, and tremors in lower or upper extremities [98].
In healthy vs motor impaired patients?	• Yes; validity higher among healthy adults, z score 3.19, *P*=.001	Insufficient data to evaluate

XSL•FO
RenderX

Are there differences in the validity of sensor-derived measures of motor function as captured	Between-study (ie, meta-analytic) findings	Within-study findings
Using different feature detection algorithms?	• Insufficient data to evaluate	No; One study was able to detect movement best when using random forests relative to support vector machines and naïve Bayes [55]. A second study found that both neural networks and boosting outperformed support vector machines and Fisher linear discriminant analysis [90]. A third study found neural networks performed better than other bagging algorithms including random forest, multilayer perception, decision tree, support vector machine, and naïve Bayes [64]. A fourth study found support vector machines performed better than logistic regression and decision trees [80]. A fifth study found that random forests based on Ridge regression outperformed those based on Lasso, or Gini impurity, and that linear support vector machines outperformed logistic regression and boosting [103]. The sole consistent pattern that emerged was that supervised machine learning techniques performed better than unsupervised techniques (eg, naïve Bayes).
Using particular motion sensor signal types?	• Insufficient data to evaluate	Insufficient data to evaluate
Using all vs a subset of features?	• Insufficient data to evaluate	No; One study found AUC values >0.90 for 998 detected features, with a drop to 0.75 when based on the top 30 features [49]. A second study concluded "Accuracies obtained using the 30 most salient features were broadly comparable with the corresponding sensitivity and specificity values obtained using all 998 features" [42].
With the thresholds held constant across patients vs patient-specific thresholds?	• No; *P*=.48	No; Although algorithm training typically occurred across a sample, several studies took the approach of starting the algorithm (feature detection) using data across all participants but then allowing each patient to vary in later stages such as feature selection or determining thresholds [34,54,63,68]. Validity estimates from this smaller group of studies were similar in magnitude to those studies that applied the same features and thresholds to the classification of all participants.
Using clinically supervised vs nonsupervised assessments of patient clinical status?	• No; *P*=.16	Insufficient data to evaluate
With outliers trimmed vs retained in the feature detection stage?	• Yes; trimming outliers is beneficial, z score 2.10, *P*=.04	Insufficient data to evaluate
With transformed data vs untransformed data?	• No; *P*=.74	Insufficient data to evaluate
With standardized data vs unstandardized data?	• No; *P*=.60	Insufficient data to evaluate

[a]AUC: area under the curve.

Discussion

Principal Findings

To our knowledge, this is the first systematic literature review to evaluate the degree to which wearable and mobile technologies provide clinically valid measures of motor function in mobility-impaired and healthy adults. The identified literature generally consisted of proof-of-concept studies, which aimed to pilot a device and assess whether it could validly measure motor functions. Consequently, most studies used a short follow-up period (<1 week) and had a total sample size of <50 participants. Unsurprisingly, many of the longitudinal studies prioritized nonsupervised measures. Even so, taken together, these studies provide a respectable evidence base supporting the potential these movement sensors have to inform clinical practice.

As the eligibility criteria for our review were inclusive in terms of population, we identified a large range of disease types, which were all but one (chronic obstructive pulmonary disease) nervous system condition (Table 1); however, the most common

XSL•FO
RenderX

disease was Parkinson disease, with stroke and Huntington disease coming in a very distant second and third place. The strong focus on Parkinson disease in this literature may be because of its prevalence or perhaps because motor function symptoms are a major characteristic of Parkinson disease for diagnosis and prognosis assessment purposes, making Parkinson disease an ideal model disease for testing the use of mobile technologies [127]. However, it is most probably a mixture of these 2 hypotheses. Parkinson disease is also one of the few diseases with Food and Drug Administration–approved devices (eg, NexStride and Personal KinetiGraph), which assesses motor function to inform treatment decisions. The field would benefit from additional study of mobile technology–assessed motor function among other neurological diseases, including multiple sclerosis, spinal muscular atrophy, amyotrophic lateral sclerosis, and Alzheimer disease. In addition, future studies might consider the advantages of assessing digital devices per neurological impairment (such as difficulties in ambulation or upper limbs) rather than per disease.

Successful integration of wearable-based movement data into clinical tools requires both analytic validation and clinical validation. However, most of the reviewed literature compared wearable sensor-derived motion data to omnibus measures of functioning or disease progression (ie, clinical validation). More studies need to perform analytic validation by comparing wearable sensor-derived motion data to the same motions measured by another source (eg, observer assessment and motion-capture technology). Observed motions may be highly correlated with omnibus assessments of motor skills or disease status (ie, clinical validation), but the foundation of approval as a clinical end point can only be met if the motions identified using the sensor have been shown to be the exact motions that have been approved by the governing or regulatory body. Using as an example the EMA's recent approval of 95% stride velocity as an approved secondary end point in Duchenne muscular dystrophy, appeal to the EMA's approval of wearable sensor stride velocity data as an end point for a given study requires evidence that when the used algorithm claims to measure stride velocity (95th percentile), there be evidence that the algorithm has, in truth, measured stride velocity. Future research in this area should focus their attention on analytic validation.

There was considerable variation in methodological approaches. The review revealed one of the key reasons why this field may still show such inconsistency in analytic approach; it is still developing. Evidence of this is seen in which motion variables could be identified by the algorithms. Despite the hundreds of motion-derived outcome variables identified across these studies, not all theoretically meaningful motions could be recovered. One study of participants with Parkinson disease concluded, "Unfortunately, we failed to find parameters that reflected fatigue (decrement response) and hesitation (intertap irregularity), which are characteristics of motor dysfunction in Parkinson's disease" [110]. Those authors offered that more precise definitions of fatigue and hesitation may be needed to recover them in clinical settings with a smartphone-based tapping test similar to the one used in that study. In addition, the motor functions viewed by some authors as theoretically relevant were occasionally overshadowed by nonmotor signals. The tendency for studies to report diminishing returns after a certain point for additional motion signals is statistically analogous to other clinical efforts to identify causal markers from a multitude of candidates, which revealed many initially flagged markers as spurious [128]. Future studies should include graphical displays to identify inflection points (similar to the scree plot in factor analysis or the elbow plot in latent class analysis) to help show where the statistical signal (or true score) from additional motions becomes outweighed by statistical noise.

The moderate to high validity coefficients reported in the identified literature may support the potential for sensor-derived motor function data from digital health technology tools to eventually contribute to screening, diagnosis, and monitoring of neurological diseases in particular. No significant differences in analytic or clinical validity estimates were found when comparing data generated by mass market devices (eg, smartphones, smartwatches, and Fitbits), game consoles (ie, Nintendo Wii and Microsoft Kinect or Xbox), and marketed motion sensors (eg, ActiGraph, ActivPAL, Axivity, Dynaport, KinetiSense, Opal devices, and PAMSys-X). Furthermore, the motion data provided by these technologies produced equivalent validity estimates in laboratory-based and home-based settings. This further supports the future potential for digital measurement solutions to provide clinically meaningful data and eventually become the gold standard for assessing motor behaviors. The degree and rate of application for motor function data from these devices to clinical practice will depend on how soon clear evidence bases are established for given sensor locations for given movements of interest.

Translation of these motor signals into clinical application is aided by demonstrating sufficient validity outside the scripted protocols of a controlled laboratory setting. The reviewed literature showed that scripted motion tasks were important when only a few minutes of motion data were to be captured. Furthermore, motion data from unscripted everyday living with longer data collection periods were also shown to be adequate and deemed complementary, as episodic scripted assessments of confined tasks might not capture the complex spectrum of potentially altered components of motor function in an unconstrained ecologic setting [129].

As a whole, the reviewed literature revealed several best practices as well as a few cautionary tales for mobile or wearable sensor-based movement data. Although cross-validation techniques all seek to counteract the inflation of validity coefficients that can occur during machine learning techniques, they can produce different results [42]. Despite these best practices, there remained indirect evidence of model overfitting in the form of some abnormally high validity coefficients in the final models (ie, specificity of 1.0, which is perfect) [130,131].

The reviewed literature also highlights areas to consider during the development of any clinical application. One illustration from this review is the critical role of thresholds [132], which require researchers to decide between manual versus automatic thresholds [133] and global versus person specific [134]. Leveraging the strengths of these modeling approaches while

XSL•FO
RenderX

keeping them robust and flexible will be important to consider as they are scaled up to create clinical applications [132].

Comparison With Previous Reviews

We identified a number of similar literature reviews during our study selection [12-20]. All identified reviews synthesized their evidence qualitatively, and none provided a quantitative synthesis of the validity of motion data generated from these sensors among patients with neurological conditions. Of the 9 identified reviews, 1 was narrative [16], whereas the remaining were systematic reviews. None of the systematic reviews focused on neurological disorders. Overall, 2 reviews focused specifically on swimming motions [12,13], 2 were focused on older adult patients with no specific disease [15,19], and 2 reviews focused on only upper [14] and lower limb movements [18]. Of the remaining 2 systematic reviews with similar objectives and scope to that of our own, the paper by Díaz et al [17] aimed to review the current literature on the use of wearable sensors in gait, balance, and range of motion analysis. Diseases of participants also varied across their 56 included studies and included a mix of neurological disorders (eg, Parkinson disease, Alzheimer disease, and multiple sclerosis), as well as stroke, amputees, and healthy participants. Similar to our own review, the authors found that most body-worn devices were complex to use and required strong experience in data analysis to interpret the collected information. In addition, the authors pointed out a need for further validation and improvements in sensor systems for them to be used as reliable and accurate clinical devices. A second systematic review conducted by Kristoffersson and Lindén [20] provided a qualitative synthesis of 73 published articles on wearable body sensors used for health monitoring. Similar to our review, the authors found that included studies were generally observational in design and small in sample size. These methodological considerations should be taken into account for future studies testing clinical devices for assessing motor function.

Strengths and Limitations

One strength of this review is that it includes more studies than any other review of similar scope that we identified during our study selection process [12-20]. This review is unique relative to other reviews on this same topic because it summarizes the validity estimates across the included studies instead of simply describing the characteristics of the samples and sensors involved [15-20]. This provides an evaluation of the degree of validity produced by such sensors. An additional strength was that we identified several meaningful patterns in this literature (eg, an absence of consistency in analytic approaches, equivalent validity of motion data collected at home or in a laboratory, and higher validity coefficients for healthy adults), which can help guide future research in this area. A final strength of this review is that it addresses statistical issues in this field. Although most reviews in this research area are silent as to statistical concerns, the findings of this review are consistent with the small group of previous reviews, which have also noted the statistical challenges present in this literature [12-14].

A limitation of this review is insufficient statistical power to address several questions of interest because of the methodological inconsistency and resulting sparseness across studies. A second limitation of this review is that the literature showed some signs of potential bias, which could limit the trustworthiness of the aggregate effect sizes. Examples of potential bias identified during the study quality assessment were that few studies provided a clear description of whether data were available for all participants throughout the study, and no studies corrected for potential selection biases in their analyses. In addition, it is unclear whether the patterns seen in the funnel plot and elsewhere are evidence of publication bias, selective outcomes, or an artifact of the dominant analytic approaches in this field. Much of the reviewed literature has focused on clinical validation, with less than one-fourth of the studies performing analytical validation. As important as clinical validation is for establishing the clinical and real-world utility of sensor-derived motion data, more studies are needed that focus on the fundamental step of analytic validation. An additional limitation may be the fact that some diseases are not as prevalent or well-studied than others, which may have impacted their representation in our analyses. Finally, our review was restricted to publications available in the English language. Therefore, some technologies being investigated for motor function assessment in non–English-speaking countries may have been missed.

Considerations for Future Research

Several questions we initially hoped to answer in this review could not be addressed because of lack of consistency across studies (eg, which technology or sensor is used, where the sensor is placed, which motions are required by participants, preprocessing steps, feature detection and selection algorithms, and number of motion features retained for the prediction algorithm). Even within studies examining the same disease state, there was limited consistency in these characteristics. As a result, we cannot say which movements and motion outcomes produce the most valid indicators of different neurological disease states, or what data preprocessing, feature processing engineering, and analysis should be considered best practices for converting raw sensor-derived motion data into meaningful digital measurements or biomarkers. It was notable that many of the most common movements from the larger clinical literature (eg, reaching, sit-to-stand, tracing, and pointing) appeared so infrequently in this literature. This lack of consistency in the literature could have affected the validity estimates [135-139], and the lack of harmonization across studies limits any inference about methodological or analytic decisions [140].

An earlier review described continuous monitoring using movement-detecting wearable sensors as a potential source of ground truth for motor function data, which were previously available only through participant self-reports [141]. On the basis of the reviewed literature, the field cannot yet provide this type of objective truth. An existing algorithm needs to be applied to multiple samples without additional adjustments or enhancements and show an aggregate performance that approximates the estimates provided by the studies included in this review. No analytic technique will solve this issue; the only true solution is replication attempts in new samples. Researchers should report how many of the detected features were moved to feature selection to give readers a sense of how many features

XSL•FO
RenderX

were excluded, a sense of the parsimony of the resultant model, and an awareness of how likely it is that the model may have been overfit. Care must be taken to design the classification algorithm in a way that maximizes the likelihood that it can perform equally well in future samples. This priority needs to be evaluated at each stage of the analysis: data set preparation, preprocessing, feature extraction, algorithm development, model development or validation, and analytical or clinical validation.

Conclusions

In conclusion, sensor-derived motion data can be leveraged to validly predict disease status for a variety of neurological conditions. Future research will elucidate to what extent sensor-derived motion data may yield robust and transformative digital measurements intended to quantify, diagnose, and predict neurological disease state and its longitudinal change.

Acknowledgments

The authors would like to thank Gráinne Donnellan, Sarah Hill, and Ana Howarth of Evidinno Outcomes Research Inc (Vancouver, British Columbia, Canada) for their contributions to conducting this systematic literature review.

Conflicts of Interest

AS, PAC, CM, and SB report employment with Biogen. During completion of the work related to this manuscript, CCG was an employee of Biogen. CCG's current affiliation is ActiGraph, LLC, Pensacola, Florida, US, which was not involved in this work. TS, KH, and MSF report employment with Evidinno Outcomes Research Inc, which was contracted by Biogen to conduct this study.

Multimedia Appendix 1
Supplemental information on the methods and results of the systematic literature review.
[DOCX File , 288 KB - jmir_v24i11e37683_app1.docx]

References

1. Laguna MA, Finat J. Remote monitoring and fall detection: multiplatform Java based mobile applications. In: Proceedings of the 3rd International Workshop on Ambient Assisted Living. 2011 Presented at: IWAAL '11; June 8-10, 2011; Torremolinos-Málaga, Spain p. 1-8. [doi: 10.1007/978-3-642-21303-8_1]
2. Reiman MP, Manske RC. The assessment of function: how is it measured? A clinical perspective. J Man Manip Ther 2011 May;19(2):91-99 [FREE Full text] [doi: 10.1179/106698111X12973307659546] [Medline: 22547919]
3. Toosizadeh N, Mohler J, Lei H, Parvaneh S, Sherman S, Najafi B. Motor performance assessment in Parkinson's disease: association between objective in-clinic, objective in-home, and subjective/semi-objective measures. PLoS One 2015 Apr 24;10(4):e0124763 [FREE Full text] [doi: 10.1371/journal.pone.0124763] [Medline: 25909898]
4. Duff SV, He J, Nelsen MA, Lane CJ, Rowe VT, Wolf SL, et al. Interrater reliability of the Wolf Motor Function Test-Functional Ability Scale: why it matters. Neurorehabil Neural Repair 2015 Jun;29(5):436-443 [FREE Full text] [doi: 10.1177/1545968314553030] [Medline: 25323459]
5. Ko J, Kim M. Reliability and responsiveness of the gross motor function measure-88 in children with cerebral palsy. Phys Ther 2013 Mar;93(3):393-400. [doi: 10.2522/ptj.20110374] [Medline: 23139425]
6. Patel S, Lorincz K, Hughes R, Huggins N, Growdon J, Standaert D, et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. IEEE Trans Inf Technol Biomed 2009 Nov;13(6):864-873 [FREE Full text] [doi: 10.1109/TITB.2009.2033471] [Medline: 19846382]
7. Saadeh W, Altaf MA, Butt SA. A wearable neuro-degenerative diseases detection system based on gait dynamics. In: Proceedings of the 2017 IFIP/IEEE International Conference on Very Large Scale Integration. 2017 Presented at: VLSI-SoC '17; October 23-25, 2017; Abu Dhabi, United Arab Emirates p. 1-6. [doi: 10.1109/vlsi-soc.2017.8203488]
8. Rigas G, Tzallas AT, Tsipouras MG, Bougia P, Tripoliti EE, Baga D, et al. Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors. IEEE Trans Inf Technol Biomed 2012 May;16(3):478-487. [doi: 10.1109/TITB.2011.2182616] [Medline: 22231198]
9. Qualification opinion on stride velocity 95th centile as a secondary endpoint in Duchenne Muscular Dystrophy measured by a valid and suitable wearable device*. European Medicines Agency. 2019 Apr 26. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-secondary-endpoint-duchenne-muscular-dystrophy_en.pdf [accessed 2021-04-09]
10. Aronson JK, Heneghan C, Ferner RE. Medical devices: definition, classification, and regulatory implications. Drug Saf 2020 Feb;43(2):83-93. [doi: 10.1007/s40264-019-00878-3] [Medline: 31845212]
11. H.R.34 - 21st Century Cures Act. US Congress. Washington, DC, USA: Library of Congress; 2016. URL: https://www.congress.gov/bill/114th-congress/house-bill/34 [accessed 2021-04-09]

XSL•FO
RenderX

12. de Magalhaes FA, Vannozzi G, Gatta G, Fantozzi S. Wearable inertial sensors in swimming motion analysis: a systematic review. J Sports Sci 2015;33(7):732-745. [doi: 10.1080/02640414.2014.962574] [Medline: 25356682]
13. Mooney R, Corley G, Godfrey A, Quinlan LR, ÓLaighin G. Inertial sensor technology for elite swimming performance analysis: a systematic review. Sensors (Basel) 2015 Dec 25;16(1):18 [FREE Full text] [doi: 10.3390/s16010018] [Medline: 26712760]
14. Walmsley CP, Williams SA, Grisbrook T, Elliott C, Imms C, Campbell A. Measurement of upper limb range of motion using wearable sensors: a systematic review. Sports Med Open 2018 Nov 29;4(1):53 [FREE Full text] [doi: 10.1186/s40798-018-0167-7] [Medline: 30499058]
15. Baig MM, Afifi S, GholamHosseini H, Mirza F. A systematic review of wearable sensors and IoT-based monitoring applications for older adults - a focus on ageing population and independent living. J Med Syst 2019 Jun 15;43(8):233. [doi: 10.1007/s10916-019-1365-7] [Medline: 31203472]
16. Chen M, Ma Y, Song J, Lai CF, Hu B. Smart clothing: connecting human with clouds and big data for sustainable health monitoring. Mobile Netw Appl 2016 Jul 7;21(5):825-845. [doi: 10.1007/s11036-016-0745-1]
17. Díaz S, Stephenson JB, Labrador MA. Use of wearable sensor technology in gait, balance, and range of motion analysis. Appl Sci 2020;10(1):234. [doi: 10.3390/app10010234]
18. Fong DT, Chan YY. The use of wearable inertial motion sensors in human lower limb biomechanics studies: a systematic review. Sensors (Basel) 2010;10(12):11556-11565 [FREE Full text] [doi: 10.3390/s101211556] [Medline: 22163542]
19. Kekade S, Hseieh CH, Islam MM, Atique S, Mohammed Khalfan A, Li YC, et al. The usefulness and actual use of wearable devices among the elderly population. Comput Methods Programs Biomed 2018 Jan;153:137-159. [doi: 10.1016/j.cmpb.2017.10.008] [Medline: 29157447]
20. Kristoffersson A, Lindén M. A systematic review on the use of wearable body sensors for health monitoring: a qualitative synthesis. Sensors (Basel) 2020 Mar 09;20(5):1502 [FREE Full text] [doi: 10.3390/s20051502] [Medline: 32182907]
21. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020 Apr 14;3:55 [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
22. Manta C, Mahadevan N, Bakker J, Ozen Irmak S, Izmailova E, Park S, et al. EVIDENCE publication checklist for studies evaluating connected sensor technologies: explanation and elaboration. Digit Biomark 2021 May 18;5(2):127-147 [FREE Full text] [doi: 10.1159/000515835] [Medline: 34179682]
23. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1. London, UK: Cochrane Collaboration; 2020.
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. Boukhvalova AK, Kowalczyk E, Harris T, Kosa P, Wichman A, Sandford MA, et al. Identifying and quantifying neurological disability via smartphone. Front Neurol 2018 Sep 4;9:740 [FREE Full text] [doi: 10.3389/fneur.2018.00740] [Medline: 30233487]
26. Montalban X, Mulero P, Midaglia L, Graves J, Hauser SL, Julian L, et al. FLOODLIGHT: smartphone-based self-monitoring is accepted by patients and provides meaningful, continuous digital outcomes augmenting conventional in-clinic multiple sclerosis measures (P3.2-024). Neurology 2019 May 7;92(15 Supplement):P3.2-P024.
27. Abrami A, Heisig S, Ramos V, Thomas KC, Ho BK, Caggiano V. Using an unbiased symbolic movement representation to characterize Parkinson's disease states. Sci Rep 2020 Apr 30;10(1):7377 [FREE Full text] [doi: 10.1038/s41598-020-64181-3] [Medline: 32355166]
28. Adams RJ, Lichter MD, Krepkovich ET, Ellington A, White M, Diamond PT. Assessing upper extremity motor function in practice of virtual activities of daily living. IEEE Trans Neural Syst Rehabil Eng 2015 Mar;23(2):287-296 [FREE Full text] [doi: 10.1109/TNSRE.2014.2360149] [Medline: 25265612]
29. Alberts JL, Hirsch JR, Koop MM, Schindler DD, Kana DE, Linder SM, et al. Using accelerometer and gyroscopic measures to quantify postural stability. J Athl Train 2015 Jun;50(6):578-588 [FREE Full text] [doi: 10.4085/1062-6050-50.2.01] [Medline: 25844853]
30. Brooks C, Shaafi Kabiri N, Mortazavi F, Auerbach S, Bonato P, Erb MK, et al. Variations in rest-activity rhythm are associated with clinically measured disease severity in Parkinson's disease. Chronobiol Int 2020 May;37(5):699-711. [doi: 10.1080/07420528.2020.1715998] [Medline: 31959001]
31. Christiansen C, Moore C, Schenkman M, Kluger B, Kohrt W, Delitto A, et al. Factors associated with ambulatory activity in de novo Parkinson disease. J Neurol Phys Ther 2017 Apr;41(2):93-100 [FREE Full text] [doi: 10.1097/NPT.0000000000000169] [Medline: 28263256]
32. Dehbandi B, Barachant A, Harary D, Long JD, Tsagaris KZ, Bumanlag SJ, et al. Using data from the Microsoft Kinect 2 to quantify upper limb behavior: a feasibility study. IEEE J Biomed Health Inform 2017 Sep;21(5):1386-1392. [doi: 10.1109/JBHI.2016.2606240] [Medline: 28113385]
33. Dowling A, Brys M, Paskavitz J, Auclair V, McLaren D, Postuma R, et al. Quantitative assessment of appendicular bradykinesia in Parkinson's disease using wearable sensors. Mov Disord 2018;33(Supplement 2):S522.

XSL•FO
RenderX

34. Hssayeni MD, Burack MA, Jimenez-Shahed J, Ghoraani B. Assessment of response to medication in individuals with Parkinson's disease. Med Eng Phys 2019 May;67:33-43. [doi: 10.1016/j.medengphy.2019.03.002] [Medline: 30876817]
35. Hughes CM, Baye M, Gordon-Murer C, Louie A, Sun S, Belay GJ, et al. Quantitative assessment of upper limb motor function in Ethiopian acquired brain injured patients using a low-cost wearable sensor. Front Neurol 2019 Dec 12;10:1323 [FREE Full text] [doi: 10.3389/fneur.2019.01323] [Medline: 31920943]
36. Levy IG, Yadav V, Abbas A, Koesmahargyo V, Kalali A. Digital markers of motor activity captured over smartphone is associated with negative symptoms of schizophrenia: results from a pilot observational study. Neuropsychopharmacology 2019 Dec 1;44(Suppl 1):348-349.
37. Ma M, Proffitt R, Skubic M. Validation of a Kinect V2 based rehabilitation game. PLoS One 2018 Aug 24;13(8):e0202338 [FREE Full text] [doi: 10.1371/journal.pone.0202338] [Medline: 30142631]
38. Mitsi G, Mendoza EU, Wissel BD, Barbopoulou E, Dwivedi AK, Tsoulos I, et al. Biometric digital health technology for measuring motor function in Parkinson's disease: results from a feasibility and patient satisfaction study. Front Neurol 2017 Jun 13;8:273 [FREE Full text] [doi: 10.3389/fneur.2017.00273] [Medline: 28659858]
39. Ozinga SJ, Linder SM, Alberts JL. Use of mobile device accelerometry to enhance evaluation of postural instability in Parkinson disease. Arch Phys Med Rehabil 2017 Apr;98(4):649-658 [FREE Full text] [doi: 10.1016/j.apmr.2016.08.479] [Medline: 27670925]
40. Pradhan S, Kelly VE. Quantifying physical activity in early Parkinson disease using a commercial activity monitor. Parkinsonism Relat Disord 2019 Sep;66:171-175 [FREE Full text] [doi: 10.1016/j.parkreldis.2019.08.001] [Medline: 31420310]
41. Taylor-Piliae RE, Mohler MJ, Najafi B, Coull BM. Objective fall risk detection in stroke survivors using wearable sensor technology: a feasibility study. Top Stroke Rehabil 2016 Dec;23(6):393-399. [doi: 10.1179/1074935715Z.00000000059] [Medline: 26382725]
42. Arora S, Baig F, Lo C, Barber TR, Lawton MA, Zhan A, et al. Smartphone motor testing to distinguish idiopathic REM sleep behavior disorder, controls, and PD. Neurology 2018 Oct 16;91(16):e1528-e1538 [FREE Full text] [doi: 10.1212/WNL.0000000000006366] [Medline: 30232246]
43. Bennasar M, Hicks YA, Clinch SP, Jones P, Holt C, Rosser A, et al. Automated assessment of movement impairment in Huntington's disease. IEEE Trans Neural Syst Rehabil Eng 2018 Oct;26(10):2062-2069 [FREE Full text] [doi: 10.1109/TNSRE.2018.2868170] [Medline: 30334742]
44. Coates L, Shi J, Rochester L, Del Din S, Pantall A. Entropy of real-world gait in Parkinson's disease determined from wearable sensors as a digital marker of altered ambulatory behavior. Sensors (Basel) 2020 May 05;20(9):2631 [FREE Full text] [doi: 10.3390/s20092631] [Medline: 32380692]
45. Del Din S, Barber T, Lo C, Rolinski M, Baig F, Ju M, et al. Free-living gait in REM sleep behaviour disorder: measures of prodromal Parkinson's disease? Mov Disord 2018;33(Supplement 2):S524-S525.
46. Del Din S, Elshehabi M, Galna B, Hansen C, Hobert M, Suenkel U, et al. Instrumented gait analysis identifies potential predictors for Parkinson's disease converters. Mov Disord 2018;33(Supplement 2):S505.
47. Hasan H, Burrows M, Athauda DS, Hellman B, James B, Warner T, et al. The BRadykinesia Akinesia INcoordination (BRAIN) tap test: capturing the sequence effect. Mov Disord Clin Pract 2019 Jun 25;6(6):462-469 [FREE Full text] [doi: 10.1002/mdc3.12798] [Medline: 31392247]
48. Kassavetis P, Saifee TA, Roussos G, Drougkas L, Kojovic M, Rothwell JC, et al. Developing a tool for remote digital assessment of Parkinson's disease. Mov Disord Clin Pract 2015 Oct 20;3(1):59-64 [FREE Full text] [doi: 10.1002/mdc3.12239] [Medline: 30363542]
49. Lo C, Arora S, Baig F, Lawton MA, El Mouden C, Barber TR, et al. Predicting motor, cognitive and functional impairment in Parkinson's. Ann Clin Transl Neurol 2019 Aug;6(8):1498-1509 [FREE Full text] [doi: 10.1002/acn3.50853] [Medline: 31402628]
50. Pantall A, Del Din S, Rochester L. Longitudinal changes over thirty-six months in postural control dynamics and cognitive function in people with Parkinson's disease. Gait Posture 2018 May;62:468-474. [doi: 10.1016/j.gaitpost.2018.04.016] [Medline: 29674286]
51. Pantall A, Suresparan P, Kapa L, Morris R, Yarnall A, Del Din S, et al. Postural dynamics are associated with cognitive decline in Parkinson's disease. Front Neurol 2018 Dec 5;9:1044 [FREE Full text] [doi: 10.3389/fneur.2018.01044] [Medline: 30568629]
52. Prince J, Arora S, de Vos M. Big data in Parkinson's disease: using smartphones to remotely detect longitudinal disease phenotypes. Physiol Meas 2018 Apr 26;39(4):044005. [doi: 10.1088/1361-6579/aab512] [Medline: 29516871]
53. van Wamelen DJ, Hota S, Podlewska A, Leta V, Trivedi D, Rizos A, et al. Non-motor correlates of wrist-worn wearable sensor use in Parkinson's disease: an exploratory analysis. NPJ Parkinsons Dis 2019 Oct 2;5:22 [FREE Full text] [doi: 10.1038/s41531-019-0094-4] [Medline: 31602393]
54. Capecci M, Pepa L, Verdini F, Ceravolo MG. A smartphone-based architecture to detect and quantify freezing of gait in Parkinson's disease. Gait Posture 2016 Oct;50:28-33. [doi: 10.1016/j.gaitpost.2016.08.018] [Medline: 27567449]

XSL•FO
RenderX

55. Cavallo F, Moschetti A, Esposito D, Maremmani C, Rovini E. Upper limb motor pre-clinical assessment in Parkinson's disease using machine learning. Parkinsonism Relat Disord 2019 Jun;63:111-116 [FREE Full text] [doi: 10.1016/j.parkreldis.2019.02.028] [Medline: 30826265]
56. Lopane G, Mellone S, Chiari L, Cortelli P, Calandra-Buonaura G, Contin M. Dyskinesia detection and monitoring by a single sensor in patients with Parkinson's disease. Mov Disord 2015 Aug;30(9):1267-1271. [doi: 10.1002/mds.26313] [Medline: 26179817]
57. Scano A, Chiavenna A, Malosio M, Molinari Tosatti L, Molteni F. Kinect V2 implementation and testing of the reaching performance scale for motor evaluation of patients with neurological impairment. Med Eng Phys 2018 Jun;56:54-58. [doi: 10.1016/j.medengphy.2018.04.005] [Medline: 29681441]
58. Vianello A, Chittaro L, Burigat S, Budai R. MotorBrain: a mobile app for the assessment of users' motor performance in neurology. Comput Methods Programs Biomed 2017 May;143:35-47. [doi: 10.1016/j.cmpb.2017.02.012] [Medline: 28391817]
59. Arroyo-Gallego T, Ledesma-Carbayo MJ, Butterworth I, Matarazzo M, Montero-Escribano P, Puertas-Martín V, et al. Detecting motor impairment in early Parkinson's disease via natural typing interaction with keyboards: validation of the neuroQWERTY approach in an uncontrolled at-home setting. J Med Internet Res 2018 Mar 26;20(3):e89 [FREE Full text] [doi: 10.2196/jmir.9462] [Medline: 29581092]
60. Pérez-López C, Samà A, Rodríguez-Martín D, Català A, Cabestany J, Moreno-Arostegui J, et al. Assessing motor fluctuations in Parkinson's disease patients based on a single inertial sensor. Sensors (Basel) 2016 Dec 15;16(12):2132 [FREE Full text] [doi: 10.3390/s16122132] [Medline: 27983675]
61. Rodríguez-Martín D, Pérez-López C, Samà A, Català A, Moreno Arostegui JM, Cabestany J, et al. A waist-worn inertial measurement unit for long-term monitoring of Parkinson's disease patients. Sensors (Basel) 2017 Apr 11;17(4):827 [FREE Full text] [doi: 10.3390/s17040827] [Medline: 28398265]
62. Rodríguez-Molinero A, Samà A, Pérez-Martínez DA, Pérez López C, Romagosa J, Bayés À, et al. Validation of a portable device for mapping motor and gait disturbances in Parkinson's disease. JMIR Mhealth Uhealth 2015 Feb 02;3(1):e9 [FREE Full text] [doi: 10.2196/mhealth.3321] [Medline: 25648406]
63. Chae SH, Kim Y, Lee KS, Park HS. Development and clinical evaluation of a web-based upper limb home rehabilitation system using a smartwatch and machine learning model for chronic stroke survivors: prospective comparative study. JMIR Mhealth Uhealth 2020 Jul 09;8(7):e17216 [FREE Full text] [doi: 10.2196/17216] [Medline: 32480361]
64. Kim HB, Lee HJ, Lee WW, Kim SK, Jeon HS, Park HY, et al. Validation of freezing-of-gait monitoring using smartphone. Telemed J E Health 2018 Nov;24(11):899-907. [doi: 10.1089/tmj.2017.0215] [Medline: 29708870]
65. Lee J, Park S, Shin H. Detection of hemiplegic walking using a wearable inertia sensing device. Sensors (Basel) 2018 May 28;18(6):1736 [FREE Full text] [doi: 10.3390/s18061736] [Medline: 29843413]
66. Seok HY, Kim JW, Kim YH, Park MH, Kwon DY. Quantitative evaluation of hand motor function using a gyrosensor in mild and moderate carpal tunnel syndrome. Muscle Nerve 2019 Apr;59(4):465-469. [doi: 10.1002/mus.26424] [Medline: 30677150]
67. Louter M, Maetzler W, Prinzen J, van Lummel RC, Hobert M, Arends JB, et al. Accelerometer-based quantitative analysis of axial nocturnal movements differentiates patients with Parkinson's disease, but not high-risk individuals, from controls. J Neurol Neurosurg Psychiatry 2015 Jan;86(1):32-37. [doi: 10.1136/jnnp-2013-306851] [Medline: 24777169]
68. Ossig C, Gandor F, Fauser M, Bosredon C, Churilov L, Reichmann H, et al. Correlation of quantitative motor state assessment using a kinetograph and patient diaries in advanced PD: data from an observational study. PLoS One 2016 Aug 24;11(8):e0161559 [FREE Full text] [doi: 10.1371/journal.pone.0161559] [Medline: 27556806]
69. Otte K, Kayser B, Mansow-Model S, Verrel J, Paul F, Brandt AU, et al. Accuracy and reliability of the Kinect Version 2 for clinical measurement of motor function. PLoS One 2016 Nov 18;11(11):e0166532 [FREE Full text] [doi: 10.1371/journal.pone.0166532] [Medline: 27861541]
70. Horigome T, Sumali B, Kitazawa M, Yoshimura M, Liang KC, Tazawa Y, et al. Evaluating the severity of depressive symptoms using upper body motion captured by RGB-depth sensors and machine learning in a clinical interview setting: a preliminary study. Compr Psychiatry 2020 Feb 20;98:152169 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152169] [Medline: 32145559]
71. Terashi H, Taguchi T, Ueta Y, Okubo Y, Mitoma H, Aizawa H. Analysis of non-invasive gait recording under free-living conditions in patients with Parkinson's disease: relationship with global cognitive function and motor abnormalities. BMC Neurol 2020 Apr 29;20(1):161 [FREE Full text] [doi: 10.1186/s12883-020-01729-w] [Medline: 32349688]
72. Terui Y, Iwakura M, Suto E, Kawagoshi A, Sugawara K, Takahashi H, et al. New evaluation of trunk movement and balance during walking in COPD patients by a triaxial accelerometer. Int J Chron Obstruct Pulmon Dis 2018 Dec 7;13:3957-3962 [FREE Full text] [doi: 10.2147/COPD.S184212] [Medline: 30584295]
73. Beange KH, Chan AD, Beaudette SM, Graham RB. Concurrent validity of a wearable IMU for objective assessments of functional movement quality and control of the lumbar spine. J Biomech 2019 Dec 03;97:109356. [doi: 10.1016/j.jbiomech.2019.109356] [Medline: 31668717]

XSL•FO
RenderX

74. Campos C, DePaul VG, Knorr S, Wong JS, Mansfield A, Patterson KK. Validity of the ActiGraph activity monitor for individuals who walk slowly post-stroke. Top Stroke Rehabil 2018 May;25(4):295-304. [doi: 10.1080/10749357.2018.1446487] [Medline: 29557275]
75. Evers LJ, Raykov YP, Krijthe JH, Silva de Lima AL, Badawy R, Claes K, et al. Real-life gait performance as a digital biomarker for motor fluctuations: the Parkinson@home validation study. J Med Internet Res 2020 Oct 09;22(10):e19068 [FREE Full text] [doi: 10.2196/19068] [Medline: 33034562]
76. Zach H, Dirkx M, Pasman JW, Bloem BR, Helmich RC. The patient's perspective: the effect of levodopa on Parkinson symptoms. Parkinsonism Relat Disord 2017 Feb;35:48-54. [doi: 10.1016/j.parkreldis.2016.11.015] [Medline: 27919585]
77. do Carmo Vilas-Boas M, Choupina HM, Rocha AP, Fernandes JM, Cunha JP. Full-body motion assessment: concurrent validation of two body tracking depth sensors versus a gold standard system during gait. J Biomech 2019 Apr 18;87:189-196. [doi: 10.1016/j.jbiomech.2019.03.008] [Medline: 30914189]
78. do Carmo Vilas-Boas M, Rocha AP, Choupina HM, Cardoso MN, Fernandes JM, Coelho T, et al. Validation of a single RGB-D camera for gait assessment of polyneuropathy patients. Sensors (Basel) 2019 Nov 12;19(22):4929 [FREE Full text] [doi: 10.3390/s19224929] [Medline: 31726742]
79. Aghanavesi S, Westin J, Bergquist F, Nyholm D, Askmark H, Aquilonius SM, et al. A multiple motion sensors index for motor state quantification in Parkinson's disease. Comput Methods Programs Biomed 2020 Jun;189:105309. [doi: 10.1016/j.cmpb.2019.105309] [Medline: 31982667]
80. Aghanavesi S, Bergquist F, Nyholm D, Senek M, Memedi M. Motion sensor-based assessment of Parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge. IEEE J Biomed Health Inform 2020 Jan;24(1):111-119. [doi: 10.1109/JBHI.2019.2898332] [Medline: 30763248]
81. Hsu CY, Tsai YS, Yau CS, Shie HH, Wu CM. Gait and trunk movement characteristics of chronic ischemic stroke patients. Int J Gerontol 2019;13(2):144-148.
82. Huang YP, Singh A, Chen S, Sun FJ, Huang CR, Liu SI. Validity of a novel touch screen tablet-based assessment for mild cognitive impairment and probable AD in older adults. Assessment 2019 Dec;26(8):1540-1553. [doi: 10.1177/1073191117748395] [Medline: 29251514]
83. Lee W, Evans A, Williams DR. Validation of a smartphone application measuring motor function in Parkinson's disease. J Parkinsons Dis 2016 Apr 02;6(2):371-382. [doi: 10.3233/JPD-150708] [Medline: 27061062]
84. de Paula JN, de Mello Monteiro CB, da Silva TD, Capelini CM, de Menezes LD, Massetti T, et al. Motor performance of individuals with cerebral palsy in a virtual game using a mobile phone. Disabil Rehabil Assist Technol 2018 Aug;13(6):609-613. [doi: 10.1080/17483107.2017.1392620] [Medline: 29092683]
85. Simonsen D, Nielsen IF, Spaich EG, Andersen OK. Design and test of an automated version of the modified Jebsen test of hand function using Microsoft Kinect. J Neuroeng Rehabil 2017 May 02;14(1):38 [FREE Full text] [doi: 10.1186/s12984-017-0250-1] [Medline: 28464927]
86. Lepetit K, Ben Mansour K, Boudaoud S, Kinugawa-Bourron K, Marin F. Evaluation of the kinetic energy of the torso by magneto-inertial measurement unit during the sit-to-stand movement. J Biomech 2018 Jan 23;67:172-176. [doi: 10.1016/j.jbiomech.2017.11.028] [Medline: 29269002]
87. Bernad-Elazari H, Herman T, Mirelman A, Gazit E, Giladi N, Hausdorff JM. Objective characterization of daily living transitions in patients with Parkinson's disease using a single body-fixed sensor. J Neurol 2016 Aug;263(8):1544-1551. [doi: 10.1007/s00415-016-8164-6] [Medline: 27216626]
88. Iakovakis D, Hadjidimitriou S, Charisis V, Bostantzopoulou S, Katsarou Z, Hadjileontiadis LJ. Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage Parkinson's disease. Sci Rep 2018 May 16;8(1):7663 [FREE Full text] [doi: 10.1038/s41598-018-25999-0] [Medline: 29769594]
89. Iakovakis D, Diniz JA, Trivedi D, Chaudhuri RK, Hadjileontiadis LJ, Hadjidimitriou S, et al. Early Parkinson's disease detection via touchscreen typing analysis using convolutional neural networks. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3535-3538. [doi: 10.1109/EMBC.2019.8857211] [Medline: 31946641]
90. Lauraitis A, Maskeliūnas R, Damaševičius R, Krilavičius T. A mobile application for smart computer-aided self-administered testing of cognition, speech, and motor impairment. Sensors (Basel) 2020 Jun 06;20(11):3236 [FREE Full text] [doi: 10.3390/s20113236] [Medline: 32517223]
91. Hiorth YH, Larsen JP, Lode K, Tysnes OB, Godfrey A, Lord S, et al. Impact of falls on physical activity in people with Parkinson's disease. J Parkinsons Dis 2016;6(1):175-182. [doi: 10.3233/JPD-150640] [Medline: 26639446]
92. Mastoras RE, Iakovakis D, Hadjidimitriou S, Charisis V, Kassie S, Alsaadi T, et al. Touchscreen typing pattern analysis for remote detection of the depressive tendency. Sci Rep 2019 Sep 16;9(1):13414 [FREE Full text] [doi: 10.1038/s41598-019-50002-9] [Medline: 31527640]
93. Ahlrichs C, Samà A, Lawo M, Cabestany J, Rodríguez-Martín D, Pérez-López C, et al. Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients. Med Biol Eng Comput 2016 Jan;54(1):223-233. [doi: 10.1007/s11517-015-1395-3] [Medline: 26429349]
94. Klingelhoefer L, Bostanjopoulou S, Trivedi D, Hadjidimitriou S, Mayer S, Katsarou Z, et al. Medical evaluation as gold standard to control iPrognosis application derived data for early Parkinson’s disease detection. Mov Disord 2019;34(Supplement 2):S913-S914.

XSL•FO
RenderX

95. Lalvay L, Lara M, Mora A, Alarcón F, Fraga M, Pancorbo J, et al. Quantitative measurement of akinesia in Parkinson's disease. Mov Disord Clin Pract 2016 Aug 3;4(3):316-322 [FREE Full text] [doi: 10.1002/mdc3.12410] [Medline: 30363442]
96. Lipsmeier F, Simillion C, Bamdadian A, Smith A, Schobel S, Czech C, et al. Preliminary reliability and validity of a novel digital biomarker smartphone application to assess cognitive and motor symptoms in Huntington's disease (HD) (P1.8-042). Neurology 2019 Apr;92(15 Supplement):P1.8-P042.
97. Lipsmeier F, Taylor KI, Postuma RB. Preliminary validation smartphone application to assess motor symptoms in recently diagnosed parkinson patients. Neurology 2019 May 8;92(15 Supplement):P4.7-005.
98. Rodríguez-Molinero A, Samà A, Pérez-López C, Rodríguez-Martín D, Quinlan LR, Alcaine S, et al. Analysis of correlation between an accelerometer-based algorithm for detecting Parkinsonian gait and UPDRS subscales. Front Neurol 2017 Sep 1;8:431 [FREE Full text] [doi: 10.3389/fneur.2017.00431] [Medline: 28919877]
99. Andrzejewski KL, Dowling AV, Stamler D, Felong TJ, Harris DA, Wong C, et al. Wearable sensors in Huntington disease: a pilot study. J Huntingtons Dis 2016 Jun 18;5(2):199-206. [doi: 10.3233/JHD-160197] [Medline: 27341134]
100. Horne MK, McGregor S, Bergquist F. An objective fluctuation score for Parkinson's disease. PLoS One 2015 Apr 30;10(4):e0124522 [FREE Full text] [doi: 10.1371/journal.pone.0124522] [Medline: 25928634]
101. Wissel BD, Mitsi G, Dwivedi AK, Papapetropoulos S, Larkin S, López Castellanos JR, et al. Tablet-based application for objective measurement of motor fluctuations in Parkinson disease. Digit Biomark 2018 Jan 9;1(2):126-135 [FREE Full text] [doi: 10.1159/000485468] [Medline: 32095754]
102. Shaafi Kabiri N, Best A, Johnson S, Ho B, Eden G, Dupee B, et al. Utility of a smartphone app in discretely assessing and monitoring symptoms of Parkinson's disease. Mov Disord 2019;34(Supplement 2):S298.
103. Arroyo-Gallego T, Ledesma-Carbayo MJ, Sanchez-Ferro A, Butterworth I, Mendoza CS, Matarazzo M, et al. Detection of motor impairment in Parkinson's disease via mobile touchscreen typing. IEEE Trans Biomed Eng 2017 Sep;64(9):1994-2002. [doi: 10.1109/TBME.2017.2664802] [Medline: 28237917]
104. Lowes L, Miller N, Iammarino M, Dugan M, Alfano L. Activity monitoring in neuromuscular disease: successes, challenges, and a path forward (P5.6-016). Neurology 2019 Apr;92(15 Supplement):P5.6-P016.
105. Zhou H. Instrumented trail-making task to identify cognitive-motor impairment and assess cognitive frailty. J Am Geriatr Soc 2019 Apr 02;67(S1):S127-S128. [doi: 10.1111/jgs.15898]
106. Ren P, Bosch Bayard JF, Dong L, Chen J, Mao L, Ma D, et al. Multivariate analysis of joint motion data by Kinect: application to Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng 2020 Jan;28(1):181-190. [doi: 10.1109/TNSRE.2019.2953707] [Medline: 31751278]
107. Gordon MF, Grachev ID, Mazeh I, Dolan Y, Reilmann R, Loupe PS, et al. Quantification of motor function in Huntington disease patients using wearable sensor devices. Digit Biomark 2019 Sep 6;3(3):103-115 [FREE Full text] [doi: 10.1159/000502136] [Medline: 32095771]
108. Shawen N, Knott J, Lee J, Simuni T, Daeschler D, Jayaraman A. Automatic scoring of Parkinson's disease motor symptoms using a smartwatch (P2.8-001). Neurology 2019 Apr;92(15 Supplement):P2.8-P001.
109. Seiffert P, Kawa J, Derejczyk J, Czernek M, Stepien P, Marcisz C. Mobile test of manual dexterity in the diagnostics of frailty in older patients with mild parkinsonian signs. Eur Geriatr Med 2019;10(Supplement 1):S258. [doi: 10.1007/s41999-019-00221-0]
110. Lee CY, Kang SJ, Hong SK, Ma HI, Lee U, Kim YJ. A validation study of a smartphone-based finger tapping application for quantitative assessment of bradykinesia in Parkinson's disease. PLoS One 2016 Jul 28;11(7):e0158852 [FREE Full text] [doi: 10.1371/journal.pone.0158852] [Medline: 27467066]
111. Kim H, Lee HJ, Lee W, Kwon S, Kim SK, Jeon HS, et al. Unconstrained detection of freezing of Gait in Parkinson's disease patients using smartphone. Annu Int Conf IEEE Eng Med Biol Soc 2015 Aug;2015:3751-3754. [doi: 10.1109/EMBC.2015.7319209] [Medline: 26737109]
112. Sigcha L, Costa N, Pavón I, Costa S, Arezes P, López JM, et al. Deep learning approaches for detecting freezing of gait in Parkinson's disease patients through on-body acceleration sensors. Sensors (Basel) 2020 Mar 29;20(7):1895 [FREE Full text] [doi: 10.3390/s20071895] [Medline: 32235373]
113. Bonnechère B, Jansen B, Haack I, Omelina L, Feipel V, Van Sint Jan S, et al. Automated functional upper limb evaluation of patients with Friedreich ataxia using serious games rehabilitation exercises. J Neuroeng Rehabil 2018 Oct 04;15(1):87 [FREE Full text] [doi: 10.1186/s12984-018-0430-7] [Medline: 30286776]
114. Hssayeni MD, Adams JL, Ghoraani B. Deep learning for medication assessment of individuals with Parkinson's disease using wearable sensors. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:1-4. [doi: 10.1109/EMBC.2018.8513344] [Medline: 30440318]
115. Zhan A, Mohan S, Tarolli C, Schneider RB, Adams JL, Sharma S, et al. Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score. JAMA Neurol 2018 Jul 01;75(7):876-880 [FREE Full text] [doi: 10.1001/jamaneurol.2018.0809] [Medline: 29582075]
116. ROBINS-E Development Group, Higgins J, Morgan R, Rooney A, Taylor K, Thayer K, et al. Risk Of Bias In Non-randomized Studies - of Exposure (ROBINS-E). Launch Version. Risk Of Bias. 2022. URL: https://www.riskofbias.info/welcome/robins-e-tool [accessed 2022-06-01]

XSL•FO
RenderX

117. Rosenthal R. Meta-Analytic Procedures for Social Research. Applied Social Research Methods Series 6. Newbury Park, CA, USA: Sage Publications; 1991.
118. Cooper H, Hedges LV, Valentine JC. The Handbook of Research Synthesis and Meta-Analysis. 3rd edition. New York, NY, USA: Russell Sage Foundation; 2019.
119. Fisher RA. On the 'probable error' of a coefficient of correlation deduced from a small sample. The University of Adelaide. 1921. URL: https://digital.library.adelaide.edu.au/dspace/handle/2440/15169 [accessed 2021-04-09]
120. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 1915 May;10(4):507-521. [doi: 10.2307/2331838]
121. Lipsey M, Wilson D. Practical Meta-Analysis. Thousand Oaks, CA, USA: Sage Publications; 2001.
122. Alexander S, Peryer G, Gray E, Barkhof F, Chataway J. Wearable technologies to measure clinical outcomes in multiple sclerosis: a scoping review. Mult Scler 2021 Oct;27(11):1643-1656 [FREE Full text] [doi: 10.1177/1352458520946005] [Medline: 32749928]
123. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd edition. Hoboken, NJ, USA: Wiley Online Library; 2000.
124. Lee EH, Lee SJ, Hwang ST, Hong SH, Kim JH. Reliability and validity of the beck depression inventory-II among Korean adolescents. Psychiatry Investig 2017 Jan;14(1):30-36 [FREE Full text] [doi: 10.4306/pi.2017.14.1.30] [Medline: 28096872]
125. Marrie RA, Goldman M. Validity of performance scales for disability assessment in multiple sclerosis. Mult Scler 2007 Nov;13(9):1176-1182. [doi: 10.1177/1352458507078388] [Medline: 17623733]
126. Martinez-Martin P, Rodriguez-Blazquez C, Alvarez-Sanchez M, Arakaki T, Bergareche-Yarza A, Chade A, et al. Expanded and independent validation of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). J Neurol 2013 Jan;260(1):228-236. [doi: 10.1007/s00415-012-6624-1] [Medline: 22865238]
127. Reich SG, Savitt JM. Parkinson's disease. Med Clin North Am 2019 Mar;103(2):337-350. [doi: 10.1016/j.mcna.2018.10.014] [Medline: 30704685]
128. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012 Jan 13;90(1):7-24 [FREE Full text] [doi: 10.1016/j.ajhg.2011.11.029] [Medline: 22243964]
129. Bottou L, Peters J, Quiñonero-Candela J, Charles DX, Chickering M, Portugaly E, et al. Counterfactual reasoning and learning systems: the example of computational advertising. J Mach Learn Res 2013;14(101):3207-3260.
130. Fiedler K. Voodoo correlations are everywhere-not only in neuroscience. Perspect Psychol Sci 2011 Mar;6(2):163-171. [doi: 10.1177/1745691611400237] [Medline: 26162135]
131. Vul E, Harris C, Winkielman P, Pashler H. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspect Psychol Sci 2009 May;4(3):274-290. [doi: 10.1111/j.1745-6924.2009.01125.x] [Medline: 26158964]
132. Sculley D, Holt G, Golovin D, Davydov E, Phillips T, Ebner D, et al. Machine learning: the high-interest credit card of technical debt. In: Proceedings of the 28th Conference on Neural Information Processing Systems. 2014 Presented at: NeurIPS '14; December 8-13, 2014; Montreal, Canada.
133. Sculley D, Otey ME, Pohl M, Spitznagel B, Hainsworth J, Zhou Y. Detecting adversarial advertisements in the wild. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2011 Presented at: KDD '11; August 21-24, 2011; San Diego, CA, USA p. 274-282. [doi: 10.1145/2020408.2020455]
134. Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. Gigascience 2017 May 01;6(5):1-9 [FREE Full text] [doi: 10.1093/gigascience/gix019] [Medline: 28327985]
135. Vul E, Pashler H. Suspiciously high correlations in brain imaging research. In: Lilienfeld SO, Waldman ID, editors. Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions. Hoboken, NJ, USA: John Wiley & Sons; 2017:196-220.
136. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLoS Biol 2015 Mar;13(3):e1002106 [FREE Full text] [doi: 10.1371/journal.pbio.1002106] [Medline: 25768323]
137. Ioannidis JP. What have we (not) learnt from millions of scientific papers with P values? Am Stat 2019 Mar 20;73(sup1):20-25. [doi: 10.1080/00031305.2018.1447512]
138. Mayo-Wilson E, Li T, Fusco N, Bertizzolo L, Canner JK, Cowley T, et al. Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. J Clin Epidemiol 2017 Nov;91:95-110 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.07.014] [Medline: 28842290]
139. Li T, Mayo-Wilson E, Fusco N, Hong H, Dickersin K. Caveat emptor: the combined effects of multiplicity and selective reporting. Trials 2018 Sep 17;19(1):497 [FREE Full text] [doi: 10.1186/s13063-018-2888-9] [Medline: 30223876]
140. Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K, MUDS investigators. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. J Clin Epidemiol 2017 Jun;86:39-50 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.05.007] [Medline: 28529187]
141. Dobkin BH. Wearable motion sensors to continuously measure real-world physical activities. Curr Opin Neurol 2013 Dec;26(6):602-608 [FREE Full text] [doi: 10.1097/WCO.0000000000000026] [Medline: 24136126]

Abbreviations

AUC: area under the curve

XSL•FO
RenderX

EMA: European Medicines Agency
PICO: Population, Intervention, Comparator, Outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROBINS-E: Risk Of Bias In Nonrandomized Studies of Exposures

Edited by R Kukafka; submitted 03.03.22; peer-reviewed by C Guinemer, K Radhakrishnan, M Bardus; comments to author 09.06.22; revised version received 18.07.22; accepted 11.10.22; published 21.11.22.

Please cite as:
Guo CC, Chiesa PA, de Moor C, Fazeli MS, Schofield T, Hofer K, Belachew S, Scotland A
Digital Devices for Assessing Motor Functions in Mobility-Impaired and Healthy Populations: Systematic Literature Review
J Med Internet Res 2022;24(11):e37683
URL: https://www.jmir.org/2022/11/e37683
doi:10.2196/37683
PMID:36409538

©Christine C Guo, Patrizia Andrea Chiesa, Carl de Moor, Mir Sohail Fazeli, Thomas Schofield, Kimberly Hofer, Shibeshih Belachew, Alf Scotland. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Views, Use, and Experiences of Web-Based Access to Pediatric Electronic Health Records for Children, Adolescents, and Parents: Scoping Review

Josefin Hagström[1], MSc; Charlotte Blease[2], PhD; Barbara Haage[3], MSc; Isabella Scandurra[4], PhD; Scharlett Hansson[3], MSc; Maria Hägglund[1], PhD

[1]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[2]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

[3]Department of Health Technologies, Tallinn University of Technology, Tallinn, Estonia

[4]Informatics, Örebro University School of Business, Örebro, Sweden

Corresponding Author:
Josefin Hagström, MSc
Department of Women's and Children's Health
Uppsala University
Dag Hammarskjölds väg 14B, 1st Floor
Uppsala, 75237
Sweden
Phone: 46 734697474
Email: josefin.hagstrom@kbh.uu.se

Abstract

Background: Ongoing efforts worldwide to provide patients with patient-accessible electronic health records (PAEHRs) have led to variability in adolescent and parental access across providers, regions, and countries. There is no compilation of evidence to guide policy decisions in matters such as access age and the extent of parent proxy access. In this paper, we outline our scoping review of different stakeholders' (including but not limited to end users) views, use, and experiences pertaining to web-based access to electronic health records (EHRs) by children, adolescents, and parents.

Objective: The aim of this study was to identify, categorize, and summarize knowledge about different stakeholders' (eg, children and adolescents, parents, health care professionals [HCPs], policy makers, and designers of patient portals or PAEHRs) views, use, and experiences of EHR access for children, adolescents, and parents.

Methods: A scoping review was conducted according to the Arksey and O'Malley framework. A literature search identified eligible papers that focused on EHR access for children, adolescents, and parents that were published between 2007 and 2021. A number of databases were used to search for literature (PubMed, CINAHL, and PsycINFO).

Results: The approach resulted in 4817 identified articles and 74 (1.54%) included articles. The papers were predominantly viewpoints based in the United States, and the number of studies on parents was larger than that on adolescents and HCPs combined. First, adolescents and parents without access anticipated low literacy and confidentiality issues; however, adolescents and parents who had accessed their records did not report such concerns. Second, the main issue for HCPs was maintaining adolescent confidentiality. This remained an issue after using PAEHRs for parents, HCPs, and other stakeholders but was not an experienced issue for adolescents. Third, the viewpoints of other stakeholders provided a number of suggestions to mitigate issues. Finally, education is needed for adolescents, parents, and HCPs.

Conclusions: There is limited research on pediatric PAEHRs, particularly outside the United States, and on adolescents' experiences with web-based access to their records. These findings could inform the design and implementation of future regulations regarding access to PAEHRs. Further examination is warranted on the experiences of adolescents, parents, and HCPs to improve usability and utility, inform universal principles reducing the current arbitrariness in the child's age for own and parental access to EHRs among providers worldwide, and ensure that portals are equipped to safely and appropriately manage a wide variety of patient circumstances.

International Registered Report Identifier (IRRID): RR2-10.2196/36158

XSL•FO
RenderX

(J Med Internet Res 2022;24(11):e40328) doi:10.2196/40328

KEYWORDS

electronic health record; patient-accessible electronic health record; adolescents; parents; children; patient experience; patient portal; electronic portal; review; scoping review; youth; patient perspective; user experience; patient access; mobile phone

Introduction

Background

Patients being enabled to read their health records on the web is a growing phenomenon. Patient-accessible electronic health records (PAEHRs) commonly include clinical information (eg, physician visit notes, laboratory test results, medications, diagnoses, and referrals), and enabling patients to access their electronic health records (EHRs) is thought to promote patient empowerment by involving patients in their own care [1]. The term *open notes* is often used to describe the specific practice of giving patients access to the free-text entries written by clinicians [2] and is considered an important part of any PAEHR. The websites that host PAEHRs, commonly developed by so-called EHR vendors, are often referred to as patient portals and, for the purposes of this study, *patient portals* will refer to tethered, secure websites that hold any type of health information recorded by a health care provider that users have access to. Today, health institutions in >15 countries are developing patient portals [3], and there is continuous adaptation of legal frameworks at a national level to improve use and ensure patients' privacy [3,4].

An often-cited challenge to PAEHR implementation concerns how to manage access for parents, children, and adolescents [3]. The transfer of proxy access being managed by the parent or guardian (hereafter referred to as *parents*) into own access for the child is often conducted during adolescence, with the aim of protecting the adolescent's privacy as well as to support the transition to adulthood. The need for protection arises as the individual begins seeking care for sensitive medical conditions such as mental health or reproductivity. The child's need for autonomy in their relationship with their health care professional (HCP) is compromised during shared access with parents. So far, providers and countries have approached this dilemma in different ways. For example, the access age of the child varies, as well as when parents lose access and the age when patients gain self-access to their records. In some countries (eg, Finland and Estonia), parents are provided access (unless actively restricted by the child), whereas in other countries (eg, Sweden and Norway), parents are blocked from accessing records by law when their children reach a certain age threshold [5]. In France, adolescents receive access at the age of 18 years, when, in turn, the parents lose access. Decisions regarding earlier access in France can also depend on the perceived maturity of the minor. In many countries and regions, a lack of continuity in access to care is apparent [3]. In the United States, policies regarding age and privacy exceptions are dependent on state laws, which vary throughout the country. In 2021, the 21st Century Cures Act made it mandatory for health care providers to provide every patient with free electronic access to their clinical notes [6]. There is a possibility for withholding confidential information; however, questions still remain [7]. Evidently, the current lack of an international consensus on regulations for EHR access for parents, children, and adolescents has led to great variability.

The research of views, use, and experiences of PAEHRs to date has focused on HCPs and patients of the general adult population. The effects of PAEHRs are not conclusive, yet they indicate benefits including improved medication adherence and self-care as well as improved relationships between patients and their physicians [8-10]. However, a growing yet scarce body of literature is exploring access to EHRs for parents, children, and adolescents in particular. Although parents appreciate having access to their child's records into adolescence [11], shared access to PAEHRs for parents and adolescents runs the risk of causing ethical dilemmas for HCPs. For example, some health information may be considered sensitive by adolescents, such as health care data pertaining to the disclosure of alcohol or drug abuse, sexual activity, or stigmatized illnesses such as anxiety or depression. Adolescents have also been observed to withhold information from HCPs if they are uncertain about who may access it [12,13]. With regard to adolescents' self-access, it is thought that EHR access offers information transparency that might contribute to patient empowerment and enhanced health care; however, evidence suggests that the adolescent population requires targeted analysis. To date, one systematic review [14] has examined patient portals among pediatric patients. The review included only parents and adolescents and focused on empirical studies, and 10 of the 11 studies were based in the United States. Mostly positive feedback was found; however, there was some concern about medical literacy and its effects on the communication between adolescents, parents, and HCPs.

Study Objectives

The objective was to identify, categorize, and summarize knowledge about different stakeholders' (eg, children and adolescents, parents, HCPs, policy makers, and patient portal designers) views, use, and experiences of PAEHR access for children, adolescents, and parents. The findings will aid policy makers in designing future regulations regarding EHR access for parents and adolescents and potentially improve the design and implementation of PAEHRs to meet the needs of the end users. The concept "view" refers to attitudes, expectations, and thoughts; "use" refers to portal feature use and use rates; and "experience" includes experiences pertaining to, for example, satisfaction, concerns, and literacy. We use the definition of Davis et al [15] for a scoping review—"a synthesis and analysis of a wide range of research and nonresearch material to provide greater conceptual clarity about a specific topic or field of evidence"—with the adjustment of not including nonresearch material because of restrictions of the study search strategy. We defined *policy maker* as an agent with capacity or responsibility for deciding policies on PAEHRs (either national, regional, institutional, or as an HCP). The following research question

XSL•FO
RenderX

was examined in detail: how do different stakeholders experience children's, adolescents', and parents' web-based access to the EHRs of children and adolescents? With regard to experiences of HCPs and HCP experts (among other stakeholders) who document in the records or manage the records within their professions, we focused on how these individuals perceive or are affected by the situation where children, adolescents, and parents have access to the EHRs of children and adolescents.

Methods

Scoping Review Approach

The full protocol for this review has been published previously [16]. To summarize, a literature search on PAEHRs for children, adolescents, and parents was conducted using the Arksey and O'Malley [17] framework. The framework includes six stages: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; (5) collating, summarizing, and reporting the results; and (6) consulting with relevant stakeholders. To ensure reproducibility and traceability, the scoping review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for Scoping Reviews (PRISMA-ScR) checklist to report our results (Multimedia Appendix 1) [18].

Stage 1: Identifying the Research Question

Our research question was as follows: how do different stakeholders experience children's, adolescents', and parents' web-based access to the EHRs of children and adolescents?

Stage 2: Identifying Relevant Studies

A comprehensive literature search was conducted on June 23, 2021, by an experienced research librarian at Uppsala University, who provided the research team with the results immediately after conducting the search. The search included the following electronic literature databases: PubMed, CINAHL, and PsycINFO. The search included peer-reviewed literature published between 2005 and September 2021, where the year 2005 was chosen as a cutoff as we expected to not identify any relevant publications on pediatric PAEHRs before this. Search terms were identified using input from the research team and the literature. The references of the identified articles were scanned backward to identify prior work to consider for the research topic. The search query with Boolean operators was presented in the published protocol [16].

Stage 3: Selecting Eligible Studies

The inclusion and exclusion criteria were informed by the review process and were applied at the study selection stage.

Inclusion Criteria

Studies were included if they met the following criteria: (1) the patient user population was children, adolescents, and parents; (2) the population studied was children, adolescents, parents, HCPs, and other stakeholders; (3) outcomes were views, use, or experiences of access or proxy access to PAEHRs; and (4) the study design was all study types.

We defined patients aged ≤12 years as children, patients aged 13 to 18 years as adolescents, and those aged ≥18 years as adults. However, to increase the number of eligible studies for the adolescent population, the age of 19 to 20 years was included if a study participant group included a majority of adolescents (eg, aged 15-19 years).

Exclusion Criteria

Studies were excluded if they (1) were not written in English, (2) were published outside the study period, or (3) did not focus on pediatric PAEHRs.

Search Strategy

The search results were imported into the software program Rayyan (Rayyan Systems Inc) [19] according to the following headings: publication type, publication year, country, sample characteristics, setting, study aim, research question, and conclusions. Duplicates were removed. Titles and abstracts were screened by the authors with consideration of the eligibility criteria. The articles were divided between the investigators (excluding IS) so that each article was screened by at least 2 people. Any disagreements were resolved through group discussion and, if needed, with the addition of a third reader.

Stage 4: Charting the Data

The first author set up a Microsoft Excel spreadsheet to which all researchers added information independently, including the following study characteristics: reference ID, type of identification, title, authors, year, journal, type of publication, study design, participant description, country, treatment setting, clinical field, research question, and main conclusions. The first author held the main responsibility for verifying the accuracy of the data (Multimedia Appendix 2 [11,20-92]). If the abstract and title were insufficient for assessment, the full text was screened. Multiple authors could provide an assessment of the same paper, and instances of disagreement were resolved through discussion. In the second stage, full-text papers were evenly assigned to 2 authors. Instances of disagreement were resolved through discussion and sometimes by bringing in a third reader. The ideas that emerged during the process were discussed among the authors in regular meetings set up by the main author.

Stage 5: Collating, Summarizing, and Reporting the Results

The results reported in the included studies were compiled and read multiple times. In the Microsoft Excel spreadsheet, papers were categorized according to the stakeholder group studied: children and adolescents, parents, HCPs, or expert viewpoints. In total, 2 students categorized the viewpoints into three groups: (1) experts, such as HCPs, IT experts, or researchers; (2) policy makers; and (3) public opinion. In a meeting, 2 authors were assigned to each stakeholder group through discussion, where the first of the following authors listed was mainly responsible: children and adolescents were assigned to JH and BH, parents were assigned to MH and SH, HCPs were assigned to CB and IS, and viewpoints were assigned to JH and MH (as first and senior author, respectively). The results from the included studies were then independently analyzed and jointly drafted in a shared Google Docs. For organization of the results, key

XSL•FO
RenderX

themes were adapted from a previous scoping review of the literature on PAEHRs in mental health [93]. These were refined by the main author using thematic analysis [94]. During this process, the material was gathered according to themes, and themes were reviewed and defined. This synthetization of results was conducted primarily by the main author but was discussed in research team meetings.

Stage 6: Consultation

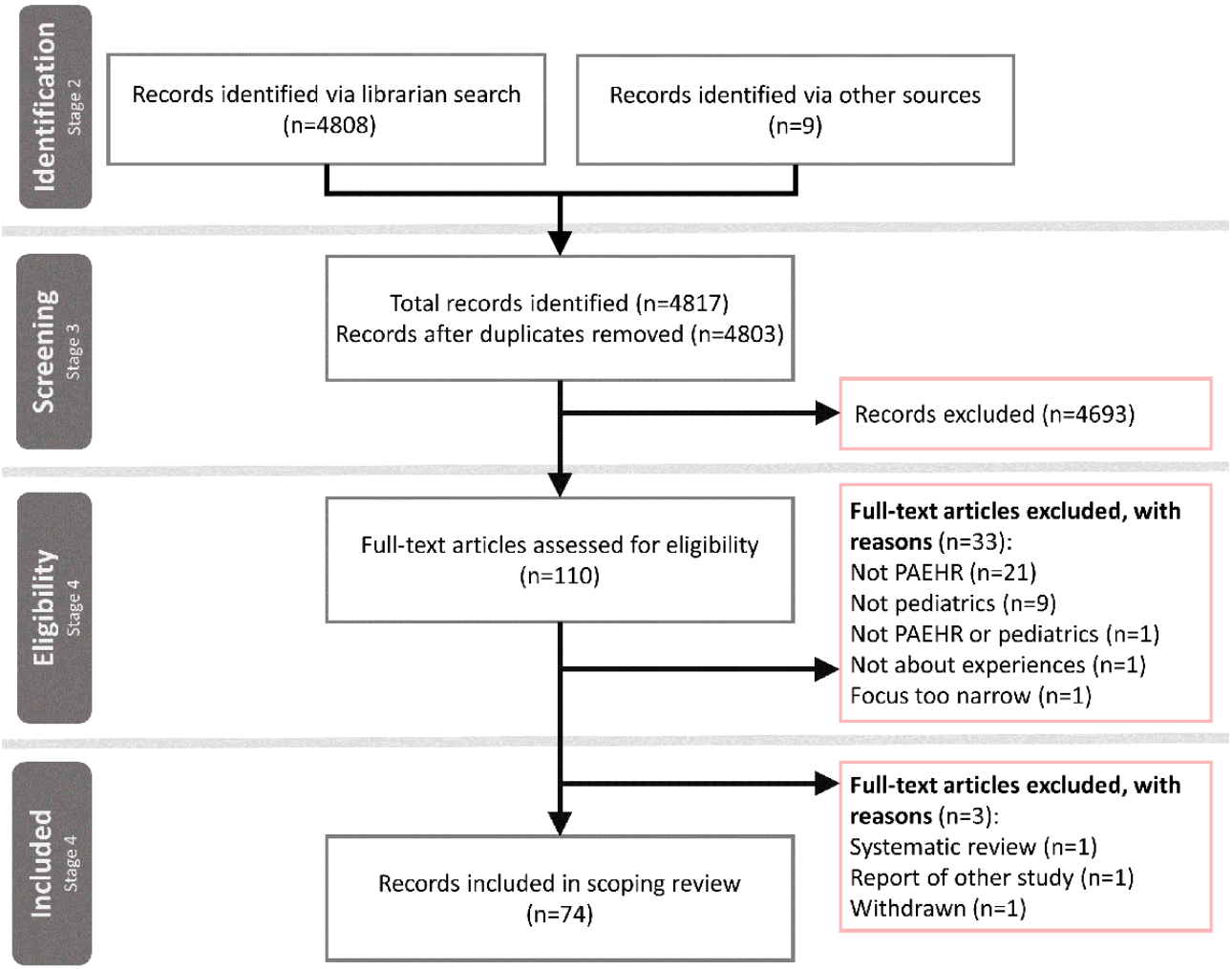
To gain further insights on the topic, the results were shared with stakeholder representatives, including a pediatric oncologist, members of a young patient council at a public hospital in Sweden, and the Ombudsman for Children in Sweden. The representatives were provided with material via email and invited to choose to provide their thoughts in text via email or verbally during a web-based meeting.

Results

Study Selection

Figure 1 shows the study selection process in a PRISMA diagram [95]. In total, 4817 records were identified, of which 4808 (99.81%) were identified via a database search and 9 (0.19%) were identified via other sources. After removing duplicates, 99.71% (4803/4817) of the records remained for screening of abstracts, titles, and keywords. In this process, 97.71% (4693/4803) of the records were excluded, resulting in 110 full-text articles to be assessed for eligibility. As a result of this, 1.6% (77/4817) of the total records identified met the inclusion criteria. During the analysis, 0.06% (3/4817) of the records were excluded, leaving 1.54% (74/4817) of articles included in the review.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram adapted from Moher et al [22]. PAEHR: patient-accessible electronic health record.



Basic Characteristics of the Body of Evidence

The included studies were mainly viewpoint papers or used quantitative methods (Table 1), and 92% (68/74) were based in the United States (Figure 2). The number of articles published in the area of PAEHRs for parents, children, and adolescents was fairly stable over time (Figure 3), ranging from 3% (2/74) of the articles in 2007 to 16% (12/74) in 2021. An increase can be observed for 2018 and 2021, and none of the articles during these years belonged to a special issue.

XSL•FO
RenderX

Table 1. Basic characteristics of the included studies (N=74).[a]

Parameter	Total, n (%)
Study design	
Viewpoint or comment	27 (36.5)
Quantitative	27 (36.5)
Qualitative	13 (17.6)
Mixed methods	7 (9.5)
Publication year	
2007-2009	7 (9.5)
2010-2012	7 (9.5)
2013-2015	13 (17.6)
2016-2018	23 (31.1)
2019-2021	24 (32.4)
Country	
Australia	3 (4.1)
Canada	1 (1.4)
New Zealand	1 (1.4)
United Kingdom	1 (1.4)
United States	68 (91.9)
Study participants[b]	
Children and adolescents	6454 (5.5)
Parents	110,184 (94.1)
Health care professionals	496 (0.4)
N/A[c] (no participants or not specified; studies)	34 (45.9)
Treatment setting	
Pediatric	34 (45.9)
Adolescent	15 (20.3)
Adult	2 (2.7)
Inpatient	15 (20.3)
Outpatient	20 (27)
Academic	1 (1.4)
N/A	7 (9.5)
Clinical field	
Chronic illnesses (cystic fibrosis, juvenile idiopathic arthritis, or diabetes mellitus)	6 (8.1)
Psychiatry	4 (5.4)
Intensive care	4 (5.4)
Gastroenterology	2 (2.7)
Hematology	2 (2.7)
Obstetrics and gynecology	2 (2.7)
Neonatal care	2 (2.7)
Cancer	1 (1.4)
Cardiology	1 (1.4)
Pulmonology	1 (1.4)

XSL•FO
RenderX

Parameter	Total, n (%)
Emergency	1 (1.4)
Hepatology	1 (1.4)
Subspeciality	1 (1.4)
Radiology	1 (1.4)
N/A	7 (9.5)

[a]Individual papers can be assigned to various subparameters at the same time, which means that percentage totals of >100% can be achieved.

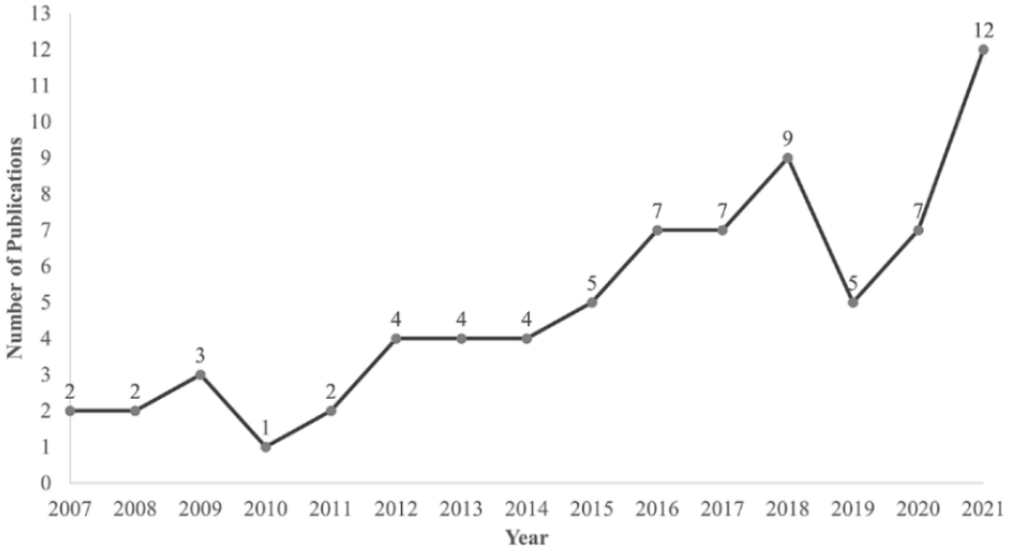
[b]The number of study participants was accumulated based on empirical and observational studies that included a reported number of study participants.

[c]N/A: not applicable.

Figure 2. Included publications by country and studied stakeholder group. HCP: health care professional; UK: United Kingdom; USA: United States of America.

XSL•FO
RenderX

Figure 3. Number of publications by year.



Search Results

The results were divided into four groups of stakeholders: (1) children and adolescents, (2) parents, (3) HCPs, and (4) other stakeholders. For children, adolescents, and parents, the identified categories were adoption and use, positive views and experiences, and concerns and negative experiences. For HCPs, the identified categories were positive views and experiences and concerns and negative experiences. For expert analysis or viewpoints, the identified categories were positive views and experiences and concerns.

Children and Adolescents

Overview

Views, use, or experiences of PAEHRs among children and adolescents comprised a relatively small part (16/74, 22%) of the included studies. Of these 16 studies, 14 (88%) were conducted in the United States, 1 (6%) was conducted in Australia [20], and 1 (6%) was conducted in Canada [21]. Most of these studies were observational (6/16, 38%), followed by surveys (4/16, 25%), qualitative studies (focus groups or interviews; 3/16, 19%), and mixed methods studies (2/16, 12%). Only 1 opinion paper was included, authored by a male patient aged 15 years [22]. Survey studies ranged from 20 [23] to 1006 [21] participants. Qualitative studies used focus group interviews (2/16, 12% of the studies) [20,24] and individual interviews (1/16, 6% of the studies) [25]. The most frequent care settings were pediatric inpatient care, primary care, psychiatry, and nonclinical care. In total, 12% (2/16) of the studies focused on the general population [20,21]. Observational studies focused on adoption and use over time, demographic data, and frequently used functions of patient portals [26-31], whereas survey studies explored satisfaction with reading the records [23,32,33], literacy [23,32], intervention effects [33], attitudes toward web-based patient portals [21,34], and barriers to adoption [34]. The studies included adolescents aged between 12 and 20 years, and 12% (2/16) of the studies included patients aged ≥18 years [21,32]. A few studies (2/16, 12%) included adolescents and proxy users and did not distinguish adolescent patients from proxy users in their analyses [26,31].

Adoption and Use

A number of studies (4/16, 25%) reported low adoption and use of patient portals among adolescents compared with other age groups [27-29,31]. In a US study, 11% of patients aged 10 to 17 years had activated an account at a patient portal implemented 3 years before [27]. Similarly, a study that allowed for surrogate access and individual accounts for patients aged >13 years with parental consent found that adolescent patients composed 16.5% of all log-ins, although use increased during late adolescence [29]. A study based in Canada identified an age-related difference where younger adolescents (aged 12-15 years) were more open than older adolescents (aged 16-19 years) to sharing their notes with parents [21]; however, a US study with a smaller sample size observed a similar tendency but no significant difference [33]. For adolescent patients with cancer, the perceived value of record access decreased during recovery [35]. Knowledge of PAEHRs was reported as low not only among adolescents without access to a patient portal in US studies [21,24,34] but also in a focus group study based in Australia where adolescents had access to their EHR from the age of 14 years [20]. The studies were inconclusive on gender differences in adolescents' PAEHR adoption, finding either no differences [30,33] or a greater inclination among female patients [27]. In 6% (1/16) of the studies, male patients aged between 12 and 17 years had the lowest percentage of viewing their results in the patient portal (<1%) [28]. A study of 96 urban, low-income African American late adolescents in outpatient care found that male patients were more likely than female patients (P=.001) to consider allowing proxy access [33]. Regarding mode of access, adolescents in 12% (2/16) of the studies reported a preference for smartphones or tablet devices over computers [34,35].

Positive Views and Experiences

Studies that explored views on PAEHRs among adolescents who had not previously accessed their records identified a strong interest in access [20,21,24,33,34]. For example, among 1006 adolescents, 84% supported the idea that adolescents should be able to read their records on the web [21]. Adolescents wished to receive information about EHRs from HCPs according to

XSL•FO
RenderX

their future needs [20,24]. Notably, an intervention study in which adolescents in primary care were informed about a patient portal observed an increase in portal account activation but not in use [33].

Positive expectations were confirmed by adolescents reading their records, with high satisfaction reported by studies in gastroenterology (9.2/10) [32], psychiatry (8.8/10) [23], and primary care (79%) [33]. In the study by Hong et al [35], adolescents with cancer and blood disorders read their records to ensure accuracy and check for updates. For these adolescents, reading their records led to reduced anxiety, enhanced knowledge about their illness, an ability to ask informed questions, and more reflection on their health. If needed, they consulted the internet or asked their parents. A US study conducted in a psychiatric ward found that having record access led patients' trust in their health provider to either increase or remain the same [23]. In total, 12% (2/16) of the studies observed adequate literacy, with almost no exceptions among patients in psychiatry [23,32]. Both adolescents with and without experience of having access to their records foresaw empowerment [22,24,25,35]; a male patient aged 17 years stated in an opinion paper [22] that access "could help my generation learn about our health care system" and "encourage [adolescents] to take more responsibility for our health." Patients with cancer anticipated that PAEHRs could support the transition from pediatric to adult care [35]. A high school senior in hematology who had used a patient portal suggested that the records could be jointly managed by themselves and their parents during the transition to adult care [25].

Better recall was an anticipated benefit among adolescents who did not access their records [20,22]. Furthermore, adolescents who did not have access to their records foresaw the utility of checking test results [21,24,34], messaging [20,24], viewing medications [20,21,34], reading visit notes [20], reviewing appointments [21,24], and viewing allergies [24]. In primary care contexts, adolescents valued being able to ask questions via email rather than in person, particularly concerning sensitive information [24]. Similarly, the most accessed information in observational studies was commonly test results [27,30,35], messaging [27,31,35], appointments [27,30,35], and medications [30,35]. Reminders were considered useful for planning around daily life [25,33,35], and a frequently asked questions section was suggested for ease of use [24].

Adolescents with cancer or a blood disorder who had accessed their records reported no concerns about what their parents would see in the EHR [35]. In an institute providing primary and mental health care that used a patient portal where a minor's consent was required when aged >10 years to release information to parents, HCPs had received no complaints about confidentiality from adolescents since the implementation [27]. In an Australian focus group study, a participant noted that, in spite of valuing privacy, timely access to medical data in a critical situation was more meaningful [20].

Concerns and Negative Experiences

Although relatively few, the leading concern was health literacy among adolescents. Adolescents without access to their EHRs expressed worry about not being able to understand and appropriately interpret the information in the EHRs [21,24]. Among patients in psychiatry who read their records, half reported not understanding the discharge criteria [23]. In studies where adolescent patients had the option to suggest note changes, edits concerned personal history and anthropometrics [23] as well as allergies and medication reconciliation [32].

Concerns about internet security and confidentiality whereby parents might access their EHR were expressed by adolescents with no access to their records and who were patients in an outpatient or nonclinical setting [21,24]. A teenager in another study suggested that the relationship with the parent may affect the teenager's feelings toward parental access, and in case of shared access, a private email option would be useful [25]. Adolescents without EHR access reported feeling uncomfortable with sharing their health information on social media [21].

Parents

Overview

Parents' and guardians' experiences with web-based access to health records comprised more than a third of the studies (33/74, 45%). Of these 33 studies, 31 (94%) were conducted in the United States, 1 (3%) was conducted in Australia [36], and 1 (3%) was conducted in the United Kingdom [37]. The most common studies were surveys (10/33, 30%), followed by qualitative (9/33, 27%), observational (9/33, 27%), and mixed methods (5/33, 15%) studies. Among these were an opinion piece coauthored by a parent [25] and a usability test where 16 parents evaluated the usability of a patient portal prototype [38]. The most common settings were pediatric inpatient care, outpatient care, in-hospital care, primary care, congenital cardiac surgery, and hematology. The observational studies focused on adoption and use over time, demographic data, and frequently used functions of patient portals. The qualitative studies included both individual interviews (5/33, 15% of the studies, whereof 1/5, 20% also included observations) and focus group interviews (3/33, 9% of the studies). The survey studies ranged from 25 [39] to 3672 [40] participants. A total of 12% (4/33) of the studies had <100 participants, and only 6% (2/33) of the studies had >500 participants. Of the survey studies, 12% (4/33) explored parents' thoughts about using a web-based patient portal [41-43] or their teenagers using such a portal [44] in the future. Of the remaining 8 survey studies, 3 (38%) focused specifically on errors in the record and patient safety issues [26,40,45]. In total, 6% (2/33) of the studies did not distinguish between parents and patients in their analyses [40,45].

Adoption and Use

The studies reported high rates of patient portal adoption and use among parents during the first years of the child's life [29,30]. In both Australia and the United States, studies identified the highest rates of patient portal activation for the youngest children of both sexes aged 0 to 11 [28] and 0 to 14 years [36]. In studies that required the assent of older adolescents for parental access, parents' use of patient portals decreased. A study of a patient portal that required such assent received no applications for unrestricted access, and 80.4% of parents or guardians who enrolled had children aged <10 years [27]. In a longitudinal study where there were no restrictions, 93.62%

XSL•FO
RenderX

(16,036/17,128) of all pediatric patients during the study period had a surrogate (parents or legal guardians), and surrogate users accounted for 83.2% of all log-ins for adolescent patients [29]. There was higher use among parents of children with chronic diseases [46]. Another study observed a 100% adoption rate among parents as proxy users for children aged 0 to 11 years, whereas merely 5.9% of parents of adolescents enrolled [30].

In an inpatient setting, a study [47] found that 27.89% (530/1900) of families created a patient portal account, 47.8% (238/498) used the portal within 3 months of registration, and 15.9% (79/498) continued using the portal 3 to 6 months after creating the account. A US study identified disparities in social demographics; parents who identified as Hispanic, Asian, or "other races" than White were less likely to use a patient portal, which was hypothesized to be related to language barriers and device accessibility [48]. The same and another study identified that privately insured parents were more likely to enroll in portal activation than those with public insurance [46,48]. In a study in which 12 children died during the study period, most families continued accessing their children's records after their death [49]. A study of parents of children with attention-deficit/hyperactivity disorder found that, although half of the participants used their home computer to read the records, one-third accessed the portal on their smartphone and that barriers to use included lack of awareness, lack of internet access, lack of time, and password problems [50]. Schneider et al [37] identified four different use styles families at a children's hospital in the United Kingdom applied to access the children's records: controlling (approach-oriented and highly motivated to use PAEHRs), collaborating (approach-oriented and motivated to use PAEHRs), co-operating (avoidance-oriented and less motivated to use PAEHRs), and avoiding (very avoidance-oriented and not motivated to use PAEHRs).

Positive Views and Experiences

Several studies (4/33, 12%) focused on parents' expectations or thoughts about PAEHR use before actually having experienced access to their child's EHR [24,42,51,52]. In a 2013 US study, parents were approached in the waiting room and given a demonstration of the patient portal. A total of 72% (46/64) of the participants had not heard of the patient portal before, and only 28% (5/18) of those who had heard of the portal had used it. Nearly 70% (44/64) of the parents intended to use the patient portal after the demonstration [42]. Expectations were mostly positive and confirmed by studies with parents who had experience of record access, yet concerns were also discussed, which will be presented in the following section.

Better recall or reinforcement of information was reported as a benefit in many studies (7/33, 21%) [24,38,45,51-54], as was improved parental knowledge and understanding of their child's health [39,51,53,55-59,96] and a sense of control [39]. In addition to access to information, parents reported enhanced communication and partnership with providers [11,39,45,51,53,55-58]. In a study on parents of hospitalized children, the addition to the PAEHR of pictures of staff taking care of the child was highly appreciated [58]. Another reported benefit was not having to bother clinicians [56,57,96]. As anticipated by parents of hospitalized children [51], having access to the child's record also improved parental empowerment and the parents' ability to advocate for their child [11,43,53,55,56]. Furthermore, parents of children with cancer or chronic illnesses described reduced anxiety as a positive result of having access to their children's records [11,96]. The benefit of error detection was both reported by parents who had experiences of accessing their child's records [35,55,56,58] and anticipated by those who did not [51].

Records were used to prepare for discussions with clinicians [35,39,56], formulate questions, and ask for explanations [35,56]. Another study observed that parents used the portal to ask questions about their children's minor illnesses and request medication refills [27]. In studies in which parents were asked for suggestions for portal improvement, they often cited more information, such as a portal use tutorial [56], more educational links and resources [57,58], medical explanations or interpretations [38], and clarification of medical jargon [38]. However, in a survey study of 25 parents with real-time access to their children's EHR, none considered notes more confusing than helpful [39].

Studies varied in the available portal features and details of reporting use. In total, 6% (2/33) of the studies provided a similar broad functionality, consequently seeing a similar use where one study [30] found high use of appointment reviews (85%), messaging (84%), test results (79%), and immunizations (79%) and the other [46] found parents to frequently access immunizations (80%), messaging (72%), appointment reviews (55%), and test results (50%).

Parents of children who were seriously ill consistently reported positive experiences, for example, parents who had immediate access to laboratory test results in an inpatient portal during a child's hospitalization [55] and parents of children diagnosed with cystic fibrosis, juvenile idiopathic arthritis, or diabetes mellitus [11]. Chung et al [43] reported that 92% (78/85) of parents of hospitalized neonatal children wished to receive information even if it was "bad news." A study among parents of patients in pediatric radiology found that, although only 12.1% (18/104) accessed a web-based portal to check their children's test results, 65% prioritized minimal waiting time as the most important aspect for receiving results [60].

Some studies explored parents' views on their teenagers accessing their own records, and parents saw it as a way for the teenagers to take better control of their own health care [24,96]. When parents of adolescents in juvenile detention were asked for their opinion on giving their teenagers access to their health records, 70% were positive, and 100% felt that the adolescent should be able to share this information with their parents through the web-based system [61]. Parents also felt that the PAEHR would be useful when transitioning to adult care or another care provider [35,51].

Concerns and Negative Experiences

Before having access to the record, parents worried about information being released without face-to-face communication [51,53]. When it came to adolescents having access to their own records, parents had privacy concerns that the portal might be hacked [61], that the teenager would be pressured to share

XSL•FO
RenderX

information [61], or that billing of confidential services would cause privacy breaches [24]. Some requested that parents be required to consent to teenagers having access to portals [24,61] and were worried that teenagers would make appointments without parents knowing and wanted to be informed about email conversations [24]. Moreover, parents worried that adolescents might not reveal sensitive information if they knew it would be visible to their parents [51]. In a US survey study of 93 parents where half were parents of adolescents, 68% were negative about their children receiving information from their HCP through a secure web portal [44]. In a study in which parents of children in an intensive care unit were provided with real-time EHR access on a large monitor, parents expressed concern about visibility to bypassers [56]. Issues around parents' loss of access to the record as the child enters adolescence were highlighted by Carlson et al [25], suggesting that record access needs to be an integrated part of the transition from childhood through adolescence and into adult health care. In the study by Hong et al [35], it was found that parents of teenagers with cancer would act as intermediaries in communication with HCPs as teenagers preferred to discuss their health with their parents rather than with clinicians. Thus, proxy access was considered essential. Parents in this study also expressed concerns about negative results being immediately available to teenagers, worrying that they might cause anxiety [35].

Some felt that teenagers may not understand all medical information, including test results [24], and that they might use the portal inappropriately and would need education [24,25]. Medical jargon was reported as an expected challenge in several studies (4/33, 12%) [38,43,51,96] as well as not being able to interpret complex results without context or explanation [56]. Parents of teenagers with cancer reported searching on the web to help make sense of the medical record and seeking additional information not readily available on MyChart [35]. Among 270 parents in a pediatric outpatient setting, 52.5% expected to read the medical records if they had access to them, with one-third indicating that they "sometimes" needed help reading health materials [41]. In another US study, 5% of surveyed parents of children with cancer reported understanding the notes to be "somewhat" or "very difficult" [59]. However, a study found that, among patients and families finding a serious mistake in visit notes, only approximately half reported the mistake, barriers including lack of knowledge of how to report but also fear or retribution [40].

Among concerns about PAEHRs, increased confusion, distress, or anxiety were anticipated by parents with no access [53]. Both parents with and without experience of PAEHRs worried about record access impairing the parents' relationship with the provider [11,53] and, in turn, negatively affecting collaboration [53]. Another concern stemmed from empathy with HCPs, worrying that parental record access could increase the workload and lead to complications [51,53] or restrict communication between HCPs through the record [51].

HCP Stakeholders

Overview

Comparatively fewer studies (11/74, 15%) explored HCPs' experiences of or opinions on web-based access to EHRs for children, adolescents, and parents. Of these 11 studies, 8 (73%) were conducted in the United States, 2 (18%) were conducted in Australia [20,62], and 1 (9%) was conducted in the United Kingdom [37]. Most of these studies (6/11, 55%) were qualitative (focus groups or interviews), although the sample sizes were small; 18% (2/11) of the studies had a sample size of 1 [23,25]. In total, 9% (1/11) of the studies used a web-based survey [63], and 18% (2/11) used paper-based surveys [43,52]. Many studies (6/11, 55%) recruited representatives from a wide variety of clinicians, including, for example, specialist physicians, general practitioners, hospitalists, nurse practitioners, nurses, mental health clinicians, physician assistants, dietitians, physiotherapists, and pharmacists [20,37,43,53,62,63]. Survey studies ranged from 1 [23] to 212 [63] participants. Notably, only 18% (2/11) of the studies exclusively solicited the views of pediatric health professionals [62,64]. Several studies explored HCPs' broad experiences with sharing PAEHRs with patients and parents [37,52,62]; a few focused on HCPs' anticipation of the practice among children or adolescents and parents [25,43,53]. In total, 18% (2/11) of survey studies exclusively focused on providers' perspectives on adolescent confidentiality with PAEHRs [63,64]. Only 12% (1/8) of the US studies reported on both accessibility and age of access: of 212 clinicians, 87.6% reported that their institution offered PAEHRs to both the adolescent and their parent or guardian, and most (69.1%) reported a minimum age requirement, with most (42.2%) citing between 12 and 14 years [63].

Positive Views and Experiences

Studies that explored HCPs' experiences with PAEHRs among children or adolescent patients and parents reported positive experiences. For example, among 96 providers with experience sharing access at a children's hospital, Kelly et al [52] found that 92% wanted patients and parents to continue to use the portal. They reported that patients and parents asked questions about the information they read, including laboratory test results (45%), medications (13%), and errors or mistakes in their care (3%). Exploring the views of HCPs in pediatric settings, Janssen et al [62] found that staff appreciated enhanced communication with patients, especially regarding coordinating appointments with parents and the potential for families or patients to ask questions. A study soliciting the views of 1 provider working in an adolescent inpatient psychiatric setting reported that clinical note sharing helped inpatient counseling sessions and compliance [23]. A study including 25 physicians identified experiences of increased transparency, improved documentation, reassurance or validation of concerns, and enhanced care plan clarity [39].

Among the anticipated benefits of sharing PAEHRs with child or adolescent patients and parents among HCPs with no experience of the practice, Kelly et al [53] reported that clinicians (including 8 nurses, 5 residents, and 7 hospitalists) predicted reinforced information, improved parental knowledge and empowerment, enhanced parent communication and partnership with providers, and increased provider accountability and documentation quality. Among 133 surveyed medical professionals, Chung et al [43] reported that 63.2% (84/133) believed that parental access may help identify incorrect information, and 61.7% (82/133) believed that access may

XSL•FO
RenderX

reassure parents of the care provided to their child. In a qualitative study based in Australia by Beaton et al [20], school-based clinicians anticipated that adolescent patients with multiple providers would benefit from reduced duplication of investigations, ineffective treatment strategies, and more timely access to information.

Concerns and Negative Experiences

In several studies (4/11, 36%), patient confidentiality breaches and managing private patient information among children and adolescents was the leading concern [20,25,63,64]; as 1 surveyed clinician noted, "Privacy is just the biggest thing" [20]. In 18% (2/11) of the studies, HCPs reported that, despite sharing PAEHRs with other patients, they had precluded sharing information with adolescents because of privacy concerns, such as that savvy parents would be able to access it [20,25]; attesting to this, lack of clinician familiarity with PAEHR utility and technical implementation among minors was another expressed concern in both studies. Among clinicians with experience of PAEHRs, in a US study of 212 clinicians, nearly 4 in 10 (39.6%) were not at all confident that their EHR maintained privacy for minors, with 81.7% expressing concerns about maintaining confidentiality [63]. In another US study of 26 pediatric health care providers with experience of sharing PAEHRs with adolescents, Stablein et al [64] reported that confidentiality concerns affected documentation practices, such as worries that all HCPs involved in the child's care will not be aware of what information in the record is private from parents versus what the parent needs to know, in addition to the fact that the record has a multifold purpose (eg, billing and communication with families). As a result, providers reported selectively omitting or concealing information and using codes on the EHR designed to alert other providers to confidential information.

Kelly et al [53] reported that HCPs with no experience of the practice (including 8 nurses, 5 residents, and 7 hospitalists) foresaw increased provider workload, heightened parental confusion, distress or anxiety, impaired parental relationship with providers, and compromised note quality and purpose. In a US study, 34% (17/50) of attending and intern physicians were concerned that parents would be confused by reading their child's notes [39]. Among 133 surveyed medical professionals, Chung et al [43] reported that 114 (85.7%) believed that parental access may make medical professionals apprehensive about charting certain information, and 75 (56.4%) believed that parental access may increase the time spent updating parents, with approximately half (64/133, 48.1%) believing that parental access may increase the probability of a lawsuit. A study of inpatient pediatric physicians with experience of access found that 11% reported increased workload and 4% reported not being satisfied with portal use by patients or families [52].

Other Stakeholders

Overview

The viewpoints of other stakeholders on pediatric PAEHRs constituted most of the included studies (30/74, 41%). These studies comprised three types of stakeholders: (1) experts (27/30, 90%) such as HCPs, IT experts, or researchers; (2) policy makers (4/30, 13%); and (3) the public (1/30, 3%). Of these 30 studies, 28 (93%) were conducted in the United States, 1 (3%) was conducted in Canada [21], and 1 (3%) was conducted in New Zealand [65]. The aim of many studies (15/30, 50%) was focused on ethical issues related to adolescent PAEHRs, and a few (2/30, 7%) described the development of a portal solution [66,67].

Positive Views and Experiences

Viewpoints focused mainly on concerns (which we describe in the following section) but included a number of positive views of PAEHRs for a pediatric population. Among informants from 25 medical organizations, it was stated that adolescent patients with chronic diseases benefited the most from parents having access [68]. In fact, pediatricians claimed that parents of children with chronic diseases should be offered full access to their children's EHRs [69]. Jasik [70] advocated that PAEHRs could be useful in health education, in support of care transition for adolescents with chronic illnesses, and in risk behavior screening. Several viewpoints (3/30, 10%) argued that adding educational materials to the PAEHRs may facilitate literacy and comprehension for families [67,71,72]. Some noted unfulfilled potential for pediatric PAEHRs, for example, in the areas of patient data contribution [66], developmental screening [73], and research trial participation [74].

Green-Shook [75] anticipated that HCPs' control of their schedule may increase with PAEHRs because of communication with patients via messaging rather than telephone, an anticipation that was subsequently observed in a primary care setting [48]. Several papers (4/11, 36%) reported a need for availability on mobile devices to increase accessibility and practicality for users [67,69,70,74], and a medical director developing a mobile PAEHR app advocated for the integration of various functions in one app [67].

Concerns

Most viewpoint papers included concerns about adolescent confidentiality [72,73,76-84]. HCPs in gynecology and psychiatry reported that adolescents may be less willing to seek health care if they are uncertain about confidentiality [82,83], and 83% of respondents in a public opinion survey [85] deemed adolescents less likely to discuss sensitive issues with HCPs when parents had access to their EHRs. An American organization advocating for adolescents' health warned that adolescent aversion toward PAEHRs caused by confidentiality concerns and an uneven internet access could increase health disparities [86].

The studies described concerns in terms of portal functionality. Many insisted on an option for HCPs to label information as confidential [68,69,76,83,87] and enable adolescents to restrict parental access [80,86], some pointing to the variable definition of "sensitive" [68,81], which portal features contain such information [83], and division into "portions" of notes [84]. Psychiatric PAEHRs have been noted as unique in need of confidentiality, and Kendrick and Benson [83] listed portal functions that may hold information pertaining to sensitive topics in mental health, noting that sexual activity, gender identity, and substance abuse may be accessed in all portal areas. Bayer et al [80] posited that the release of sensitive information

XSL•FO
RenderX

to the parent should require the adolescent's consent, whereas Bourgeois et al [88] urged HCPs to carefully review notes to prevent leakage of sensitive information. Not only concerning children, 7% (2/30) of the studies noted the need for protecting caregiver privacy [73,76]. In fact, medical professionals favored customizable controls of information display for both parents and adolescents [69], and several studies prompted considering family circumstances [65,89]. A group of pediatricians suggested that structured data content could improve efficiency and consistency [73].

Jasik [70] asserted a lack of stakeholder investment in PAEHR development for adolescents and that current portals are usually not designed to deal with privacy issues. Attesting to this, pediatricians noted that adolescent access to patient portals is hindered by time-consuming decision-making and lacking technology and manpower and that implementation variability is a result of absent guidelines and vague laws [68]. Anoshiravani et al [69] proposed that portal access for adolescents should be limited until the privacy functionality is more robust.

Set age limits for patient and parental access to mitigate confidentiality issues has raised concerns and been the topic of much debate. Taylor et al [89] suggested different content access for three subgroups in the pediatric population (aged <13 years, 13-18 years, and >18 years) based on information sensitivity. Various studies (2/11, 18%) held that default ages may enable long-term consistency [65,68], allow for automated notifications, and facilitate policy making [65]. Conversely, viewpoint papers cautioned that age-based loss of access could seriously affect families reliant on EHR access in the care for a child [68]. With regard to the transition from child to adult, Sittig and Singh [78] discussed the transfer of EHRs created when the patient was a child, whereas Bourgeois et al [88] reported that their institution provided access "prospectively," keeping confidential information suppressed also when the individual became an adult.

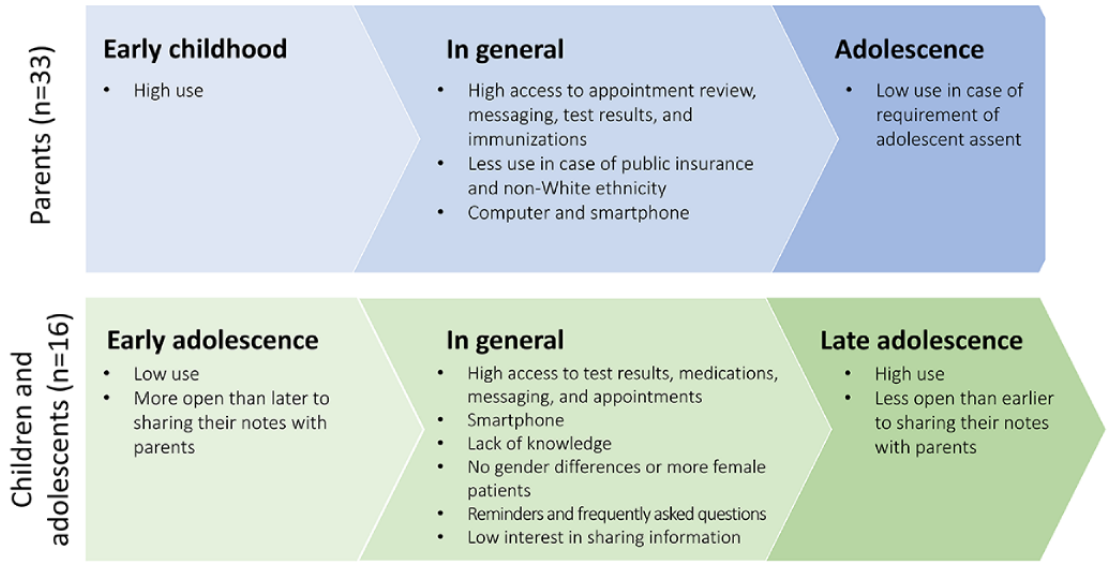
Several viewpoints (3/11, 27%) advocated for education on PAEHRs for various stakeholders [75,86,87], for example, that early HCP-initiated conversations with parents and adolescents may reduce parental concerns, increase acceptance [65], and set clear expectations [88]. Obstetrician-gynecologists have argued that adolescents should be informed if parents will have access to the EHRs [84], and Sherek and Gray [90] stated that, when possible, parents need to be informed of how to extend access to the child's record. In a short paper, the American Academy of Child and Adolescent Psychiatry [91] provided advice for parents on questions for their child's psychologist. In total, 13% (4/30) of the studies noted that insurance claims can lead to confidentiality issues [75,80,84,87], especially with uninformed use of the PAEHRs. The importance of guidance for staff has also been stated [69,88,92] as well as communication between staff and EHR vendors [68]. In pediatric psychiatry, Nielsen [81] advocated for training graduate students in penning PAEHRs. On the topic, a group of pediatric gastroenterologists recommended removing irrelevant details, not labeling emotions, and spell-checking [71].

Among other concerns, Gracy et al [73] described the divergent needs of pediatric portals compared with those of adult populations. Spooner [72] listed the critical areas for pediatric PAEHRs as immunizations, growth tracking, medication dosing, patient identification, norms for pediatric data, and privacy.

Visual Summary of Stakeholders' Expectations and Experiences

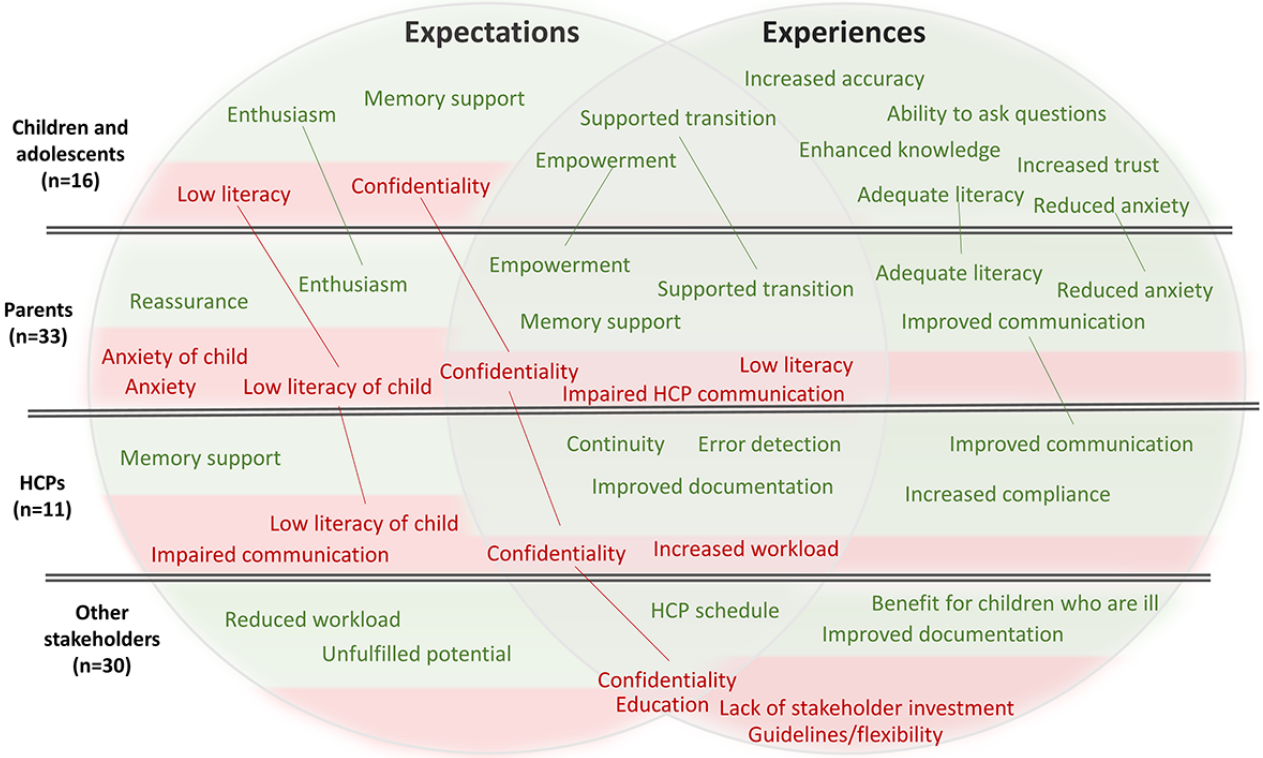
Figure 4 presents a visualization of the findings on adoption and use among adolescents and parents. Furthermore, Figure 5 provides a visualization of the findings based on expectations and experiences. Here, "expectations" is, as mentioned previously, a type of view in which the stakeholder has no previous experience of web-based record access.

Figure 4. Summary of results of adoption and use of patient-accessible electronic health records among parents and adolescents. FAQ: frequently asked questions.



XSL•FO
RenderX

Figure 5. Summary of results of expectations and experiences of patient-accessible electronic health records among children/ and adolescents, parents, health care professionals (HCPs), and other stakeholders. Green text depicts positive views and experiences, and red text depicts negative views and experiences. Color in the various boxes illustrates the distribution of positive and negative views and experiences for the stakeholder group.



Discussion

Principal Findings

The results of the 74 studies included in this scoping review contribute to the understanding of factors associated with stakeholders' views, use, and experiences of children's, adolescents', and parents' web-based access to the EHRs of children and adolescents. The reviewed studies consistently observed positive views and experiences on the part of parents and particularly of adolescents, whereas HCPs and other stakeholders held many concerns. In this section, we will (1) compare stakeholders' views on and experiences with PAEHRs, (2) discuss some of the challenges that are unique to the PAEHRs of children and adolescents, (3) comment on the implications for design and implementation, and (4) suggest future research.

Limitations

Although it followed the scoping review methodology, the review was limited by not assessing the quality of the included studies. By only including studies written in English, we may have missed important papers written in other languages. Considering that 92% (68/74) of the included studies were based in the United States, we do not know whether an information bias affected the findings. Among the identified studies, some merged adolescents with young adults or parent proxies, which complicated the analysis of specific groups. Furthermore, the studies' definitions of adolescents varied, with the upper age limit ranging from 17 to 20 years. The studies did not always distinguish between positive and negative views or experiences. For example, the provision of education and guidance could be deemed as both a benefit and a concern. Furthermore, several expert viewpoints provided recommendations for the future based on concerns about PAEHRs, omitting to mention benefits. For the purpose of this study, we referred to the effects of PAEHRs that appeared beneficial to the patient as "benefits." Finally, conducting stakeholder consultations after completing the review prevented any integration of their results into the study. Future scoping reviews may wish to invite stakeholders to a more active participation earlier or to provide input throughout the process.

Expectations Versus Experiences Among Adolescents and HCPs

The findings suggest a similar pattern for adolescents to that previously observed in adult populations [9,10,93,97-99], where adolescents' positive experiences contrast with HCPs' concerns. For example, HCPs and parents imagined that adolescents would not understand the information in their notes and experience negative emotions as a result. However, adolescents reported high satisfaction and literacy even in the much-debated field of psychiatry. Another interesting aspect was that, although adolescents who had not previously accessed their EHR notes did have concerns about not understanding the notes and what parents may have access to, those with experience of accessing their records reported no such concerns.

A possible explanation for this might be a different perspective as most nonobservational studies exploring adolescents' experiences with PAEHRs (4/5, 80%) included patients with serious illness or in inpatient settings. It might be that children and teenagers with serious illnesses may have a better understanding of medical jargon. In addition, they may be

XSL•FO
RenderX

familiar with being dependent on parental insight into their care and involving parents in their health care issues. Thus, the adolescent's desire for privacy is likely to depend on many factors, and there is still a need to provide confidentiality for those who require it, which was mentioned in many viewpoint papers.

The one existing review in the field [14] did not include expectations of PAEHRs; however, its findings in terms of experiences were aligned with our included evidence. For example, there was enthusiasm among adolescents and interest among parents in using patient portals, whereas medical literacy and confidentiality were the main concerns. Similarities are not surprising as, of the 11 included papers in the aforementioned review, only 3 were not included in this review (because they did not have a focus on pediatric PAEHRs). Except for not focusing on expectations, among the differences were that the previous review included use barriers and clinical outcomes and did not include the perspectives of HCPs and other stakeholders.

Interestingly, all but one of the parents' concerns about adolescents' confidentiality referred to external parties rather than the self as a parent as a threat to their adolescent's privacy. It is difficult to explain this result, but it might be related to the fact that parents have been found to value the importance of their involvement highly out of concerns about not being apprised of important information and uncertainty of the child's ability to manage their own care [100]. Instructing HCPs to engage parents and adolescents in a dialogue on confidentiality has been mentioned previously as a strategy to mitigate parents' worries, although current extensive pressure on HCPs may necessitate new approaches to such education.

Special Challenges for Pediatrics

A key challenge for PAEHRs is balancing confidentiality and information privacy for adolescent patients with the need for parental involvement in the adolescent's care. Several viewpoint papers focused on guidelines regarding when and how to grant access to parents and adolescents. The results are inconclusive and reflect the complexity of this issue. A health institute argued that allowing for manual changes to parental access can signal that the child has received some type of sensitive care [27]. Set age limits for automatic gain and loss of access could be beneficial, yet an extensive variety of potential circumstances do call for customizability according to the situation. A lack of investment and priority of portal development for adolescents and parents was indicated, which one could argue causes a waste of potential of PAEHRs and a loss for the health care system in the long term. One such function advocated by numerous viewpoints was the possibility of designating information as confidential. Still, efforts to hide sensitive information from parental view could be counteracted by parents evading the system to access their adolescents' accounts directly. If a parent perceives their adolescent incapable or unwilling to manage their own health care, they may consider it necessary and part of their parental responsibility to find a work-around. A recent UK article published outside the search period indicated that more than half of the messages to adolescents' accounts were accessed by guardians [101]. In addition to protecting the adolescent, a few papers stressed the importance of considering caregiver privacy in cases where parents disclose confidential information with regard to the child's care. Furthermore, modern family constellations vary, which may require the consideration of access provision based on the type of parental or legal guardianship. In a case study, health data coordinators at a US medical center described using different rules of access for a "natural or adoptive parent," legal guardian, or stepparent [90]. The same institution denied parents aged <18 years access to their child's EHR before becoming an adult, highlighting another potential issue. Differences between countries further complicate the issue of PAEHRs in pediatrics; for example, the definition of policy maker in the PAEHR context varies considerably by country, whereby HCPs in some countries are required to decide on policies themselves.

Consultation With Stakeholder Representatives in Sweden

We consulted on the findings with a pediatric oncologist, a young patient council at a public hospital in Sweden, and the Ombudsman for Children in Sweden. All reported their feedback via email. First, the pediatric oncologist reported not missing any aspect in the results. She reported that she considered the findings highly interesting and the biggest takeaway to be the positive effects on adolescents and parents of reading the PAEHRs and that security seemed to be the main cause of worry. Second, the young patient council reported to the first author, after discussion in a meeting, that the findings "looked very good" and dovetailed with their own experiences of having access to the EHRs. They reported that they had nothing to add to our findings. Third, the Ombudsman for Children in Sweden expressed positivity toward this overview as none has so far been done. He had questions about the findings, such as about results that confirmed his suspicions (eg, that male patients were more likely than female patients to consider allowing proxy access), as well as whether there was a complete lack of Swedish studies. He also asked for clarification of one case of unclear wording. An area that he saw as missing was the perceptions of shared access among adolescents and parents. As a result, we clarified some wordings and included the perspective of shared access in the Results and Discussion sections.

Implications of the Findings

Consistent findings can be summarized into four implications for PAEHR implementation: (1) adolescents and parents should be educated on PAEHR use and confidentiality (eg, information visibility for children, adolescents, and parents; possibility to restrict information; reasons for age limits; children's and adolescents' need for privacy; the moment when parents will lose access; and procedures for parents to stay involved in the child's care); (2) HCPs should communicate with EHR vendors and be educated on PAEHRs (eg, use; updates; privacy functionality; and information visibility for children, adolescents, and parents); (3) PAEHRs should be available on mobile devices, and functions need to be integrated; and (4) there should be options on a portal for HCPs and patients to label information as confidential.

XSL•FO
RenderX

Future Research

There is a lack of studies examining the effects of PAEHRs among children and adolescents. Although the Nordic countries are often considered to be at the forefront of PAEHR implementation [1] and access has been available longer at the national level in Sweden than in most other countries, no survey studies targeting a pediatric population in Sweden have been published to date. However, there is ongoing research within the NORDeHEALTH project [1] (with some of the authors' involvement) that aims to rectify this situation. One way is to explore approaches that have already been implemented and conduct comparative studies on the benefits and risks of access or exclusions among children and adolescents. Owing to the current scarcity, investigations with focus on literacy and confidentiality in adolescent outpatient or nonclinical populations are suggested. In addition, there is a need to explore the anticipation of parents and adolescents that shared access may support the transition to adulthood. Furthermore, there is little evidence on the efficiency of PAEHRs in the pediatric population, and work should be undertaken to better understand the effects on documentation time for HCPs and the potential cost-effectiveness of PAEHRs for families and adolescents in the long term. Finally, questions remain with respect to how PAEHRs affect the quality of documentation [102]. In this area, the approach of natural language processing has been increasingly used to quantitatively examine note changes, for example, according to ethnicity and disease chronicity [103].

Conclusions

This study consisted of a scoping review of 74 studies on PAEHRs for parents, children, and adolescents. Most studies (27/74, 36%) were comment papers as, despite the urgency of the matter, there is limited research, particularly regarding adolescents' experiences with web-based access to their records and outside the United States. Existing literature on pediatric PAEHRs indicates a pattern similar to that observed in adult populations, whereby adolescents' and parents' strong interest and positive experiences of accessing the records are juxtaposed with and obstructed by concerns among HCPs and other stakeholders, confidentiality being the key issue. Our findings could inform the design and implementation of future regulations regarding access to PAEHRs. Further examination of the experiences of adolescents, parents, and HCPs is warranted to improve usability and utility, inform universal principles reducing the current arbitrariness in the child's age for own and parental access to EHRs among providers worldwide, and ensure that portals are equipped to safely and appropriately manage a wide variety of patient circumstances.

Acknowledgments

The authors gratefully acknowledge the young patient council at Uppsala University Hospital, the Ombudsman for Children, and Dr Arja Harila for their participation in the consultation of the findings of this review. Furthermore, the authors thank students Isabel Paul and Maria Kirss for their contribution to charting the data.

Authors' Contributions

JH wrote the manuscript, created the tables, and designed the figures. MH, CB, BH, and SH contributed to study design, results analysis, and writing of the Results section. MH was responsible for the conception of the study. All authors read, provided feedback, and approved the paper for submission.

Conflicts of Interest

CB is employed by OpenNotes, a research and advocacy unit that investigates and promotes patients' access to their clinical records.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 168 KB - jmir_v24i11e40328_app1.pdf]

Multimedia Appendix 2
Summary of the included studies.
[PDF File (Adobe PDF File), 281 KB - jmir_v24i11e40328_app2.pdf]

References

1. Hägglund M. Nordic countries lead new initiative on patient access to EHRs. The BMJ Opinion. 2021 May 18. URL: https://blogs.bmj.com/bmj/2021/05/18/maria-hagglund-nordic-countries-lead-new-initiative-on-patient-access-to-ehrs/ [accessed 2022-01-16]
2. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. Opennotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
3. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient access to electronic health records: differences across ten countries. Health Policy Technol 2018 Mar;7(1):44-56. [doi: 10.1016/j.hlpt.2017.11.003]

XSL•FO
RenderX

4. Scandurra I, Pettersson M, Eklund B, Lyttkens L. Analysis of the updated Swedish regulatory framework of the patient accessible electronic health record in relation to usage experience. Stud Health Technol Inform 2017;245:798-802. [Medline: 29295208]
5. Hagström J, Scandurra I, Moll J, Blease C, Haage B, Hörhammer I, et al. Minor and parental access to electronic health records: differences across four countries. Stud Health Technol Inform 2022 May 25;294:495-499. [doi: 10.3233/SHTI220508] [Medline: 35612129]
6. Lam BD, Bourgeois F, DesRoches CM, Dong Z, Bell SK. Attitudes, experiences, and safety behaviours of adolescents and young adults who read visit notes: opportunities to engage patients early in their care. Future Healthc J 2021 Nov 29;8(3):e585-e592 [FREE Full text] [doi: 10.7861/fhj.2021-0118] [Medline: 34888446]
7. Murugan A, Gooding H, Greenbaum J, Boudreaux J, Blanco R, Swerlick A, et al. Lessons learned from opennotes learning mode and subsequent implementation across a pediatric health system. Appl Clin Inform 2022 Jan 26;13(1):113-122. [doi: 10.1055/s-0041-1741483] [Medline: 35081655]
8. Neves AL, Freise L, Laranjo L, Carter AW, Darzi A, Mayer E. Impact of providing patients access to electronic health records on quality and safety of care: a systematic review and meta-analysis. BMJ Qual Saf 2020 Dec 12;29(12):1019-1032 [FREE Full text] [doi: 10.1136/bmjqs-2019-010581] [Medline: 32532814]
9. Moll J, Rexhepi H, Cajander A, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
10. Rexhepi H, Åhlfeldt RM, Cajander A, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: a qualitative study. Health Informatics J 2018 Jun 19;24(2):115-124 [FREE Full text] [doi: 10.1177/1460458216658778] [Medline: 27440056]
11. Britto MT, Hesse EA, Kamdar OJ, Munafo JK. Parents' perceptions of a patient portal for managing their child's chronic illness. J Pediatr 2013 Jul;163(1):280-1.e1. [doi: 10.1016/j.jpeds.2013.02.041] [Medline: 23541773]
12. Campos-Castillo C, Anthony DL. The double-edged sword of electronic health records: implications for patient disclosure. J Am Med Inform Assoc 2015 Apr;22(e1):e130-e140. [doi: 10.1136/amiajnl-2014-002804] [Medline: 25059953]
13. Klein JD, McNulty M, Flatau CN. Adolescents' access to care: teenagers' self-reported use of services and perceived access to confidential care. Arch Pediatr Adolesc Med 1998 Jul 01;152(7):676-682. [doi: 10.1001/archpedi.152.7.676] [Medline: 9667540]
14. Bush RA, Connelly CD, Fuller M, Pérez A. Implementation of the integrated electronic patient portal in the pediatric population: a systematic review. Telemed J E Health 2016 Feb;22(2):144-152 [FREE Full text] [doi: 10.1089/tmj.2015.0033] [Medline: 26258289]
15. Davis K, Drey N, Gould D. What are scoping studies? A review of the nursing literature. Int J Nurs Stud 2009 Oct;46(10):1386-1400. [doi: 10.1016/j.ijnurstu.2009.02.010] [Medline: 19328488]
16. Hagström J, Blease C, Haage B, Scandurra I, Hansson S, Hägglund M. Use of and experiences with online access to electronic health records for parents, children, and adolescents: protocol for a scoping review. JMIR Res Protoc 2022 Jun 15;11(6):e36158 [FREE Full text] [doi: 10.2196/36158] [Medline: 35704386]
17. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Social Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
18. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
20. Beaton L, Williams I, Sanci L. Exploring adolescent and clinician perspectives on Australia. Aust J Prim Health 2021 Mar 15;27(2):102. [doi: 10.1071/PY20169] [Medline: 33715770]
21. Frazer C, Ratchford F, Roch J. Enabling adolescent electronic access to personal health information. Stud Health Technol Inform 2017;234:115-119. [Medline: 28186026]
22. Ransom NR. A teen's perspective: adolescent access to their own electronic medical records. Int J Adolesc Med Health 2016 Feb;28(1):123-124. [doi: 10.1515/ijamh-2015-0081] [Medline: 26656608]
23. Dohil I, Cruz R, Sweet H, Huang JS. Sharing notes with adolescents and young adults admitted to an inpatient psychiatry unit. J Am Acad Child Adolesc Psychiatry 2021 Mar;60(3):317-320. [doi: 10.1016/j.jaac.2020.09.016] [Medline: 33035620]
24. Bergman DA, Brown NL, Wilson S. Teen use of a patient portal: a qualitative study of parent and teen attitudes. Perspect Health Inf Manag 2008;5:13 [FREE Full text] [Medline: 18923702]
25. Carlson JL, Goldstein R, Buhr T, Buhr N. Teenager, parent, and clinician perspectives on the electronic health record. Pediatrics 2020 Mar;145(3):e20190193. [doi: 10.1542/peds.2019-0193] [Medline: 32024749]
26. Bourgeois FC, Fossa A, Gerard M, Davis ME, Taylor YJ, Connor CD, et al. A patient and family reporting system for perceived ambulatory note mistakes: experience at 3 U.S. healthcare centers. J Am Med Inform Assoc 2019 Dec 01;26(12):1566-1573 [FREE Full text] [doi: 10.1093/jamia/ocz142] [Medline: 31504576]
27. Calman N, Pfister HR, Lesnewski R, Hauser D, Shroff N. Electronic access to adolescents' health records: legal, policy, and practice implications. Fam Pract Manag 2015;22(2):11-14 [FREE Full text]

XSL•FO
RenderX

28. Krasowski MD, Grieme CV, Cassady B, Dreyer NR, Wanat KA, Hightower M, et al. Variation in results release and patient portal access to diagnostic test results at an academic medical center. J Pathol Inform 2017;8:45 [FREE Full text] [doi: 10.4103/jpi.jpi_53_17] [Medline: 29226008]
29. Steitz B, Cronin R, Davis S, Yan E, Jackson G. Long-term patterns of patient portal use for pediatric patients at an academic medical center. Appl Clin Inform 2017 Dec 20;08(03):779-793. [doi: 10.4338/aci-2017-01-ra-0005]
30. Thompson LA, Martinko T, Budd P, Mercado R, Schentrup AM. Meaningful use of a confidential adolescent patient portal. J Adolesc Health 2016 Feb;58(2):134-140. [doi: 10.1016/j.jadohealth.2015.10.015] [Medline: 26802988]
31. Wood KE, Greene HR, Krasowski MD. Patient portal activation and use in hospitalized children at an academic medical center. Hosp Pediatr 2021 Jun;11(6):587-594. [doi: 10.1542/hpeds.2020-003707] [Medline: 34006533]
32. Huang JS, Yueh R, Ma S, Cruz R, Bauman L, Choi LJ. Adolescents' and young adults' satisfaction with and understanding of medical notes from a pediatric gastroenterology practice: a cross-sectional cohort study. J Pediatr 2019 Dec;215:264-266. [doi: 10.1016/j.jpeds.2019.06.052] [Medline: 31377044]
33. Ramsey A, Lanzo E, Huston-Paterson H, Tomaszewski K, Trent M. Increasing patient portal usage: preliminary outcomes from the MyChart genius project. J Adolesc Health 2018 Jan;62(1):29-35 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.08.029] [Medline: 29169768]
34. Miklin DJ, Vangara SS, Delamater AM, Goodman KW. Understanding of and barriers to electronic health record patient portal access in a culturally diverse pediatric population. JMIR Med Inform 2019 Apr 26;7(2):e11570 [FREE Full text] [doi: 10.2196/11570] [Medline: 31066681]
35. 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;2016:628-637 [FREE Full text] [Medline: 28269859]
36. Torrens E, Walker SM. Demographic characteristics of Australian health consumers who were early registrants for opt-in personally controlled electronic health records. Health Inf Manag 2017 Sep;46(3):127-133. [doi: 10.1177/1833358317699341] [Medline: 28537210]
37. Schneider H, Hill S, Blandford A. Patients know best: qualitative study on how families use patient-controlled personal health records. J Med Internet Res 2016 Feb 24;18(2):e43 [FREE Full text] [doi: 10.2196/jmir.4652] [Medline: 26912201]
38. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh W. Usability testing finds problems for novice users of pediatric portals. J Am Medical Informatic Assoc 2009 Sep 01;16(5):660-669. [doi: 10.1197/jamia.m3154]
39. Zellmer BM, Nacht CL, Coller RJ, Hoonakker PL, Smith CA, Sklansky DJ, et al. BedsideNotes: sharing physicians' notes with parents during hospitalization. Hosp Pediatr 2021 May;11(5):503-508 [FREE Full text] [doi: 10.1542/hpeds.2020-005447] [Medline: 33795371]
40. Lam BD, Bourgeois F, Dong ZJ, Bell SK. Speaking up about patient-perceived serious visit note errors: patient and family experiences and recommendations. J Am Med Inform Assoc 2021 Mar 18;28(4):685-694 [FREE Full text] [doi: 10.1093/jamia/ocaa293] [Medline: 33367831]
41. Dalrymple PW, Rogers M, Zach L, Luberti A. Understanding internet access and use to facilitate patient portal adoption. Health Informatics J 2018 Dec 29;24(4):368-378 [FREE Full text] [doi: 10.1177/1460458216675497] [Medline: 27821606]
42. Ahlers-Schmidt CR, Nguyen M. Parent intention to use a patient portal as related to their children following a facilitated demonstration. Telemed J E Health 2013 Dec;19(12):979-981. [doi: 10.1089/tmj.2013.0041] [Medline: 24050608]
43. Chung RK, Kim UO, Basir MA. Differing perspectives on parent access to their child's electronic medical record during neonatal intensive care hospitalization: a pilot study. J Matern Fetal Neonatal Med 2018 Apr 10;31(8):1078-1084. [doi: 10.1080/14767058.2017.1306853] [Medline: 28316266]
44. Thompson LA, Black EW, Saliba H, Schentrup AM. Parents' knowledge of and opinions about healthcare laws and technology in primary care. Inform Prim Care 2012 Dec 20;20(1):69-74 [FREE Full text] [doi: 10.14236/jhi.v20i1.49] [Medline: 23336837]
45. Bell SK, Folcarelli P, Fossa A, Gerard M, Harper M, Leveille S, et al. Tackling ambulatory safety risks through patient engagement: what 10,000 patients and families say about safety-related knowledge, behaviors, and attitudes after reading visit notes. J Patient Saf 2021 Dec 01;17(8):e791-e799. [doi: 10.1097/PTS.0000000000000494] [Medline: 29781979]
46. Tom JO, Mangione-Smith R, Solomon C, Grossman DC. Integrated personal health record use: association with parent-reported care experiences. Pediatrics 2012 Jul 11;130(1):e183-e190. [doi: 10.1542/peds.2011-1786] [Medline: 22689872]
47. Byczkowski TL, Munafo JK, Britto MT. Variation in use of internet-based patient portals by parents of children with chronic disease. Arch Pediatr Adolesc Med 2011 May;165(5):405-411. [doi: 10.1001/archpediatrics.2011.55] [Medline: 21536954]
48. Ketterer T, West DW, Sanders VP, Hossain J, Kondo MC, Sharif I. Correlates of patient portal enrollment and activation in primary care pediatrics. Acad Pediatr 2013;13(3):264-271. [doi: 10.1016/j.acap.2013.02.002] [Medline: 23680344]
49. Burke RP, Rossi AF, Wilner BR, Hannan RL, Zabinsky JA, White JA. Transforming patient and family access to medical information: utilisation patterns of a patient-accessible electronic health record. Cardiol Young 2010 May 11;20(05):477-484. [doi: 10.1017/s1047951110000363]

XSL•FO
RenderX

50. Ronis SD, Baldwin CD, McIntosh S, McConnochie K, Szilagyi PG, Dolan J. Caregiver preferences regarding personal health records in the management of ADHD. Clin Pediatr (Phila) 2015 Jul 07;54(8):765-774 [FREE Full text] [doi: 10.1177/0009922814565883] [Medline: 25567294]
51. Smith CA, Coller RJ, Dean SM, Sklansky D, Hoonakker PL, Smith W, et al. Parent perspectives on pediatric unpatient OpenNotes. AMIA Annu Symp Proc 2019;2019:812-819 [FREE Full text] [Medline: 32308877]
52. Kelly MM, Coller RJ, Hoonakker PL, Nacht CL, Dean SM. Provider experiences with offering families bedside health record access across a children's hospital. Hosp Pediatr 2020 Nov;10(11):1002-1005 [FREE Full text] [doi: 10.1542/hpeds.2020-0044] [Medline: 33109520]
53. Kelly MM, Smith CA, Hoonakker PL, Nacht CL, Dean SM, Sklansky DJ, et al. Stakeholder perspectives in anticipation of sharing physicians' notes with parents of hospitalized children. Acad Pediatr 2021 Mar;21(2):259-264 [FREE Full text] [doi: 10.1016/j.acap.2020.11.018] [Medline: 33259951]
54. Bourgeois F, Lowe E, Wachenheim D. OpenNotes: sharing visit notes with patients and families. Pediatric Nurs 2018;44(1):45-48.
55. Kelly MM, Thurber AS, Coller RJ, Khan A, Dean SM, Smith W, et al. Parent perceptions of real-time access to their hospitalized child's medical records using an inpatient portal: a qualitative study. Hosp Pediatr 2019 Apr;9(4):273-280 [FREE Full text] [doi: 10.1542/hpeds.2018-0166] [Medline: 30894398]
56. Asan O, Scanlon MC, Crotty B, Holden RJ, Flynn KE. Parental perceptions of displayed patient data in a PICU. Pediatric Critical Care Med 2019;20(5):435-441. [doi: 10.1097/pcc.0000000000001895]
57. 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]
58. Kelly MM, Hoonakker PL, Dean SM. Using an inpatient portal to engage families in pediatric hospital care. J Am Med Inform Assoc 2017 Jan;24(1):153-161 [FREE Full text] [doi: 10.1093/jamia/ocw070] [Medline: 27301746]
59. Sarabu C, Lee T, Hogan A, Pageler N. The value of opennotes for pediatric patients, their families and impact on the patient-physician relationship. Appl Clin Inform 2021 Jan 10;12(1):76-81 [FREE Full text] [doi: 10.1055/s-0040-1721781] [Medline: 33567464]
60. Edwards EA, Cote A, Phelps AS, Naeger DM. Parents of pediatric radiology patients prefer timely reporting and discussing results with referring providers. Acad Radiol 2020 May;27(5):739-743. [doi: 10.1016/j.acra.2019.07.010] [Medline: 31420160]
61. Gaskin GL, Bruce J, Anoshiravani A. Understanding parent perspectives concerning adolescents' online access to personal health information. J Particip Med 2016;8:e3 [FREE Full text] [Medline: 27595043]
62. Janssen A, Keep M, Selvadurai H, Kench A, Hunt S, Simonds S, et al. Factors that influence use of a patient portal by health professionals. Int J Environ Res Public Health 2021 Feb 15;18(4):1877 [FREE Full text] [doi: 10.3390/ijerph18041877] [Medline: 33671953]
63. Goldstein RL, Anoshiravani A, Svetaz MV, Carlson JL. Providers' perspectives on adolescent confidentiality and the electronic health record: a state of transition. J Adolesc Health 2020 Mar;66(3):296-300. [doi: 10.1016/j.jadohealth.2019.09.020] [Medline: 31831320]
64. Stablein T, Loud KJ, DiCapua C, Anthony DL. The catch to confidentiality: the use of electronic health records in adolescent health care. J Adolesc Health 2018 May;62(5):577-582. [doi: 10.1016/j.jadohealth.2017.11.296] [Medline: 29422435]
65. Wong D, Morgan-Lynch S. Patient portals and young people: addressing the privacy dilemma of providing access to health information. J Prim Health Care 2017;9(4):240. [doi: 10.1071/hc17037]
66. Bialostozky M, Huang JS, Kuelbs CL. Are you in or are you out? Provider note sharing in pediatrics. Appl Clin Inform 2020 Jan 04;11(1):166-171 [FREE Full text] [doi: 10.1055/s-0040-1701679] [Medline: 32131116]
67. Weinstock M. Modern Healthcare. 2018 Apr 7. URL: https://www.modernhealthcare.com/article/20180407/NEWS/180409956/q-a-with-gina-altieri-on-nemours-children-s-health-system-s-digital-agenda [accessed 2022-11-14]
68. Sharko M, Wilcox L, Hong MK, Ancker JS. Variability in adolescent portal privacy features: how the unique privacy needs of the adolescent patient create a complex decision-making process. J Am Med Inform Assoc 2018 Aug 01;25(8):1008-1017 [FREE Full text] [doi: 10.1093/jamia/ocy042] [Medline: 29788423]
69. Anoshiravani A, Gaskin GL, Groshek MR, Kuelbs C, Longhurst CA. Special requirements for electronic medical records in adolescent medicine. J Adolesc Health 2012 Nov;51(5):409-414. [doi: 10.1016/j.jadohealth.2012.08.003] [Medline: 23084160]
70. Jasik CB. Unlocking the potential of the patient portal for adolescent health. J Adolesc Health 2016 Feb;58(2):123-124. [doi: 10.1016/j.jadohealth.2015.12.003] [Medline: 26802987]
71. Lee JA, Miller SD, Mezoff EA, Screws J, Sauer C, Huang JS. The 21st Century CURES act in pediatric gastroenterology: problems, solutions, and preliminary guidance. J Pediatr Gastroenterol Nutr 2021 May 01;72(5):700-703. [doi: 10.1097/MPG.0000000000003117] [Medline: 33720090]
72. Spooner SA, Council on Clinical Information Technology, American Academy of Pediatrics. Special requirements of electronic health record systems in pediatrics. Pediatrics 2007 Mar;119(3):631-637. [doi: 10.1542/peds.2006-3527] [Medline: 17332220]

XSL•FO
RenderX

73. Gracy D, Weisman J, Grant R, Pruitt J, Brito A. Content barriers to pediatric uptake of electronic health records. Adv Pediatr 2012;59(1):159-181. [doi: 10.1016/j.yapd.2012.04.004] [Medline: 22789578]
74. Park K, Park MD, Longhurst CA. Patient and family access to electronic health records: a key ingredient for a pediatric learning health system. J Particip Med 2015;7:e2 [FREE Full text] [Medline: 25664199]
75. Green-Shook S. Parental proxy access via web portals. J AHIMA 2009 Jul;80(7):60-61. [Medline: 19663148]
76. Bourgeois FC, Taylor PL, Emans SJ, Nigrin DJ, Mandl KD. Whose personal control? Creating private, personally controlled health records for pediatric and adolescent patients. J Am Medical Informatic Assoc 2008 Nov 01;15(6):737-743. [doi: 10.1197/jamia.m2865]
77. Berlan ED, Bravender T. Confidentiality, consent, and caring for the adolescent patient. Curr Opin Pediatr 2009 Aug;21(4):450-456. [doi: 10.1097/MOP.0b013e32832ce009] [Medline: 19474734]
78. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
79. Swartz MK. Protecting the privacy rights of adolescents. J Pediatr Health Care 2013 May;27(3):161. [doi: 10.1016/j.pedhc.2012.11.007] [Medline: 23611455]
80. Bayer R, Santelli J, Klitzman R. New challenges for electronic health records: confidentiality and access to sensitive health information about parents and adolescents. JAMA 2015 Jan 06;313(1):29-30. [doi: 10.1001/jama.2014.15391] [Medline: 25562260]
81. Nielsen BA. Confidentiality and electronic health records: keeping up with advances in technology and expectations for access. Clin Practice Pediatric Psychol 2015 Jun;3(2):175-178. [doi: 10.1037/cpp0000096]
82. Taylor JF, Williams RL, Blythe BJ. Healthcare reform, EHRs, and adolescent confidentiality. Contemporary OB/GYN. 2015 Jul 29. URL: https://www.contemporaryobgyn.net/view/healthcare-reform-ehrs-and-adolescent-confidentiality [accessed 2022-11-03]
83. Kendrick EJ, Benson C. Patient portals in child and adolescent psychiatry. Child Adolesc Psychiatr Clin N Am 2017 Jan;26(1):43-54. [doi: 10.1016/j.chc.2016.07.005] [Medline: 27837941]
84. Confidentiality in adolescent health care: ACOG Committee Opinion, Number 803. Obstet Gynecol 2020 Apr;135(4):e171-e177. [doi: 10.1097/AOG.0000000000003770] [Medline: 32217979]
85. Ancker JS, Sharko M, Hong M, Mitchell H, Wilcox L. Should parents see their teen's medical record? Asking about the effect on adolescent-doctor communication changes attitudes. J Am Med Inform Assoc 2018 Dec 01;25(12):1593-1599 [FREE Full text] [doi: 10.1093/jamia/ocy120] [Medline: 30247699]
86. Society for Adolescent HealthMedicine, Gray SH, Pasternak RH, Gooding HC, Woodward K, Hawkins K, et al. Recommendations for electronic health record use for delivery of adolescent health care. J Adolesc Health 2014 Apr;54(4):487-490. [doi: 10.1016/j.jadohealth.2014.01.011] [Medline: 24656534]
87. Williams RL, Taylor JF. Four steps to preserving adolescent confidentiality in an electronic health environment. Curr Opin Obstet Gynecol 2016 Oct;28(5):393-398. [doi: 10.1097/GCO.0000000000000305] [Medline: 27454851]
88. Bourgeois FC, DesRoches CM, Bell SK. Ethical challenges raised by OpenNotes for pediatric and adolescent patients. Pediatrics 2018 Jun 18;141(6):e20172745. [doi: 10.1542/peds.2017-2745] [Medline: 29776979]
89. Taylor P, Bourgeois FC, Mandl KD. Access controls for a pediatric personally controlled health record. AMIA Annu Symp Proc 2007 Oct 11:1131. [Medline: 18694228]
90. Sherek PD, Gray E. Case study: managing pediatric health information in a patient portal. J AHIMA 2014 Apr;85(4):46-47. [Medline: 24834555]
91. Knopf A. Advocating for your child in the digital age. The Brown University Child Adolescent Behavior Letter 2021 Feb 03;37(3):9-10. [doi: 10.1002/cbl.30531]
92. Schapiro NA, Mihaly LK. The 21st Century cures act and challenges to adolescent confidentiality. J Pediatr Health Care 2021 Jul;35(4):439-442. [doi: 10.1016/j.pedhc.2021.03.005] [Medline: 33865680]
93. Schwarz J, Bärkås A, Blease C, Collins L, Hägglund M, Markham S, et al. Sharing clinical notes and electronic health records with people affected by mental health conditions: scoping review. JMIR Ment Health 2021 Dec 14;8(12):e34170 [FREE Full text] [doi: 10.2196/34170] [Medline: 34904956]
94. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
95. 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]
96. Schultz CL, Alderfer MA. Are on-line patient portals meeting test result preferences of caregivers of children with cancer? A qualitative exploration. Pediatr Blood Cancer 2018 Nov 14;65(11):e27306. [doi: 10.1002/pbc.27306] [Medline: 30007016]
97. Petersson L, Erlingsdóttir G. Open Notes in Swedish Psychiatric Care (Part 2): survey among psychiatric care professionals. JMIR Ment Health 2018 Jun 21;5(2):e10521 [FREE Full text] [doi: 10.2196/10521] [Medline: 29929946]
98. Petersson L, Erlingsdóttir G. Open Notes in Swedish Psychiatric Care (Part 1): survey among psychiatric care professionals. JMIR Ment Health 2018 Feb 02;5(1):e11 [FREE Full text] [doi: 10.2196/mental.9140] [Medline: 29396386]

XSL•FO
RenderX

99. Blease C, Salmi L, Rexhepi H, Hägglund M, DesRoches CM. Patients, clinicians and open notes: information blocking as a case of epistemic injustice. J Med Ethics 2021 May 14;48(10):785-793 [FREE Full text] [doi: 10.1136/medethics-2021-107275] [Medline: 33990427]
100. Duncan RE, Vandeleur M, Derks A, Sawyer S. Confidentiality with adolescents in the medical setting: what do parents think? J Adolesc Health 2011 Oct;49(4):428-430. [doi: 10.1016/j.jadohealth.2011.02.006] [Medline: 21939875]
101. Ip W, Yang S, Parker J, Powell A, Xie J, Morse K, et al. Assessment of prevalence of adolescent patient portal account access by guardians. JAMA Netw Open 2021 Sep 01;4(9):e2124733 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.24733] [Medline: 34529064]
102. Blease C, Torous J, Hägglund M. Does patient access to clinical notes change documentation? Front Public Health 2020 Nov 27;8:577896 [FREE Full text] [doi: 10.3389/fpubh.2020.577896] [Medline: 33330320]
103. Himmelstein G, Bates D, Zhou L. Examination of stigmatizing language in the electronic health record. JAMA Netw Open 2022 Jan 04;5(1):e2144967 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.44967] [Medline: 35084481]

Abbreviations

EHR: electronic health record
HCP: health care professional
PAEHR: patient-accessible electronic health record
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

Edited by R Kukafka; submitted 23.06.22; peer-reviewed by C Grünloh, G Myreteg, T Risling, J Maxwell; comments to author 22.09.22; revised version received 12.10.22; accepted 25.10.22; published 22.11.22.

Please cite as:
Hagström J, Blease C, Haage B, Scandurra I, Hansson S, Hägglund M
Views, Use, and Experiences of Web-Based Access to Pediatric Electronic Health Records for Children, Adolescents, and Parents: Scoping Review
J Med Internet Res 2022;24(11):e40328
URL: https://www.jmir.org/2022/11/e40328
doi:10.2196/40328
PMID:36413382

©Josefin Hagström, Charlotte Blease, Barbara Haage, Isabella Scandurra, Scharlett Hansson, Maria Hägglund. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing the Content and Quality of Digital Tools for Managing Gestational Weight Gain: Systematic Search and Evaluation

Bonnie R Brammall[1], MHumNutr; Rhonda M Garad[1], PhD; Jacqueline A Boyle[1,2], MBBS, PhD; Melanie J Hayman[3], PhD; Susan J de Jersey[4,5], PhD; Helena J Teede[1,6], MBBS, PhD; Quoc V Hong[1], MBBS; Alayna Carrandi[1], BSc; Cheryce L Harrison[1,6], PhD

[1]Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia
[2]Department of Obstetrics and Gynecology, Monash Health, Clayton, Australia
[3]Appleton Institute, School of Health, Medical and Applied Sciences, Central Queensland University, Rockhampton, Australia
[4]Department of Nutrition and Dietetics, The Royal Brisbane and Women's Hospital, Metro North Health, Herston, Australia
[5]Perinatal Research Centre, Centre for Clinical Research, Faculty of Medicine, The University of Queensland, Herston, Australia
[6]Diabetes and Vascular Research, Monash Health, Clayton, Australia

Corresponding Author:
Cheryce L Harrison, PhD
Monash Centre for Health Research and Implementation
School of Public Health and Preventive Medicine
Monash University
Kanooka Grove
Clayton, 3168
Australia
Phone: 61 3 8572 2662
Email: cheryce.harrison@monash.edu

Abstract

Background: Digital health resources have the potential to assist women in optimizing gestational weight gain (GWG) during pregnancy to improve maternal health outcomes.

Objective: In this study, we aimed to evaluate the quality and behavior change potential of publicly available digital tools (websites and apps) that facilitate GWG tracking.

Methods: Digital tools were identified using key search terms across website search engines and app stores and evaluated using the Mobile App Rating Scale, the App Behavior Change Scale, as well as criteria to evaluate the rigor and safety of GWG information.

Results: Overall, 1085 tools were screened for inclusion (162 websites and 923 apps), and 19 were deemed eligible. The mean Mobile App Rating Scale quality score was 3.31 (SD 0.53) out of 5, ranging from 2.26 to 4.39, and the mean App Behavior Change Scale score was 6 (SD 3.4) out of 21, ranging from 19 to 0. Of the 19 items used to evaluate rigor of GWG advice, most tools (n=11, 57.9%) contained ≤3 items.

Conclusions: This review emphasizes the substantial limitations in current digital resources promoting the monitoring and optimization of GWG. Most tools were of low quality, had minimal behavior change potential, and were potentially unsafe, with minimal linkage to evidence-based information or partnership with health care.

(J Med Internet Res 2022;24(11):e37552) doi:10.2196/37552

KEYWORDS

digital; gestational; weight; tracking; pregnancy

XSL•FO
RenderX

Introduction

Gestational Weight Gain: An Overview

During pregnancy, gestational weight gain (GWG) is essential to ensure the development of a healthy fetus [1]. However, GWG below or above the recommendations is associated with an increased risk of negative pregnancy outcomes and neonatal conditions or complications [1-6]. Epidemiological data in over 1 million pregnancies globally reported GWG below or above the recommended thresholds in 23% and 47% of all pregnancies, respectively [7]. The associated risks of GWG below recommendations include preterm birth and the delivery of a small-for-gestational-age infant, whereas excessive GWG above recommendations was associated with cesarean section, macrosomia, and the delivery of a large-for-gestational-age infant [6-8]. Long-term, excessive GWG is associated with intergenerational adverse health risks, including obesity, cardiovascular disease, and type 2 diabetes [1-5]. Therefore, optimizing GWG during pregnancy in line with recommendations is a global health priority.

Digital Health Engagement During Pregnancy

Digital health, including internet-based information and mobile Health (mHealth) apps, have become popular and widely used sources of health information for pregnant women, often replacing traditional paper-based and supplementing face-to-face health professional consultations [9-12]. However, the attainment of credible internet health or mHealth information is reliant on consumer health literacy and the ability to judge the quality and accuracy of information. Given the tendency of consumers to trust digital health information [13], this is problematic, as health information is not always reliable or current and can be confusing, overwhelming, and at times potentially harmful [12].

During pregnancy, freely accessible web-based resources, including trackers, calculators, or graphs, to record and self-monitor GWG have the potential to assist women in identifying whether weight gain is outside the recommended thresholds. In conjunction with the promotion of healthy lifestyle behaviors, these web-based resources have the potential to assist women in achieving healthy GWG [14-16]. However, there is currently limited information about the type of tools available, their format (ie, web-based application or mobile app) and functionality, credibility of the information provided, or their ability to guide behavior change to positively impact GWG. Evaluating digital tools that are publicly available to women to monitor GWG during pregnancy is a critical gap to address, given the risk of complications associated with excessive or inadequate GWG and the need to ensure credible and reliable self-monitoring tools for women during this time. Previous research in this area is limited to evaluations of mobile apps only and is primarily based on functionality [17] or a narrow evaluation of selected apps based on predefined pregnancy topics [18].

In this study, we aimed to evaluate the quality and behavior change potential of publicly available digital tools (websites and apps) that facilitate GWG tracking. Given the benefits of self-weighing for weight management [16] and the high use of digital health information during pregnancy [9-11], there is a need to examine and review what is currently available to ensure that pregnant women are being provided with evidence-based information and tools that align with GWG recommendations.

Methods

The methods of this study have been informed by previous reviews exploring the quality, features, functions, behavior change capacity, and quality of digital applications and resources [19-23].

Systematic Search

Searches were conducted in an Australian web browser using website search engines (Google, BING, and Yahoo) and mobile app stores (Apple AppStore, iOS and Google Play, Android) using a combination of search terms emulating terms likely used by end users, including *pregnancy weight*, *pregnancy weight tracker*, *pregnancy weight gain calculator*, *pregnancy weight graph* (website searches), and *pregnancy weight*, *pregnancy weight tracker,* and *gestational weight tracker* (app searches). Search terms were developed by a multidisciplinary team comprising obstetrics and gynecology (O&G), midwifery, nursing, dietetics and nutrition, and exercise physiology. Each search term combination was entered individually in the search engine. For websites, the first 2 pages of results for each search term were screened for inclusion, similar to previous studies [19,24,25]. For apps, searches were entered into the Google Play and Apple App Store databases without any specified search categories. All the retrieved app search results were screened for inclusion. One reviewer (BRB) independently reviewed all results, with a 100% cross-check of websites and 50% cross-check of apps completed by 2 additional independent reviewers (CLH and RMG).

Inclusion Criteria and Selection Process

Websites and apps were included according to the following criteria: publicly available or ability to download (free or paid, but with free discovery capacity); written in or available in English; title or description suggested inclusion of tools or advice or resources relating to pregnancy weight gain; and weight-tracking tool enabled multiple logs or entries of weight across pregnancy (ie, not just 1 static weight log).

Apps that met the inclusion criteria were further filtered using the following app-specific inclusion criteria: updated within 18 months from the search date, May 2021; user rating of ≥4.0 stars if ≥6 months old (apps <6 months were included irrespective of user rating) as a proxy for app popularity per previous research [21]; incorporation of a graph or chart or illustration of GWG (ie, does not merely display the weight as a numerical value); and presence of surrounding content about pregnancy health and well-being. Apps that required downloading to complete this step were screened for inclusion by 2 researchers (BRB and RMG). If the apps available on Google Play and Apple AppStore had contrasting user ratings, the higher rating was carried forward and documented in the app description results.

XSL•FO
RenderX

Resource Evaluation

Overview

Eligible websites and apps were randomly allocated to 2 reviewers and independently reviewed on a mobile device. All reviewers (AC, BRB, MJH, QVH, RMG, and SJdJ) have expertise in public health and form a multidisciplinary team (ie, O&G, midwifery, nursing, dietetics and nutrition, and exercise physiology). Where the same app was available on both Google Play Store and Apple App Store, app details and descriptions were reviewed to ensure consistency across the 2 platforms and downloaded for review on an Apple device. The reviews were conducted from June to July 2021. Apps were user tested for evaluation using numerous validated scales and relevant questions (Multimedia Appendices 1-4) using a mock user profile. Each app was explored until the reviewer had familiarized themselves with the functionality and features of the app, with a user experience consistent with other studies [21]. Reviewers noted whether the app stopped functioning or whether the features were not accessible. Following the review, if there was a contradiction in reviewer responses, a third independent reviewer was assigned to resolve item or items of disagreement and establish consensus (BRB, CLH, and RMG).

Collections of user demographic and pregnancy-specific data were recorded, including username; contact details (name, email, phone, or other); date of birth or age; country of origin; gestation (due date, last menstrual cycle, or date of conception); type of pregnancy (singleton, twin, triplet, etc); parity (first, second, third, etc); and preconception weight and height.

GWG Criteria

To evaluate the rigor and safety aspects of GWG management information, GWG-specific criteria were developed by a multidisciplinary team (Multimedia Appendix 1). The criteria encompassed 19 items, including reference to published international guidelines for GWG (ie, National Academy of Medicine, previously Institute of Medicine [26]) with personalization according to BMI; warnings, notifications, or alerts for weight gain detected outside of recommendations; direction or advice to consult a health professional if logged GWG was outside of the recommendations; and dietary and physical activity content and the development of content in consultation with relevant health professionals (O&G, midwifery, allied health, etc).

Mobile App Rating Scale

The Mobile App Rating Scale (MARS) is a 23-item evaluation tool comprising 6 domains (Multimedia Appendix 2): *engagement*, *functionality*, *aesthetics*, *information quality*, *subjective quality*, and *health topic specific* [27,28]. Each item is scored using a 5-point ordinal scale, with a mean score derived for each domain. The first 4 domains, including *engagement* (ie, incorporation of interesting, customizable, and interactive—eg, sends alerts, messages, reminders, and feedback and enables sharing—features targeted at the audience); *functionality* (ie, ease of use, navigation, flow logic, and gestural design); *aesthetics* (ie, graphic design, overall visual appeal, color scheme, and stylistic consistency); and *information quality* (ie, contains high-quality information from a credible source), are combined and averaged to provide an overall app quality score out of 5. A *subjective quality* score between 0 and 20 is allocated by each reviewer. This section requires the reviewer to rate whether they would recommend the app to people who may benefit from using the app, how many times over 12 months they would use the app if it was relevant to them, whether they would be willing to pay for the app, and their overall app star rating. The *health topic–specific* domain is an optional 5-item section that can be adjusted to suit the topic area researched (ie, GWG). This domain aims to assess whether the app is likely to *increase awareness of the importance of addressing GWG*, *increase knowledge or understanding of GWG*, *change attitudes toward improving GWG*, *increase intention or motivation to address GWG*, and *encourage further help seeking for GWG*.

The MARS also includes an App Classification section to obtain information about technical features (Multimedia Appendix 2). These items were recorded for descriptive purposes but did not form part of the functionality rating. These features include the app rating, obtained via the Google Play or Apple App Store; the number of app downloads (derived from the Google Play Store only as of August 2021; the Apple App Store does not provide app download information, so this information is precluded); whether the digital tool presented or required agreement to terms and conditions or a disclaimer; required log-in; allowed password protection; allowed sharing to social media; allowed data export; had an app community; sent reminders; required web access to function; and whether the digital tool sent push notifications. All applicable criteria were used for website evaluation, excluding ratings and downloads.

The App Behavior Change Scale

The App Behavior Change Scale (ABACUS) is designed to evaluate the behavior change potential of smartphone apps and websites across 4 domains (Multimedia Appendix 3) [22,23]. These include knowledge and information (ie, customized and personal features, collection of baseline information, and consequences for continuing or discontinuing behavior); goals and planning (ie, goal setting, goal reviewing, updating or changing, and willingness for behavior change); feedback and monitoring (ie, easy-to-use self-monitoring tools and data exporting or rewards or incentives); and actions (ie, reminders, prompts or cues, planning for barriers, and assistance with distractions or avoidance).

Quality Evaluation

Criteria to evaluate the quality of the health-related digital tools were developed (Multimedia Appendix 4) and modified from app review studies in the field authored by our group [19,20]. The criteria include statement of purpose of the app or website; contact details provided (email, phone, or fax); ownership disclosure (who owns the app or website); copyright; general disclosures; general disclaimers; advertisement disclosures; sponsorship disclosures; author or developer disclosures; author or developer credentials (credentials and affiliations); independence of sponsors or funders; references provided; and type of references provided (a list of types provided, including meta-analysis, randomized controlled trial, media, government guideline, or option piece).

XSL•FO
RenderX

Statistical Analysis

Descriptive statistics (mean and SD) and frequencies (numbers and percentages) were calculated for all scales applied. The reported percentages were rounded to the nearest whole number. Intraclass correlation (ICC) scores were calculated to determine the agreement between the MARS rating using SPSS statistical software (version 25; IBM Corp). All analyses were conducted using SPSS for Windows, with a significance level set at $P<.05$. The following previously established categories for expressing levels of reliability for ICC results were used: high reliability, 0.90 to 0.99; good reliability, 0.80 to 0.89; fair reliability, 0.70 to 0.79; and poor reliability, 0.69 or less [29].

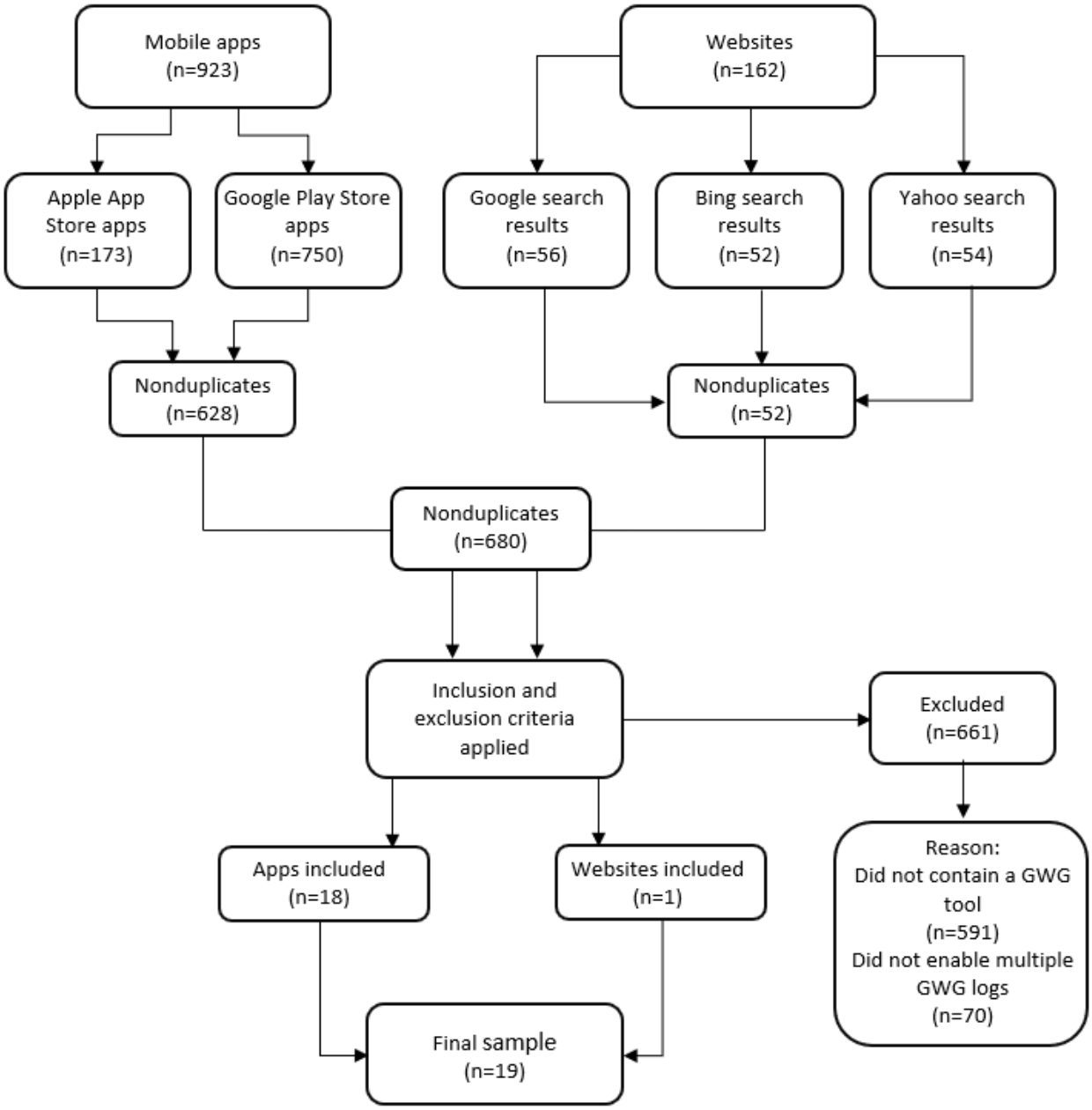
Ethical Considerations

This study does not meet the criteria for human research and thus did not require oversight from the authors' institutions.

Results

A total of 1085 digital tools were screened for inclusion across 162 websites and 923 apps. After excluding duplicates, 89 digital tools were retained for potential inclusion with 19 digital tools eligible for analysis (Figure 1).

Figure 1. Flowchart of gestational weight gain (GWG) digital tool selection.



Characteristics and Overview of Digital Tools

Table 1 presents the main characteristics of the websites and apps included in this study; further descriptions of tools are available in Multimedia Appendix 5. All apps (n=18) were available on the Google Play Store and 9 were available on both the Google Play and Apple App Store. The 18 apps had a Google Play or Apple App Store user-rating score ranging from 4.10 to 4.90, with a mean score of 4.64 (SD 0.22), and had been

XSL•FO
RenderX

downloaded over 25 million times from the Google Play Store alone. Most digital tools were associated with commercial enterprises (17/19, 89%), whereas few were affiliated with government services (1/19, 5%) and universities (1/19, 5%). All apps had a free discovery capacity (ability to download and use without payment), with total downloads per app ranging from >500 to >10,000,000. Overall, 50% (9/18) of apps had costs for app subscription and in-app purchases ranging from Aus $1.99 to $79.99 (US $1.49 to $59.99; Multimedia Appendix 5); however, this did not impact the discoverability of content or tools reviewed. The website (1/1, 100%) was free to access.

All digital tools were based on information or education (19/19, 100%) and monitoring or tracking (19/19, 100%), and the majority included advice, tips, and strategies (15/19, 79%). A small number of tools used assessment (3/19, 16%), feedback (3/19, 15%), and goal setting (1/19, 5%). Technical aspects included reminders (11/19, 58%), log-in requirements (11/19, 58%), app communities (5/19, 26%), password protection (4/19, 21%), and sharing options (eg, social media, app to app, or email; 3/19, 16%). Only the website required web access to function, with all apps able to be used offline. All collected information about gestation (19/19, 100%) and most, but not all, collected preconception weight (16/19, 84%) and height (14/19, 74%; Table 1).

XSL•FO
RenderX

Table 1. Technical aspects and characteristics of digital tools for GWG management.

	Value, n (%)	App[a]																		Web[b]
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	01
Theoretical background or strategies																				
Advice or tips or strategies or skills training	14 (74)	✓[c]	✓			✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓
Assessment	4 (16)			✓			✓													✓
Feedback	4 (16)						✓											✓		✓
Goal setting	1 (5)						✓													
Information or education	18 (95)	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Monitoring or tracking	19 (100)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Technical aspects																				
Allows sharing (social media, app to app, or email)	4 (21)				✓		✓						✓							✓
App community	5 (26)					✓	✓	✓		✓						✓				
Needs web access to function	1 (5)																			✓
Password protected	3 (16)					✓	✓									✓				
Requires log-in	10 (53)		✓	✓		✓	✓					✓	✓			✓	✓	✓	✓	
Sends reminders	11 (58)		✓	✓	✓	✓	✓		✓	✓	✓		✓		✓		✓			

XSL•FO
RenderX

	Value, n (%)	App[a]																		Web[b]
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	01
Information collected																				
Contact details	9 (47)		✓		✓	✓	✓		✓			✓	✓			✓		✓		
Country or location	4 (21)					✓	✓		✓				✓							
Date of birth or age	6 (32)		✓			✓	✓		✓				✓		✓					
Name	11 (58)	✓	✓		✓	✓	✓		✓				✓			✓	✓	✓	✓	
Pregnancy-related information collected																				
Gestation	19 (100)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Height	14 (74)	✓		✓		✓	✓		✓	✓	✓	✓		✓		✓	✓	✓	✓	✓
Pregnancy number (ie, first or second etc)	4 (21)		✓			✓	✓		✓											
Pregnancy type (single or twins etc)	3 (16)					✓	✓													✓
Preconception weight	15 (79)	✓	✓	✓		✓	✓		✓	✓	✓	✓		✓		✓	✓	✓	✓	✓

[a]App: apps included in ths study.

[b]Web: website included in this study.

[c]✓: indicates technical aspects or characteristics present in the digital tool.

GWG Criteria

Gestational weight tracking was a major feature of most digital tools, displayed prominently to users (15/19, 79%), in line with our inclusion criteria (Multimedia Appendix 6). In total, 58% (11/19) of digital tools provided weight recommendations based on preconception weight and height. All other criteria were present in less than half of the digital tools. Overall, 47% (9/19) of tools encouraged an unspecified, healthy diet for optimal GWG, and 37% (7/19) encouraged nonspecific, regular moderate physical activity for optimal GWG. Very few (2/19, 11%) tools alerted the user when their weight gain was outside of the recommended range, and none directed the user to consult a health professional if their weight entry was outside the recommended range. Overall, of the 19 criteria, the majority (11/19, 58%) contained ≤3 items, with 11% (2/19) having 0 items. The tool that met the most criteria for GWG was Web01 (9 of 19 criteria), followed by App17 (7 of 19 criteria) and App06 (6 of 19 criteria); the name and description of tools can be viewed in Multimedia Appendix 5. Refer to Multimedia Appendix 1 for the complete GWG criteria and definitions.

MARS Results

The specific MARS scores for each digital tool are presented in Table 2. The overall mean MARS quality score (comprising engagement, functionality, aesthetics of tool, and the quality of

XSL•FO
RenderX

general pregnancy-related information domains) ranged from 2.26 (lowest-rated app) to 4.39 (highest-rated app), with a mean score of 3.17 (SD 0.75). Subjective ratings (ie, reviewer recommendations, rating, and perceived monetary value) ranged from mean 3.25 (SD 0.00) to mean 15.50 (SD 0.71), from a potential score of 20; app-specific ratings (ie, GWG awareness, knowledge, and understanding of GWG; attitudes toward improving GWG; intention and motivation to address GWG; and help seeking for GWG) ranged from 1.00 to 4.50, with most (15/19, 79%) scores being 2.50 or less. Overall, the best-rated section was functionality (mean 3.94, SD 0.63), followed by aesthetics (mean 3.61, SD 0.69) and engagement (mean 3.19, SD 0.63), compared with app-specific (mean 2.24, SD 0.84) and information (mean 2.49, SD 0.68) domains, which scored the lowest. ICC scores ranged from 0.671 (95% CI −0.169 to 0.946) to 0.996 (95% CI 0.076-0.999). Most ICC results showed either high (10/19, 53%) or good (6/19, 32%) reliability.

Table 2. Mobile App Rating Scale (MARS) scoring.

App or website name	Overall MARS quality score (A-D), mean (SD)	A (engagement), mean (SD)	B (functionality), mean (SD)	C (aesthetics), mean (SD)	D (information), mean (SD)	E (subjective), mean (SD)	F (app specific), mean (SD)	ICC[a] (95% CI)
App06	4.39 (0.54)	4.50 (0.71)	4.50 (0.71)	4.50 (0.24)	4.07 (0.51)	15.50 (0.71)	2.70 (2.40)	0.935 (0.615 to 0.991)
App12	4.07 (0.15)	3.80 (0.57)	4.75 (0.35)	5.00 (0.00)	2.72 (0.40)	13.00 (0.00)	2.30 (0.42)	0.973 (0.824 to 0.996)
Web01	4.00 (0.18)	3.20 (0.28)	5.00 (0.00)	4.00 (0.47)	3.79 (0.91)	12.50 (3.54)	4.50 (0.14)	0.836 (0.222 to 0.975)
App08	3.60 (0.04)	4.10 (0.14)	4.00 (0.00)	4.00 (0.00)	2.29 (0.00)	3.25 (0.00)	2.60 (0.57)	0.996 (0.976 to 0.999)
App02	3.56 (0.11)	3.30 (0.42)	4.13 (0.88)	4.34 (0.47)	2.50 (0.51)	8.00 (1.41)	2.00 (0.57)	0.858 (0.349 to 0.981)
App01	3.54 (0.01)	3.10 (0.14)	4.25 (0.00)	4.00 (0.00)	2.79 (0.10)	8.50 (0.71)	2.50 (0.42)	0.972 (0.817 to 0.996)
App14	3.51 (0.24)	3.50 (0.14)	3.50 (0.35)	4.34 (0.94)	2.72 (0.21)	10.00 (2.83)	2.30 (0.99)	0.856 (0.285 to 0.978)
App05	3.41 (0.06)	3.70 (0.14)	3.88 (0.18)	3.33 (0.00)	2.72 (0.21)	11.50 (0.71)	3.20 (0.28)	0.972 (0.999 to 0.817)
App03	3.39 (0.27)	3.10 (0.42)	4.25 (0.35)	4.00 (0.00)	2.22 (0.30)	7.50 (2.12)	1.70 (0.42)	0.873 (0.349 to 0.981)
App17	3.38 (0.06)	3.60 (0.00)	4.00 (0.00)	3.50 (0.24)	2.43 (0.00)	12.50 (0.71)	3.50 (0.14)	0.995 (0.962 to 0.999)
App09	3.34 (0.08)	3.20 (0.57)	3.63 (0.53)	3.33 (0.00)	3.22 (0.30)	9.50 (0.71)	2.60 (0.57)	0.957 (0.729 to 0.994)
App15	3.25 (0.10)	3.50 (0.71)	3.50 (0.00)	3.50 (0.24)	2.50 (0.10)	10.00 (1.41)	1.60 (0.28)	0.950 (0.693 to 0.993)
App07	3.14 (0.41)	3.20 (0.28)	3.88 (0.18)	3.50 (1.17)	2.00 (0.00)	7.00 (1.41)	1.60 (0.57)	0.859 (0.296 to 0.979)
App11	2.96 (0.69)	2.90 (0.71)	3.50 (0.71)	3.00 (0.95)	2.43 (0.40)	5.50 (2.12)	1.70 (0.14)	0.711 (−0.095 to 0.954)
App10	2.85 (0.62)	2.10 (0.14)	4.38 (0.88)	3.33 91.41)	1.57 (0.00)	6.00 (2.83)	1.30 (0.42)	0.713 (−0.090 to 0.954)
App13	2.84 (0.01)	2.60 (0.57)	3.88 (0.18)	3.00 (0.00)	1.86 (0.40)	7.00 (0.00)	1.90 (0.14)	0.972 (0.815 to 0.996)
App16	2.75 (0.02)	2.80 (0.00)	3.13 (0.18)	2.84 (0.23)	2.22 (0.50)	5.50 (0.71)	1.50 (0.71)	0.864 (0.315 to 0.980)
App18	2.60 (0.83)	2.50 (0.99)	3.63 (0.53)	2.50 (0.71)	1.79 (1.11)	6.50 (3.54)	2.00 (1.13)	0.671 (−0.169 to 0.946)
App04	2.26 (0.26)	2.00 (0.57)	3.13 (0.18)	2.50 (0.24)	1.43 (0.40)	4.00 (0.00)	1.00 (0.00)	0.938 (0.627 to 0.991)

[a]ICC: intraclass correlation; agreement between reviewers (A-F).

XSL•FO
RenderX

ABACUS Results

The overall ABACUS score was 6 (SD 3.6) of 21 (Table 3). Four behavior change techniques were most prominent, which were included in >50% of the apps. These techniques or functions included the ability to customize and personalize some features (19/19, 100%), the collection of baseline information (ie, user information or personal details; 16/19, 84%), allowing the user to easily self-monitor behavior (13/19, 68%) and providing instructions on how to perform a behavior (10/19, 53%). These and other didactic or simple techniques such as instructions, data export, and sending of reminders were much more frequent than interactive functions such as goal setting (1/19, 5%), encouragement (0/19, 0%), providing material or social rewards (0/19, 0%), and ascertaining willingness to change (0/19, 0%). The top tools for behavior change potential were App06 (scoring 16/21), App08 (scoring 11/21), App17 (scoring 9/21), and Web01 (scoring 9/21).

Table 3. Performance on App Behavior Change Scale (ABACUS) criteria (most to least frequently used).

Behavior change technique[a]	Value, n (%)
Customize and personalize some features	19 (100)
Baseline information	16 (84)
Allow the user to easily self-monitor behavior	13 (68)
Provide instruction on how to perform the behavior	10 (53)
Reminders or prompts or cues for activity (on app)	8 (42)
Data export	7 (37)
Information provided about the consequences of continuing or discontinuing behavior	7 (37)
Give user feedback (person or automatic)	5 (26)
Allow or encourage practice or rehearsal in addition to daily activities	4 (21)
Created with expertise or information consistent with national guidelines	4 (21)
Restructure the physical or social environment	4 (21)
Encourage positive habit formation	3 (16)
Provide the opportunity to plan for barriers	3 (16)
Share behaviors with others or allow for social comparison	3 (16)
Understand the difference between current action and future goals	3 (16)
Distraction or avoidance	2 (11)
Review goals, update, and change	2 (11)
Goal setting	1 (5)
Provide general encouragement	0 (0)
Material or social reward or incentive	0 (0)
Willingness for behavior change	0 (0)

[a]App Behavior Change Scale average score: mean 6 (SD 4) out of 21.

Quality Evaluation

Most (16/19, 84%) digital tools had a statement of purpose and all, with the exception of one (18/19, 95%), provided developer or author contact details. Ownership disclosure and copyright statements (14/19, 78%), advertisement disclosure (13/19, 68%), and author or developer disclosure (12/19, 63%) were present in most of the digital tools. No tool provided information to ascertain the independence of sponsors or funders (0/19, 0%); 5% (1/19) provided a sponsorship disclosure and 11% (2/19) outlined author or developer credentials, which included academics and O&G. Overall, 21% (4/19) of digital tools contained references (Multimedia Appendix 7). App06 met the most quality criteria (14 of 21), followed by App05 (9 of 21), and Web01 (9 of 21).

Discussion

Principal Findings

Women are increasingly engaging with digital resources for health guidance, including healthy lifestyles and weight gain during pregnancy. A systematic search approach identified current and publicly available websites and mobile apps that contain tools and resources to monitor GWG. Those included were reviewed based on their quality, features and functions; behavior change potential; the credibility, quality, and safety of the health-related information provided; and their ability to highlight the importance of optimizing GWG. Across 19 eligible digital tools, we found that the majority reported features including pregnancy-related education, advice, monitoring, and tracking of GWG. Despite this, the quality of information related to GWG was poor, and limited ability to guide behavior change

XSL•FO
RenderX

for optimized GWG was found. Advice related to achieving healthy GWG was present in ≤50% of the apps. Overall, this advice was nonspecific in nature and unlikely to be associated with evidence-based information. We found minimal likelihood of resources to alert, provide support, or direct women into partnerships with their health care provider if GWG was outside the recommended thresholds. These results emphasize a missed opportunity in information provision and support to safely optimize health behaviors and GWG for women. There is a critical need to improve the quality and regulation of publicly accessible web-based resources informed by health care, policy, and consumer needs during pregnancy.

Pregnancy presents a unique opportunity in which women are motivated to optimize lifestyle behaviors to ensure favorable health outcomes for themselves and their baby [30]. Optimizing diet, physical activity, and ultimately GWG during pregnancy reduces adverse outcomes for mother and baby and confers protective maternal and intergenerational benefits [30-32]. Our results support a mobilization of women during this time in engagement with health resources, with over 25 million downloads observed across the 18 apps included in this review. Associated consumer user ratings for apps were very high; however, it is not clear what aspects were most appealing and why. Recent qualitative research exploring consumer preferences and experiences with mHealth apps for maternal health reported that functionality and technical ability features were perceived to be of highest value to women [33]. Consumers reported an increased likelihood to use apps that were free or low-cost, aesthetically pleasing, and with minimal technological barriers [33]. However, little emphasis was placed on the quality or credibility of information by consumers when prompted, and there was little desire to obtain and ensure evidence-based information was received [33]. This may potentially explain the high user ratings of the apps included here. On evaluation, MARS domains related to visual appeal, engagement, and functionality scored the highest overall compared with domains related to content specificity, in line with previous research evaluating pregnancy-related apps [18]. Interestingly, although not captured on the scales applied in this study, we observed that functionality was impeded in several apps by mandatory viewing of advertisements contingent on accessing free features, information, or moving between pages. However, it is unclear whether this impacted the highly scored user ratings overall.

In the absence of availability of a framework to evaluate safety features within web-based resources, we built on our previous research [19,20] and included a checklist to rigorously evaluate the presence of features related to GWG management. These included consultation with relevant health care providers in content development, linkage to clinical practice guidelines for pregnancy care and guidelines for GWG, evaluation of surrounding content to promote healthy GWG, and in-built alerts if GWG entries are outside of the recommended range. Overall, we found that only 10% disclosed development in consultation with O&G expertise, 10% used adequate referencing for GWG guidelines, 10% included an alert for GWG outside of recommendations based on preconception weight and height, and none advised health care consultation if GWG was outside of recommendations. These results emphasize a near-complete absence of components related to safety within currently available web-based resources, mandating a critical need to improve regulatory control in this field [34,35]. Previous research in over 1 million pregnancies worldwide demonstrated an increase in adverse outcomes for both mother and baby when GWG is below or above international recommendations, compared with GWG within recommended thresholds [7]. Level 1 evidence demonstrates optimized GWG and improved maternal outcomes following antenatal lifestyle intervention, and there is now a strong mandate for the implementation of effective strategies in routine care [36]. With increased engagement in and availability of resources to monitor GWG, it is essential that evidence-based information and recommendations are made available to support women, with effective translation of health information congruent with the current guidelines to minimize potential harm.

Using the validated ABACUS framework, we evaluated the capacity of the included apps to guide and support behavior change [23] toward the optimization of GWG. Our results demonstrate that beyond the ability to personalize adaptable features within apps with user information or personal data, scores for the capacity to change behavior were poor overall. Behavioral techniques common to healthy lifestyle change [37], including goal setting, problem solving, provision of consequences related to the target behavior, habit formation, and social and environmental antecedents of behavior, were rarely present. This is reflective of findings within previous non–pregnancy-related research [38] and pregnancy-related research specific to exercise and physical activity [21]. Further research is needed to fully elucidate which behavioral components embedded within web-based resources are effective in changing behavior [38]. This is particularly significant in the context of the burgeoning availability and use of health apps, yet for developers minimal evaluation of efficacy in changing health behaviors or improving health outcomes is required [38].

Altogether, our results highlight several areas of concern, culminating in a missed opportunity to support and guide women during this formative life phase of increased health care needs. First, despite increasing awareness, there is little regulatory control currently in place for digital health resources that are publicly available, which is an area warranting improvement. A recent Australian review highlighted the complexities between developer and consumer considerations and the involvement of multiple, siloed sectors, traversing medical, privacy, advertising, finance, and digital content as barriers to improving regulations to ensure consumer safety [39]. Of the policy documentation available, the review found a focus on the commercial loss or gains related to regulation over and above consumer safety, with consumers ultimately assigned as the primary evaluator in selecting safe and credible apps [39]. Given that women may base their engagement on functionality and aesthetics aspects within apps [33], there is a need to develop resources that can inform women about the quality, credibility, and safety of apps in a reliable, easy, and transparent way. This could include independent certification or endorsements not dissimilar to currently available entities, such as Health on the Net or similar [21,40]. Second, given that resources were likely to be more

XSL•FO
RenderX

based on function and aesthetics, it is not unreasonable to conclude that entertainment and gamification came at the expense of credible information and support for women. Frequent exposure to advertisements highlights the potential for exploitation of women when using resources with exposure to potentially harmful information and imagery, underscoring the need for improved regulation and distinction between apps for entertainment and those for health information provision. Finally, in improving content quality within apps, improved partnership among commercial developers, policy makers, the health care sector, and with women, the consumers, at the forefront is required. Co-design of resources must occur to ensure a balance between the valued consumer attributes of apps alongside evidence-based information and effective behavior change techniques delivered in a way that women value as engaging, trustworthy, and safe. Previous research suggests that involving relevant expertise in app development does not compromise user downloads of apps, suggesting that quality can be optimized without compromising popularity [41].

Strengths and Limitations

This study had several strengths and limitations. To ensure that we captured the available digital health resources for GWG, we used a robust search strategy across both websites and mHealth apps with minimal exclusion criteria, reflective of our search results. By reviewing current digital tools using the validated MARS and ABACUS tools, questions specific to GWG as well as evaluation of credibility of health-related information, we were able to evaluate technical features and quality as well as the behavior change potential and health information. We applied safety criteria specific to GWG management based on our previous publications [19,20] and tested all weight trackers for their ability to digitally summarize GWG, provide personalized feedback according to GWG, and alert and direct women to health care provision if GWG was outside recommendations. Owing to inconsistent search terms used for pregnancy and weight management across Google Play Store and Apple App Store, it is possible that some apps may have been inadvertently missed. Furthermore, a search for digital resources cannot be replicated due to the rapidly changing market and time-dependent popularity, which warrants the need for the development of validated search frameworks in this field.

Conclusions

This review emphasizes the substantial limitations in publicly available consumer-facing digital resources for monitoring and optimizing GWG. Most tools reviewed were of low quality overall, had minimal ability to support behavior change, and were potentially unsafe, with minimal linkage to evidence-based information or partnership with health care. When women require increased support for health optimization, these results emphasize the minimal likelihood of currently available resources to positively influence GWG or, ultimately, health outcomes during this time. Owing to the extensive use of publicly available digital tools, these findings underscore the critical need for better linkage among health, research, and commercial sectors to design apps that are high quality across visual appeal, functionality, credibility, safety, and effectiveness in lifestyle modification and self-management of GWG.

Acknowledgments

The authors have received no specific funding for this study. Other funding support is as follows. BRB is supported by a Monash Graduate Scholarship and received funding support from Medibank. CLH is supported by a National Health and Medical Research Council Centres of Research Excellence Health in Preconception and Pregnancy Senior Postdoctoral Fellowship (APP1171142). SJdJ is supported by a Metro North Health Clinician Research Fellowship. HJT is supported by a Medical Research Future Fund and National Health and Medical Research Council Fellowship.

Authors' Contributions

BRB, CLH, JAB, and RMG conceptualized and refined the research idea. CLH and HJT were responsible for funding to support the work. BRB, CLH, MJH, JAB, and RMG designed the study. BRB, CLH, and RMG conducted the literature search and screening of tools. BRB, MJH, SJdJ, AC, QVH, RMG, and CLH conducted data extraction and preparation; BRB synthesized data and conducted statistical analyses. All authors assisted in the interpretation of the analyses, had intellectual input into manuscript and reviewed and approved the manuscript. BRB prepared the manuscript. CLH and RMG supervised this work and CLH has overall responsibility for the work and is the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Gestational weight gain criteria.
[DOCX File , 18 KB - jmir_v24i11e37552_app1.docx]

Multimedia Appendix 2
Mobile App Rating Scale.
[DOCX File , 50 KB - jmir_v24i11e37552_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
App Behavior Change Scale.
[DOCX File , 23 KB - jmir_v24i11e37552_app3.docx]

Multimedia Appendix 4
Quality evaluation criteria.
[DOCX File , 22 KB - jmir_v24i11e37552_app4.docx]

Multimedia Appendix 5
Description of digital tools for gestational weight gain management (results table).
[DOCX File , 15 KB - jmir_v24i11e37552_app5.docx]

Multimedia Appendix 6
Performance on gestational weight gain (GWG) quality questions (inclusion of GWG-specific tools or features; results figure).
[PNG File , 67 KB - jmir_v24i11e37552_app6.png]

Multimedia Appendix 7
Performance on quality evaluation (results table).
[DOCX File , 14 KB - jmir_v24i11e37552_app7.docx]

References

1. Gilmore LA, Klempel-Donchenko M, Redman LM. Pregnancy as a window to future health: excessive gestational weight gain and obesity. Semin Perinatol 2015 Jun;39(4):296-303 [FREE Full text] [doi: 10.1053/j.semperi.2015.05.009] [Medline: 26096078]
2. Curhan GC, Chertow GM, Willett WC, Spiegelman D, Colditz GA, Manson JE, et al. Birth weight and adult hypertension and obesity in women. Circulation 1996 Sep 15;94(6):1310-1315. [doi: 10.1161/01.cir.94.6.1310] [Medline: 8822985]
3. Curhan GC, Willett WC, Rimm EB, Spiegelman D, Ascherio AL, Stampfer MJ. Birth weight and adult hypertension, diabetes mellitus, and obesity in US men. Circulation 1996 Dec 15;94(12):3246-3250. [doi: 10.1161/01.cir.94.12.3246] [Medline: 8989136]
4. Whincup PH, Kaye SJ, Owen CG, Huxley R, Cook DG, Anazawa S, et al. Birth weight and risk of type 2 diabetes: a systematic review. JAMA 2008 Dec 24;300(24):2886-2897. [doi: 10.1001/jama.2008.886] [Medline: 19109117]
5. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of in utero and early-life conditions on adult health and disease. N Engl J Med 2008 Jul 03;359(1):61-73 [FREE Full text] [doi: 10.1056/NEJMra0708473] [Medline: 18596274]
6. Goldstein RF, Abell SK, Ranasinha S, Misso M, Boyle JA, Black MH, et al. Association of gestational weight gain with maternal and infant outcomes: a systematic review and meta-analysis. JAMA 2017 Jun 06;317(21):2207-2225 [FREE Full text] [doi: 10.1001/jama.2017.3635] [Medline: 28586887]
7. Goldstein RF, Abell SK, Ranasinha S, Misso ML, Boyle JA, Harrison CL, et al. Gestational weight gain across continents and ethnicity: systematic review and meta-analysis of maternal and infant outcomes in more than one million women. BMC Med 2018 Aug 31;16(1):153 [FREE Full text] [doi: 10.1186/s12916-018-1128-1] [Medline: 30165842]
8. Beyerlein A, Schiessl B, Lack N, von Kries R. Associations of gestational weight loss with birth-related outcome: a retrospective cohort study. BJOG 2011 Jan;118(1):55-61. [doi: 10.1111/j.1471-0528.2010.02761.x] [Medline: 21054761]
9. Kennedy RA, Mullaney L, Reynolds CM, Cawley S, McCartney DM, Turner MJ. Preferences of women for web-based nutritional information in pregnancy. Public Health 2017 Feb;143:71-77. [doi: 10.1016/j.puhe.2016.10.028] [Medline: 28159029]
10. Kraschnewski JL, Chuang CH, Poole ES, Peyton T, Blubaugh I, Pauli J, et al. Paging "Dr. Google": does technology fill the gap created by the prenatal care visit structure? Qualitative focus group study with pregnant women. J Med Internet Res 2014 Jun 03;16(6):e147 [FREE Full text] [doi: 10.2196/jmir.3385] [Medline: 24892583]
11. Hearn L, Miller M, Fletcher A. Online healthy lifestyle support in the perinatal period: what do women want and do they use it? Aust J Prim Health 2013;19(4):313-318. [doi: 10.1071/PY13039] [Medline: 23899373]
12. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
13. Mead N, Varnam R, Rogers A, Roland M. What predicts patients' interest in the internet as a health resource in primary care in England? J Health Serv Res Policy 2003 Jan;8(1):33-39. [doi: 10.1177/135581960300800108] [Medline: 12683432]
14. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265 [FREE Full text] [doi: 10.1002/oby.20946] [Medline: 25521523]

XSL•FO
RenderX

15. Chuang CH, Stengel MR, Hwang SW, Velott D, Kjerulff KH, Kraschnewski JL. Behaviours of overweight and obese women during pregnancy who achieve and exceed recommended gestational weight gain. Obes Res Clin Pract 2014;8(6):e577-e583 [FREE Full text] [doi: 10.1016/j.orcp.2013.12.254] [Medline: 25434913]
16. Phelan S, Jankovitz K, Hagobian T, Abrams B. Reducing excessive gestational weight gain: lessons from the weight control literature and avenues for future research. Womens Health (Lond) 2011 Nov;7(6):641-661 [FREE Full text] [doi: 10.2217/whe.11.70] [Medline: 22040207]
17. Dahl AA, Dunn CG, Boutté AK, Crimarco A, Turner-McGrievy G. Mobilizing mHealth for moms: a review of mobile apps for tracking gestational weight gain. J technol behav sci 2017 Nov 2;3(1):32-40. [doi: 10.1007/s41347-017-0030-6]
18. Musgrave LM, Kizirian NV, Homer CS, Gordon A. Mobile phone apps in Australia for improving pregnancy outcomes: systematic search on app stores. JMIR Mhealth Uhealth 2020 Nov 16;8(11):e22340 [FREE Full text] [doi: 10.2196/22340] [Medline: 33196454]
19. Chiu W, Kuczynska-Burggraf M, Gibson-Helm M, Teede H, Vincent A, Boyle J. What can you find about polycystic ovary syndrome (PCOS) online? Assessing online information on PCOS: quality, content, and user-friendliness. Semin Reprod Med 2018 Jan;36(1):50-58. [doi: 10.1055/s-0038-1667186] [Medline: 30189451]
20. Aleksova J, Kuczynska-Burggraf M, Ranasinha S, Vincent A. Information on early menopause: is the internet the place to search? Climacteric 2017 Jun;20(3):248-255. [doi: 10.1080/13697137.2017.1301920] [Medline: 28333562]
21. Hayman M, Alfrey K, Cannon S, Alley S, Rebar AL, Williams S, et al. Quality, features, and presence of behavior change techniques in mobile apps designed to improve physical activity in pregnant women: systematic search and content analysis. JMIR Mhealth Uhealth 2021 Apr 07;9(4):e23649 [FREE Full text] [doi: 10.2196/23649] [Medline: 33825693]
22. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
23. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
24. Chivers BR, Garad RM, Boyle JA, Skouteris H, Teede HJ, Harrison CL. Perinatal distress during COVID-19: thematic analysis of an online parenting forum. J Med Internet Res 2020 Sep 07;22(9):e22002 [FREE Full text] [doi: 10.2196/22002] [Medline: 32857707]
25. Chivers BR, Garad RM, Moran LJ, Lim S, Harrison CL. Support seeking in the postpartum period: content analysis of posts in web-based parenting discussion groups. J Med Internet Res 2021 Jul 15;23(7):e26600 [FREE Full text] [doi: 10.2196/26600] [Medline: 34264198]
26. Institute of Medicine (US) and National Research Council (US) Committee to Reexamine IOM Pregnancy Weight Guidelines. Weight Gain During Pregnancy: Reexamining the Guidelines. Washington (DC): National Academies Press; 2009.
27. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
28. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the mobile application rating scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
29. Kachingwe AF, Phillips BJ. Inter- and intrarater reliability of a back range of motion instrument. Arch Phys Med Rehabil 2005 Dec;86(12):2347-2353. [doi: 10.1016/j.apmr.2005.07.304] [Medline: 16344034]
30. Phelan S. Pregnancy: a "teachable moment" for weight control and obesity prevention. Am J Obstet Gynecol 2010 Feb;202(2):135.e1-135.e8 [FREE Full text] [doi: 10.1016/j.ajog.2009.06.008] [Medline: 19683692]
31. Raghavan R, Dreibelbis C, Kingshipp BL, Wong YP, Abrams B, Gernand AD, et al. Dietary patterns before and during pregnancy and maternal outcomes: a systematic review. Am J Clin Nutr 2019 Mar 01;109(Suppl_7):705S-728S. [doi: 10.1093/ajcn/nqy216] [Medline: 30982868]
32. Thangaratinam S, Rogozinska E, Jolly K, Glinkowski S, Roseboom T, Tomlinson JW, et al. Effects of interventions in pregnancy on maternal weight and obstetric outcomes: meta-analysis of randomised evidence. BMJ 2012 May 16;344:e2088 [FREE Full text] [doi: 10.1136/bmj.e2088] [Medline: 22596383]
33. Biviji R, Williams KS, Vest JR, Dixon BE, Cullen T, Harle CA. Consumer perspectives on maternal and infant health apps: qualitative content analysis. J Med Internet Res 2021 Sep 01;23(9):e27403 [FREE Full text] [doi: 10.2196/27403] [Medline: 34468323]
34. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
35. Magrabi F, Habli I, Sujan M, Wong D, Thimbleby H, Baker M, et al. Why is it so difficult to govern mobile apps in healthcare? BMJ Health Care Inform 2019 Nov;26(1):e100006 [FREE Full text] [doi: 10.1136/bmjhci-2019-100006] [Medline: 31744843]

XSL•FO
RenderX

36. Cantor AG, Jungbauer RM, McDonagh M, Blazina I, Marshall NE, Weeks C, et al. Counseling and behavioral interventions for healthy weight and weight gain in pregnancy: evidence report and systematic review for the US preventive services task force. JAMA 2021 May 25;325(20):2094-2109. [doi: 10.1001/jama.2021.4230] [Medline: 34032824]
37. 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]
38. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
39. Parker L, Bero L, Gillies D, Raven M, Grundy Q. The "Hot potato" of mental health app regulation: a critical case study of the australian policy arena. Int J Health Policy Manag 2019 Mar 01;8(3):168-176 [FREE Full text] [doi: 10.15171/ijhpm.2018.117] [Medline: 30980633]
40. HONcode Certification. Health on the Net. URL: https://www.hon.ch/en/certification.html [accessed 2021-10-18]
41. Pereira-Azevedo N, Osório L, Cavadas V, Fraga A, Carrasquinho E, Cardoso de Oliveira E, et al. Expert involvement predicts mHealth app downloads: multivariate regression analysis of urology apps. JMIR Mhealth Uhealth 2016 Jul 15;4(3):e86 [FREE Full text] [doi: 10.2196/mhealth.5738] [Medline: 27421338]

Abbreviations

ABACUS: App Behavior Change Scale
GWG: gestational weight gain
ICC: intraclass correlation
MARS: Mobile App Rating Scale
O&G: obstetrics and gynecology

Edited by G Eysenbach; submitted 24.02.22; peer-reviewed by A Paglialonga, M Waring, C Pathiravasan ; comments to author 13.07.22; revised version received 31.07.22; accepted 20.10.22; published 25.11.22.

Please cite as:
Brammall BR, Garad RM, Boyle JA, Hayman MJ, de Jersey SJ, Teede HJ, Hong QV, Carrandi A, Harrison CL
Assessing the Content and Quality of Digital Tools for Managing Gestational Weight Gain: Systematic Search and Evaluation
J Med Internet Res 2022;24(11):e37552
URL: https://www.jmir.org/2022/11/e37552
doi:10.2196/37552
PMID:36427237

©Bonnie R Brammall, Rhonda M Garad, Jacqueline A Boyle, Melanie J Hayman, Susan J de Jersey, Helena J Teede, Quoc V Hong, Alayna Carrandi, Cheryce L Harrison. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Online Ethnography for People With Chronic Conditions: Scoping Review

Yajing Gao[1*], MD; Xuemei Chen[1*], PhD; Wei Zhang[1], PhD; Qiuyi Wang[1], MD; Jing Liu[1], MD; Lanshou Zhou[1], PhD

Nursing School of Naval Medical University, Shanghai, China

*these authors contributed equally

Corresponding Author:
Lanshou Zhou, PhD
Nursing School of Naval Medical University
800 Xiangyin Road, Yangpu District, Shanghai
Shanghai, 200433
China
Phone: 86 137 6180 1446
Email: zhoulanshu@hotmail.com

Abstract

Background: Online ethnography has been making a unique contribution to people with chronic conditions as a complement to offline ethnography. It can also be used to study the complexities and contingencies of people with chronic conditions in the context of the internet. Therefore, there is a need to synthesize existing knowledge on research activities concerning online ethnography for people with chronic conditions.

Objective: This scoping review aimed to profile the existing evidence on the application of online ethnography for people with chronic conditions, focusing on the characteristics, contributions, and implementation process. This will provide recommendations for the future use of online ethnography.

Methods: We followed the scoping review methodologies developed by Arksey and O' Malley and the Joanna Briggs Institute. A comprehensive search was conducted on the PubMed, CINAHL, Embase, Scopus, and PsycInfo databases using preselected keywords. The search was limited to documents written in English and published between January 1, 2000, and February 1, 2022. After removal of duplicates, articles were screened by 2 independent reviewers reading the title, abstract, and full text. One reviewer extracted data, which were descriptively analyzed to map the existing knowledge.

Results: After 2836 titles and abstracts and 51 full texts were screened, 27 publications were included in the analysis, published between 2009 and 2022. Most studies were from the United States (11/27, 40.7%), and most articles collected data from online forums (10/27, 37.0%). Moreover, the most commonly used type of researcher involvement was passive analysis (24/27, 88.9%), and 18.5% (5/27) of the topics concerned people with mental illness. Notably, the majority of articles did not report the immersion process in detail (17/25, 63.0%). Ethical issues were mentioned in 88.9% (24/27) of the included articles.

Conclusions: We analyzed the current literature across fields and found that online ethnography can be exploited to explore the deeper experience of people with chronic conditions that are difficult to investigate using traditional ethnography. We found that there was diversity in researcher involvement, immersion process, data collection, and data analysis. However, most studies reported the insufficient immersion into the online environment. Researchers should determine the research approaches and data resources in order to complete culture immersion before researching. We also found that there was no uniform standard for ethical issues. Therefore, we recommend that researchers collect public and private data, obtain informed consent, and preserve the privacy and confidentiality of online users with chronic conditions. The findings can provide a practical reference for the use of online health care in studying chronic conditions.

(J Med Internet Res 2022;24(11):e37941) doi:10.2196/37941

KEYWORDS

online ethnography; chronic condition; scoping review; review; ethnography; online; research; online users

XSL•FO
RenderX

Introduction

There is an increasing number of patients with chronic conditions worldwide. The use of digital technology for these patients has become common due to advances in technology. People with chronic conditions include patients experiencing chronic physical pain, mental conditions, continuing conditions, chronic symptoms, or alcohol and substance abuse that normally last more than a year [1,2]. Among the chronic conditions, noncommunicable diseases alone accounted for 60.8% of all deaths in 2000, which rose to 73.6% in 2019 globally [3]. Patients with chronic conditions often face varied long-term physical and mental challenges which require long-term professional and nonprofessional care including health information, psychological care, symptom management, and social support [4,5]. Fortunately, the development and extensive use of digital technology has provided a potential avenue to meet the diverse needs of patients with chronic diseases. Indeed, the World Health Organization (WHO) Global Strategy on Digital Health was adopted in 2020 by the World Health Assembly [6]. For these patients, the most convenient way to take advantage of the available digital technology is to access online health communities, social media, government or charity websites, and live chats, which enables them to share, communicate, and search for tailored health information and implementation of health management [7-9]. One of the most important features of the digital platforms is that users in discussion boards can express themselves anonymously, thus providing them with sufficient privacy protection [7]. Moreover, patients are able to talk freely about particular topics which they would otherwise not mention [10]. In addition, the online platforms can store data about the patients' perceptions and experiences, the interactions between patients, and the interactions between doctors and patients. Moreover, memes and symbols also provide information through nonverbal interactions [11]. Thus, these features have made the online platforms important approaches in collecting qualitative data of patients with chronic conditions.

With the growth of online health communities and platforms, online ethnography has attracted more attention from health scholars, especial those who are interested in chronic diseases. Ethnography research is a qualitative methodology, which allows health professionals and researchers to study and generate a comprehensive understanding of social interaction, behavior, and perceptions that occur within groups, teams, and communities in the health care setting [12]. With a background in anthropology and sociology, ethnography offers a way to understand a particular cultural context. A researcher conducting this kind of study often spends time watching, listening, and engaging with the study participants before analyzing, interpreting, and theorizing upon the observed phenomena. As ethnographic methodologies have flourished, different frameworks supporting different research purposes, settings, and issues have emerged. These frameworks include institutional ethnography, which is focused ethnography and video-ethnography [13-15]. Online ethnography, also referred to as netnography or internet ethnography, has emerged as a new form, along with the metaverse concept, to yield great advantages in the medical field. Nowadays, a massive pool of medical information can rapidly and anonymously be found online, from which someone's commentary or experience with a specific health or medical issue can provide new research insight. As a fairly new approach, online ethnography uses ethnography in online spaces and communities to explore holistic and cultural thoughts as well as the feelings and experiences of particular groups [16]. In the past two decades, there has been an evolution from "participant-observational research" to "netnography" [17]. Earlier online ethnography held a distinction between online and offline realities and tended to search individuals' interactions with online spaces by observing them in offline settings. Whereas, nowadays there is a growing preference to use real-time, participatory netnography [18,19]. Netnography improves accessibility to geographically hard-to-reach populations who prefer anonymity, is more convenient, allows for the ultimate naturalistic inquiry, and has no recall bias [16]. Online ethnography and related methods have been used to research health-related topics, which include exploration of cyber-nursing approaches, discussions on image and performance-enhancing drugs, and examinations of UK general practitioners' views on health policy changes [20-22]. Therefore, a combination of internet-delivered ethnographic modalities and patient-centered approaches is an exciting area that could be pursued in the management of chronic illness, owing to the widespread adoption of online platforms among the patients, including those with mental illness, epilepsy, or diabetes [23-26].

Although there is a growing number of online ethnography studies for people with chronic conditions, there is a lack of reviews synthesizing their common characteristics. In addition, the field of online ethnography is relatively young and thus lacks structures and guidelines, and so many studies are in the exploratory stage. Furthermore, there is limited data on the research specification and standards within a given type of online context and participant group. Consequently, there might be bias in the development of online ethnography. A review of the network ethnography in the field of chronic conditions can help understand research status, existing problems, and issues requiring attention concerning the new research method. Among the different types of reviews, scoping reviews yield a comprehensive review of a new research area [27]. Therefore, we aimed to conduct a scoping review to explore the research status of online ethnography among people with chronic conditions to yield an in-depth understanding of this new method and provide the basis for further research.

Methods

Overview

We employed the scoping review framework by Arksey and O' Malley [28], which encompasses five stages: (1) identification of initial research questions; (2) identification of relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results.

Stage 1: Identifying the Research Question

The two main review questions were as follows: (1) What are the characteristics and contribution of online ethnography in

XSL•FO
RenderX

people with chronic conditions? (2) How should ethnography be conducted for people with chronic conditions in an online environment?

Stage 2: Identifying Relevant Studies

Initially, a literature search was manually conducted in PubMed to identify chronic disease fields where online ethnography was mostly used and developed. We searched for the term "chronic disease" to identify relevant papers. However, the retrieved results were not comprehensive because of our consideration of "chronic conditions" instead of "chronic diseases." Therefore, we decided to choose eligible articles from a wide range of articles and then searched in 5 electronic databases: PubMed, CINAHL, Embase, Scopus, and PsycInfo. The search strategy was as follows: (online* OR internet* OR cyber* OR web* OR digital OR online forum OR virtual community OR Facebook OR Twitter OR Instagram OR blog OR Youtube OR social media OR remote video OR visual) AND (nethnography OR ethnography OR netnography). The search included 2 sets of search terms: "online" or "ethnography." To capture the evolution of online ethnography in people with chronic conditions over the years, the databases were searched between January 1, 2000, and February 1, 2022. An additional list of 2 relevant articles was manually searched to identify any other potentially relevant articles.

Stage 3: Study Selection

Here, we adopted the Joanna Briggs Institute's population-concept-context framework to define our inclusion criteria. We included people with chronic conditions and searched for chronic diseases as a general concept as well as specific diseases including stroke, asthma, chronic obstructive pulmonary disease, cancer, and mental disease. We used any other relevant publications regardless of age, origin, or gender of the studied populations. The concept was use of ethnography while the context was online. We included full texts that reported the following contents: online ethnography being taken as a research purpose, collection of data with online ethnography, and analysis of data with online ethnography. This research included articles that were written in English; published between January 1, 2000, and February 1, 2022; and that evaluated online ethnography in people with chronic conditions. We excluded articles that were not directly related to our research review topic, such as those about dying patients and nursing robots. Gray literature and studies about nonhuman subjects were also excluded. In addition, we excluded articles which did not involve people with chronic conditions and those with no relevant information on online ethnography. A 2-step screening protocol was employed after duplicate removal. First, titles and abstracts were screened to determine the eligibility of each article. Full texts were then screened, and only articles that met the eligibility criteria were included. Two reviewers evaluated the articles using title and abstract analysis for relevance to online ethnography in people with chronic conditions. Full texts were retrieved and independently reviewed by 2 authors (WZ and QW) to confirm study eligibility. Consensus was reached through discussion and, where required, with consultation with a third author (LZ). One author (XC) conducted a supplementary hand search of reviews retrieved from the database search to identify additional studies to be included after consensus with the second author (YG).

Stage 4: Charting the Data

A data extraction template was developed and independently piloted by 2 authors (YG and JL), which was then refined for the purposes of this review. Data were extracted from the included studies by 1 author (YG) and verified by the second author (XC). Information relating to authors, year, study design, country, target group, type of researcher involvement, data source, methods of immersion, data collection, data analysis, study purpose, results, ethical considerations, and limitations were extracted and tabulated using Microsoft Excel 2019 software. Moreover, the type of research was divided according to Keim-Malpass et al [29], who proposed 3 types of researcher involvement in internet-based research: passive analysis, active analysis, and self-identified active analysis. Our study did not appraise methodological quality of risk of bias of the included articles since it is a scoping review.

Stage 5: Collating, Summarizing, and Reporting the Results

We performed a descriptive analysis of the included studies. The analysis included data on publication date, country of origin, type of population, and type of disease studied.

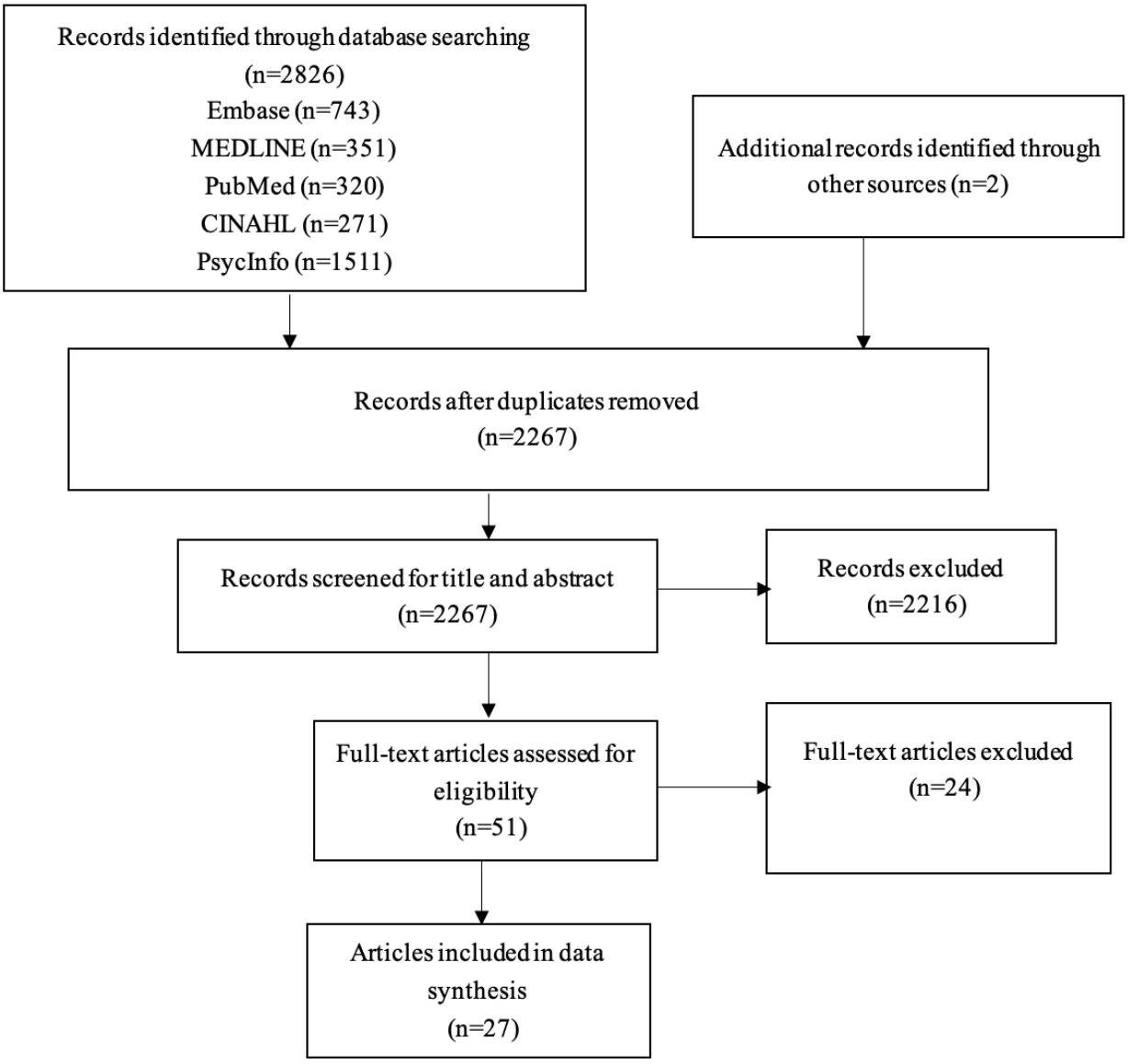
Results

Search Results

The electronic database search yielded a total of 2836 records. The selection process was summarized in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram as shown in Figure 1. Briefly, after removal of duplicates, the titles and abstracts of 2267 records were independently screened by 2 reviewers. The screening excluded 2216 irrelevant records. The full texts of the remaining 51 records were assessed for eligibility. After scanning the reference lists of 2 reviews, we evaluated the eligibility of 27 records whose full texts were screened. Out of these, 24 records were excluded for the following reasons: 9 did not involve people with chronic conditions, 10 were conference abstracts, 4 were not relevant human subjects, and 1 was a duplicate record. Finally, a total of 27 studies were included in the analyses (Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Flow diagram showing included studies.



Characteristics of the Included Studies

Publication Year and Geographic Distribution

Although our research evaluated studies from 2000 to 2022, none of the 27 included articles were conducted before 2009. As shown in Table 1, the number of publications grew through the years. This demonstrates that the application of online ethnography for people with chronic conditions has increased in the past 13 years. Most studies concentrated more on exploring the experience among the targeted population. Table 2 displays the distribution of articles by the continent of publication: 40.7% (11/27) of the articles [24,26,30-38] were from the United States, 29.6% (8/27) from England [25,39-44], 14.8% (4/27) from France [45-47], 7.4% (2/27) from Brazil [48,49], 3.7% (1/27) from Austria [50], and 3.7% (1/27) from Canada [51].

XSL•FO
RenderX

Table 1. The distribution of articles by the year of publication.

Year	Publications, n
2009	1
2011	1
2012	1
2013	1
2014	1
2016	3
2017	2
2018	1
2019	5
2020	5
2021	5
2022	1

Table 2. Distribution of publications by geographic location.

Geographic location	Publications, n
American	11
England	8
France	4
Brazil	2
Austria	1
Canada	1

Data Sources

Currently, most of the used data were from second-generation online social media, which include discussion forums, blogs, and social networking sites [52]. In our study, 4 online ethnographers collected data from more than 1 platform [25,26,38,41] (4/27, 14.8%), and most of them collected data from just 1 source. Among the retrieved articles, most data (10/27, 37.0%) were collected from online forums [33,34,40,43-47,50,53], 18.5% (5/27) were from Facebook [30], 11.1% (3/27) were from websites [30,39,54], 14.08% ((4/27) were from multisources, 7.4% (2/27) were from blogs [31,32], 3.7% (1/27) was from Instagram [37], 3.7% (1/27) was from Twitter [36], and 3.7% (1/27) was from Youtube [24].

Type of Researcher Involvement

Our analysis showed that the majority (24/27, 88.9%) of the included studies were passive analyses [24-26,30-34,36-38,40-44,46,48-51,53,54], 3.7% (1/27) were self-identified active analyses [45], and 7.4% (2/27) involved active analyses [35,39]. With regard to active analysis, Gibson et al [39] emphasized using participation-observation to complete data collection, and Frohlich [35] founded an online community and communicated with patients with inflammatory bowel disease. Troisoeufs et al [45] carried out 8 semidirective interviews from the consumers of online ethnography, and thus the study was identified as a self-identified analysis.

Domains of Chronic Conditions

On the basis of the report of the National Health Council [55], we classified the included chronic conditions into mental illness [24,41,44,51], female reproductive diseases [25,34,38,42], neurological disorders [40,45,46,50], cancer [31,32,39,53], rheumatic immune system diseases [33,37,43], common chronic diseases [26,36,48], autoimmune disorders [30,35], hereditary diseases [54], HIV [49], and thyroid disease [47]. Table 3 displays the distribution of articles by the domains of chronic conditions. Most of the studies focused on patients with mental illness (4/27, 14.8%), neurological disorders (4/27, 14.8%), cancer (4/27, 14.8%), or female reproductive diseases (4/27, 14.8%). Additionally, 11.1% (3/27) focused on diabetes, 11.1% (3/27) on rheumatic immune system diseases, 7.4% (2/27) on autoimmune disorders, 3.7% (1/27) on hereditary disease, 3.7% (1/27) on HIV, and 3.7% (1/27) on thyroid disease .

XSL•FO
RenderX

Table 3. The distribution of articles by type of disease.

Disease	Publications, n
Mental illness	4
Female reproductive disease	4
Neurological disorders	4
Cancer	4
Rheumatic immune system disease	3
Diabetes	3
Autoimmune disorders	2
Hereditary disease	1
HIV	1
Thyroid disease	1

Study Purpose and Important Results

We evaluated the effect of online ethnography on health information flow, emotional support, and interaction among patients with chronic conditions, peers, and health providers. We found the studies could be used to explore new modes of online medical practices, which include application of new medical technologies. With regard to study results, themes such as experience of self-management, experience of living with illness, and physician-patient relationship emerged (Multimedia Appendix 1).

Methods of Immersion and Data analysis

Our findings showed that the majority (17/27, 63.0%) of articles among the included studies did not report the immersion process in detail [25,30,33,34,37,38,40-42,44-46,49-51,53,54], while 37.0% (10/27) did report it [24,26,31,32,35,36,39,43,47,48]. With regard to data analysis, our results demonstrated that the studies employed various data analysis methods, including thematic analysis [31,33,38,40-44,47,48,53], grounded theory [24,35,38,50,54], content analysis [37,39,46,53], and disclose analysis [34]. The most used method was thematic analysis (11/27, 37.0%). Grounded theory (5/27, 18.5%) and content analysis (4/27, 14.8%) were also frequently used in the data analysis.

Ethical Issues

Ethical issues were mentioned in the majority of articles. Only 11.1% (3/27) of the articles did not mention any issues on ethics [26,45,50]. In general, there were 5 ethical issues in the included studies, which included public data [24,25,30-33,35-37,40,41,44,46,48,51,53], informed consent [38-40,42,43,47,49,51], privacy and confidentiality [31,32,40,41,44,46,47], naming [24,31,39-41,44,46,47,51,54], and legal consideration [24,25,30-34,36,37,39-44,47-49,54]. Moreover, 70.4% (19/27) of ethical concerns were approved or exempted by institutional review board, 55.6% (15/27) concerned public data, 37.0% (10/27) concerned naming, 25.9% (7/27) emphasized participants' privacy and confidentiality, and 29.6% (8/27) acquired users' informed consent.

Discussion

Principal Findings

Our findings demonstrated that online ethnography has great potential in the field of chronic disease research and has yielded beneficial outcomes in many chronic disease studies. This method has also attracted more attention from scholars in the field of chronic diseases**.** Most of the included articles in this scoping review were conducted from 2009 onward, and the number of included studies gradually increased with time, which may indicate growing interest in this area of research, especially since 2019. The included articles covered various chronic conditions, such as mental illness, neurological disorders, common chronic diseases, cancer, rheumatic immune system diseases, autoimmune disorders, hereditary diseases, female reproductive diseases, and HIV. These articles showed that online ethnography could highlight the patients' experiences with disease and demonstrated new insights into improving medical practice from a patient-centered perspective. These diseases are characterized by long-term physical health conditions and the high mental health needs of the patients. In our scoping review, youth with mental illnesses and women with reproductive diseases were the most concerned population, which is consistent with Bour et al's [56] study. These categories of patients are among the most stigmatized, marginalized, and vulnerable members of society and suffer from discrimination in many areas of their daily life. Online forums and social media can be a user-friendly environment where individuals can normalize their illness, assert their voice and identity with each other by validating shared experience with peers, and form social ties with a broad audience [57]. HIV patients and youth with mental illness affected by COVID-19 were also discussed in our study [44,49]. COVID-19 possibly shifts from an acute to chronic disease and also causes political controversy [58-60], while HIV has long been recognized as a health and political challenge [60,61]. Since online ethnography is currently a well-established method, which involves the social, cultural, and structural dimension of a given phenomenon [44], it is particularly well suited to dealing with personally or politically sensitive topics or illegal acts in online communities [19].

XSL•FO
RenderX

Online ethnography showed great contribution toward understanding how online users acquired health information, emotional support, and interaction with health providers, which provides thoughts on how to improve internet-based medical practices. First, the data showed that online users possibly sought health information and emotional support from peers instead of health care professionals. This could be because people tend to trust others who have similar challenges more than figures of authority from business, government, or mass media. However, there exists a controversy between the untrusted face-to-face patient-provider relationship and online support needs from health care professionals. Second, online ethnography can investigate the online health care pathways among chronically ill patients, which include individual and collective self-management and empowerment processes [53]. In addition, online ethnography can be used to evaluate social and experiential aspects of health technologies, such as deep brain stimulation and the Open Artificial Pancreas System (OpenAPS) do-it-yourself artificial pancreas technology [36,45,62], which could promote offline promotion of new medical technologies. In summary, online ethnography is a bridge that connects online and offline cultures, which affect offline-online social and medical practices among people with chronic conditions. It is also noteworthy that most of the studies were conducted in Western cultures. Therefore, there is need for research from other countries.

We also highlighted some characteristics and issues regarding the methodology and implementation process which should be dealt with. We identified online forums as the main data source, followed by Facebook, but the majority of the studies did not describe the immersion process. These data sources of online ethnography are broad and constantly evolving. Keim-Malpass et al [31] searched for blogs that explored the experience of women with cancer in the past, and now the use of online health forums for data collection for online ethnography is steadily increasing. This is because online forums can provide a rich source of primary naturalistic data about the perspectives and experiences of users with a particular health issue and can offer additional insights compared with traditional interviews [63,64]. Our analysis showed that Facebook has many users, and thus a large amount of data can potentially be harnessed. Although Facebook users' activity consists of creating written posts, the potential use of text data for research is enormous, and thus online ethnographers are increasingly using it for research [49,65]. Instagram is pictorial, Reddit is text-based, and Twitter comprises short posts, all of which can facilitate sustained discussions where people share their in-depth experiences. Additionally, YouTube and video websites are public video-sharing platforms which may serve as important avenues to gain an understanding of people's viewpoints [66-68]. Online ethnographers could choose one or more platforms as data sources depending on the research objectives. Moreover, using more than 1 research method was associated with superior outcomes compared to using a single study method. A combination of 2 methods can compare and explore deeper experiences and value perspectives [69]. For example, Lee et al [34] gathered narratives through interviews to circumvent the limitation of online ethnography where some women would keep silent or express false feelings in the infertility forum. The research methods for online ethnography are constantly evolving, and now it can stand alone or be combined with other research methods including offline interactions [70]. Furthermore, the majority of the studies did not describe the immersion process, a component that should be considered in future studies. Theoretically, the main researchers need to be immersed in the online culture of the targeted group prior to sampling. In our scoping review, only 8 articles described the immersion process; however, developing competencies as an online ethnographic fieldworker (immersive depth, prolonged engagement, researcher identification, and persistent conversations) is a key strategy for learning socially relevant events [71]. Reading relevant posts for several weeks or months, engaging researchers who are familiar with a given online environment, and being involved in the online community are 3 approaches to achieve a high quality of immersion. In addition, various qualitative data analysis methods were chosen based on specific research questions in the articles included, while quantitative methods need further exploration. Researchers believe that thematic analysis, lexical analysis, content analysis, grounded theory, narrative analysis, and "netnography" can be selected according to the research purpose [46,48]. In the future, there is a possibility that quantitative analysis could be combined with netnography [72]. Online ethnographers should take advantage of online software in data analysis under the condition that they follow the principles of ethnography.

We also highlighted 5 aspects of ethical issues and relevant limitations in the application processes of online ethnography for people with chronic conditions. First, the majority of our included studies were conducted using public data to avoid ethical controversy, which meant that individuals who had posted public messages did not have a reasonable expectation of privacy. However, such public data could lead to research bias, as passive patients with chronic conditions could have contrary opinions. Moreover, administrators of online communities sometimes would remove messages that were sad, shocking, or in violation of the group's code of conduct, and, therefore, researchers could not capture a true and meaningful experience. Second, although it seemed justifiable to waive informed consent for observational research in a public space in the past [73], Hetland et al [74] recently suggested that informed consent was one way that ethnographers could handle potential risks in internet research. In our study, we discovered that some investigators had posted announcements or contacted users by email to protect users' rights, which could be one form of balancing the issue of informed consent. Third, to protect the identities of a vulnerable population, all names were removed and post titles or specific demographic characteristics were not reported. We further recommend concealing discriminative descriptions because search engines are often capable of finding statements used in research reports, making it difficult to conceal certain data in certain avenues. Lastly, our finding showed that there was no consensus research among various national ethical committees regarding online ethnography [75]. Based on the literature, it was acknowledged that the ethics surrounding internet research is complex and multifactorial because internet research is a multidisciplinary project and has different types of environments [76,77]. Therefore, we suggest that online ethnographers should follow guidelines of local ethical

XSL•FO
RenderX

committees. In conclusion, we recommend using public and private data as well as preserving patient anonymity, privacy, and confidentiality through removal of identified information when performing online ethnography among people with chronic conditions [78].

Strengths and Limitations

The present study used a rigorous scoping review methodology based on the manual of the Joanna Briggs Institute. To ensure a broad search of the literature, the search strategy included 5 databases and employed the snowball technique. Although we highlighted significant findings, our scoping review process was limited by the fact that we might not have identified all relevant articles in the published literature despite attempts to be as comprehensive as possible. We also limited our review to documents written in English, which might have led to relevant studies being missed.

Conclusions and Recommendations

Our finding suggested that online ethnography has good potential for exploring the deeper experience of people with chronic conditions, which is difficult to investigate with traditional ethnography. These results provide practical guidance for the online health care of chronic diseases in a wide range of fields. Researchers should first determine the research type and map the online community. There was high heterogeneity in the immersion, data collection, and analysis. We therefore recognize that online ethnography is adaptable and without strict methodological limitations. However, the analyzed articles demonstrated insufficiency in the immersion process into the online environment. In addition, we observed that there is no uniform standard for ethical concerns. Therefore, we recommend preserving the privacy and confidentiality of online users.

Acknowledgments

This work is supported by The National Natural Science Foundation of China (grant #71904196) and the Shanghai Pujiang Program (grant #2019PJC111).

Authors' Contributions

YG and XC screened the title, abstract, and full-text of the identified studies. YG performed the data extraction. LZ performed the data verification. YG and XC drafted the manuscript. YG, XC, WZ, QY, and JL reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of included studies.
[DOCX File , 73 KB - jmir_v24i11e37941_app1.docx]

References

1. Cheng C, Inder K, Chan SW. Patients' experiences of coping with multiple chronic conditions: A meta-ethnography of qualitative work. Int J Ment Health Nurs 2019 Feb 19;28(1):54-70. [doi: 10.1111/inm.12544] [Medline: 30232822]
2. Delaney C, Barrere C. Bark L: A Metaphor Analysis of Patients' With Chronic Conditions Experiences With Holistic Nurse Coaching. Holist Nurs Pract 2020;34(1):A. [doi: 10.1097/hnp.0000000000000359]
3. World Health Statistics. Switzerland: World Health Organization; 2021.
4. Sherman LD, Goidel K, Bergeron CD, Smith ML. Web-based health information seeking among African American and Hispanic men living with chronic conditions: cross-sectional survey study. J Med Internet Res 2021 Jul 14;23(7):e26180 [FREE Full text] [doi: 10.2196/26180] [Medline: 34259646]
5. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Education and Counseling 2002 Oct;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0]
6. Dhingra D, Dabas A. Global strategy on digital health. Indian Pediatr 2020 Apr 25;57(4):356-358. [doi: 10.1007/s13312-020-1789-7]
7. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
8. Mendell J, Bates J, Banner-Lukaris D, Horvat D, Kang B, Singer J, et al. What do patients talk about? A qualitative analysis of online chat sessions with health care specialists during a "virtual" cardiac rehabilitation program. Telemed J E Health 2019 Jan;25(1):71-78. [doi: 10.1089/tmj.2017.0206] [Medline: 29742035]
9. Kingod N, Cleal B, Wahlberg A, Husted GR. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res 2017 Jan;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]

XSL•FO
RenderX

10. Bailey E, Robinson J, Alvarez-Jimenez M, Nedeljkovic M, Valentine L, Bendall S, et al. Moderated online social therapy for young people with active suicidal ideation: qualitative study. J Med Internet Res 2021 Apr 5;23(4):e24260. [doi: 10.2196/24260]
11. Rozenblum R, Greaves F, Bates DW. The role of social media around patient experience and engagement. BMJ Qual Saf 2017 Apr 20;26(10):845-848. [doi: 10.1136/bmjqs-2017-006457]
12. Strudwick RM. Ethnographic research in healthcare - patients and service users as participants. Disabil Rehabil 2021 Nov;43(22):3271-3275. [doi: 10.1080/09638288.2020.1741695] [Medline: 32202439]
13. Kearney GP, Corman MK, Hart ND, Johnston JL, Gormley GJ. Why institutional ethnography? Why now? Institutional ethnography in health professions education. Perspect Med Educ 2019 Feb 11;8(1):17-24. [doi: 10.1007/s40037-019-0499-0]
14. Chin S, Lopez V, Tan M, Goh Y. "I Would Like to Be Heard." Communicating With Singaporean Chinese Patients With Dementia: A Focused Ethnography Study. J Transcult Nurs 2018 Sep 18;30(4):331-339. [doi: 10.1177/1043659618800535]
15. Ajjawi R, Hilder J, Noble C, Teodorczuk A, Billett S. Using video‐reflexive ethnography to understand complexity and change practice. Med Educ 2020 May;54(10):908-914. [doi: 10.1111/medu.14156]
16. Keim-Malpass J, Steeves RH, Kennedy C. Internet ethnography: A review of methodological considerations for studying online illness blogs. International Journal of Nursing Studies 2014 Dec;51(12):1686-1692. [doi: 10.1016/j.ijnurstu.2014.06.003]
17. Morais G, Santos V, Gonçalves C. Netnography: Origins, Foundations, Evolution and Axiological and Methodological Developments and Trends. TQR 2020 Feb 18:441-455. [doi: 10.46743/2160-3715/2020.4227]
18. Winter R, Lavis A. Looking, But Not Listening? Theorizing the Practice and Ethics of Online Ethnography. Journal of Empirical Research on Human Research Ethics 2019 Jul 07;15(1-2):55-62. [doi: 10.1177/1556264619857529]
19. Costello L, McDermott M, Wallace R. Netnography. International Journal of Qualitative Methods 2017 Apr 04;16(1):160940691770064. [doi: 10.1177/1609406917700647]
20. Eriksson H, Salzmann-Erikson M. Cyber nursing?Health ?experts? approaches in the post-modern era of virtual performances: A nethnography study. International Journal of Nursing Studies 2013 Mar;50(3):335-344. [doi: 10.1016/j.ijnurstu.2012.09.014]
21. Underwood M. Exploring the social lives of image and performance enhancing drugs: An online ethnography of the Zyzz fandom of recreational bodybuilders. International Journal of Drug Policy 2017 Jan;39:78-85. [doi: 10.1016/j.drugpo.2016.08.012]
22. Hatchard JL, Quariguasi Frota Neto J, Vasilakis C, Evans-Reeves KA. Tweeting about public health policy: Social media response to the UK Government's announcement of a Parliamentary vote on draft standardised packaging regulations. PLoS ONE 2019 Feb 26;14(2):e0211758. [doi: 10.1371/journal.pone.0211758]
23. Alhusseini N, Banta J, Oh J, Montgomery S. Social Media Use for Health Purposes by Chronic Disease Patients in the United States. Saudi J Med Med Sci 2021;9(1):51. [doi: 10.4103/sjmms.sjmms_262_20]
24. Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally Occurring Peer Support through Social Media: The Experiences of Individuals with Severe Mental Illness Using YouTube. PLoS ONE 2014 Oct 15;9(10):e110171. [doi: 10.1371/journal.pone.0110171]
25. Fazekas B, Megaw B, Eade D, Kronfeld N. Insights into the real-life experiences of people living with epilepsy: A qualitative netnographic study. Epilepsy & Behavior 2021 Mar;116:107729. [doi: 10.1016/j.yebeh.2020.107729]
26. 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]
27. 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 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
29. Keim-Malpass J, Steeves RH, Kennedy C. Internet ethnography: A review of methodological considerations for studying online illness blogs. International Journal of Nursing Studies 2014 Dec;51(12):1686-1692. [doi: 10.1016/j.ijnurstu.2014.06.003]
30. Copelton DA, Valle G. "You don't need a prescription to go gluten-free": The scientific self-diagnosis of celiac disease. Social Science & Medicine 2009 Aug;69(4):623-631. [doi: 10.1016/j.socscimed.2009.05.012]
31. Keim-Malpass J, Albrecht TA, Steeves RH, Danhauer SC. Young Women's Experiences With Complementary Therapies During Cancer Described Through Illness Blogs. West J Nurs Res 2013 Jun 30;35(10):1309-1324. [doi: 10.1177/0193945913492897]
32. Keim-Malpass J, Baernholdt M, Erickson J, Ropka M, Schroen A. Steeves RH: Blogging through cancer: young women's persistent problems shared online. Cancer nursing 2013;36(2):163-172. [doi: 10.1097/ncc.0b013e31824eb879]
33. Willis E. Patients' self-efficacy within online health communities: facilitating chronic disease self-management behaviors through peer education. Health Communication 2015 Sep 11;31(3):299-307. [doi: 10.1080/10410236.2014.950019]
34. Lee M. Don't Give Up! A Cyber-ethnography and Discourse Analysis of an Online Infertility Patient Forum. Cult Med Psychiatry 2017 Jan 6;41(3):341-367. [doi: 10.1007/s11013-016-9515-6]

XSL•FO
RenderX

35. Frohlich DO. Inflammatory Bowel Disease Patient Leaders Responsibility for Disseminating Health Information Online. Gastroenterology Nursing 2019;42(1):29-40. [doi: 10.1097/sga.0000000000000361]
36. Litchman ML, Lewis D, Kelly LA, Gee PM. Twitter Analysis of #OpenAPS DIY Artificial Pancreas Technology Use Suggests Improved A1C and Quality of Life. J Diabetes Sci Technol 2018 Sep 10;13(2):164-170. [doi: 10.1177/1932296818795705]
37. Lewis D, Long J, Moss N. Fowles ER: The Online Sharing of Instagram Images for Rheumatoid Arthritis. Computers, informatics, nursing : CIN 2020;39(2):89-96. [doi: 10.1097/cin.0000000000000660]
38. Gonzalez G, Vaculik K, Khalil C, Zektser Y, Arnold C, Almario CV, et al. Using Digital Ethnography to Understand the Experience of Women With Pelvic Organ Prolapse. Female Pelvic Med Reconstr Surg 2020 Sep 9;27(2):e363-e367. [doi: 10.1097/spv.0000000000000938]
39. Gibson F, Hibbins S, Grew T, Morgan S, Pearce S, Stark D, et al. How young people describe the impact of living with and beyond a cancer diagnosis: feasibility of using social media as a research method. Psycho-Oncology 2016 Jan 08;25(11):1317-1323. [doi: 10.1002/pon.4061]
40. Kendal S, Kirk S, Elvey R, Catchpole R, Pryjmachuk S. How a moderated online discussion forum facilitates support for young people with eating disorders. Health Expect 2016 Jan 03;20(1):98-111. [doi: 10.1111/hex.12439]
41. Lavis A, Winter R. #Online harms or benefits? An ethnographic analysis of the positives and negatives of peer‐support around self‐harm on social media. J Child Psychol Psychiatr 2020 May 27;61(8):842-854. [doi: 10.1111/jcpp.13245]
42. Shah S, Bradbury-Jones C, Taylor J. Using Facebook to tell stories of premature ageing and sexual and reproductive healthcare across the life course for women with cerebral palsy in the UK and USA. BMJ Open 2020 Feb 17;10(2):e032172. [doi: 10.1136/bmjopen-2019-032172]
43. Sloan M, Bosley M, Blane M, Holloway L, Barrere C, D'Cruz D, et al. 'But you don't look sick': a qualitative analysis of the LUPUS UK online forum. Rheumatol Int 2020 Oct 26;41(4):721-732. [doi: 10.1007/s00296-020-04726-x]
44. Winter R, Lavis A. The Impact of COVID-19 on Young People's Mental Health in the UK: Key Insights from Social Media Using Online Ethnography. Int J Environ Res Public Health 2021 Dec 30;19(1):e [FREE Full text] [doi: 10.3390/ijerph19010352] [Medline: 35010612]
45. Troisoeufs A. Deep Brain Stimulation (DBS) experiences: an ethnographic approach to their expression on the Internet forums. Med Health Care Philos 2019 Sep;22(3):343-352. [doi: 10.1007/s11019-018-9874-6] [Medline: 30377912]
46. Bayen S, Carpentier C, Baran J, Cottencin O, Defebvre L, Moreau C, et al. Parkinson's disease: Content analysis of patient online discussion forums. A prospective observational study using Netnography. Patient Education and Counseling 2021 Aug;104(8):2060-2066. [doi: 10.1016/j.pec.2021.01.028]
47. Fayn M, des Garets V, Rivière A. Collective empowerment of an online patient community: conceptualizing process dynamics using a multi-method qualitative approach. BMC Health Serv Res 2021 Sep 13;21(1):958 [FREE Full text] [doi: 10.1186/s12913-021-06988-y] [Medline: 34511115]
48. Fernandes LDS, Calado C, Araujo CAS. Social networks and health practices: influence of a diabetes online community on adherence to treatment. Cien Saude Colet 2018 Oct;23(10):3357-3368 [FREE Full text] [doi: 10.1590/1413-812320182310.14122018] [Medline: 30365855]
49. Melo LPD, Cortez LCDA, Santos RDP. Is the chronicity of HIV/AIDS fragile? Biomedicine, politics and sociability in an online social network. Rev Lat Am Enfermagem 2020;28:e3298 [FREE Full text] [doi: 10.1590/1518-8345.4006.3298] [Medline: 32520246]
50. Nunes F, Andersen T, Fitzpatrick G. The agency of patients and carers in medical care and self-care technologies for interacting with doctors. Health Informatics J 2017 Jun 27;25(2):330-349. [doi: 10.1177/1460458217712054]
51. Gajaria A, Yeung E, Goodale T, Charach A. Beliefs about attention-deficit/hyperactivity disorder and response to stereotypes: youth postings in Facebook groups. J Adolesc Health 2011 Jul;49(1):15-20. [doi: 10.1016/j.jadohealth.2010.09.004] [Medline: 21700151]
52. Loureiro-Koechlin C, Córdoba-Pachón JR. Assessing online ethnography as a form of interpretive research in information systems. School of Management, Royal Holloway, University of London, Working Paper Series SoMWP–1201. 2012. URL: https://repository.royalholloway.ac.uk/file/4037eb3c-f4c3-7658-7917-d4760d6cf34c/1/SoMWP1201-Loureiro-Cordoba.pdf [accessed 2022-10-01]
53. Cherif E, Martin-Verdier E, Rochette C. Investigating the healthcare pathway through patients' experience and profiles: implications for breast cancer healthcare providers. BMC Health Serv Res 2020 Aug 11;20(1):735 [FREE Full text] [doi: 10.1186/s12913-020-05569-9] [Medline: 32781993]
54. Kirk S, Milnes L. An exploration of how young people and parents use online support in the context of living with cystic fibrosis. Health Expect 2015 Feb 17;19(2):309-321. [doi: 10.1111/hex.12352]
55. Bernell S, Howard SW. Use Your Words Carefully: What Is a Chronic Disease? Front Public Health 2016 Aug 02;4:159 [FREE Full text] [doi: 10.3389/fpubh.2016.00159] [Medline: 27532034]
56. Bour C, Ahne A, Schmitz S, Perchoux C, Dessenne C, Fagherazzi G. The Use of Social Media for Health Research Purposes: Scoping Review. J Med Internet Res 2021 May 27;23(5):e25736. [doi: 10.2196/25736]
57. Lieberman A, Schroeder J. Two social lives: How differences between online and offline interaction influence social outcomes. Current Opinion in Psychology 2020 Feb;31:16-21. [doi: 10.1016/j.copsyc.2019.06.022]

XSL•FO
RenderX

58. Horrell LN, Hayes S, Herbert LB, MacTurk K, Lawhon L, Valle CG, et al. Telemedicine Use and Health-Related Concerns of Patients With Chronic Conditions During COVID-19: Survey of Members of Online Health Communities. J Med Internet Res 2021 Feb 18;23(2):e23795 [FREE Full text] [doi: 10.2196/23795] [Medline: 33539307]
59. Higgins V, Sohaei D, Diamandis EP, Prassas I. COVID-19: from an acute to chronic disease? Potential long-term health consequences. Critical Reviews in Clinical Laboratory Sciences 2020 Dec 21;58(5):297-310. [doi: 10.1080/10408363.2020.1860895]
60. Boschele M. COVID-19 Is a Crisis in Planetary Health and Politics of Expertise: Time to Think Critically and Innovate Both. OMICS 2021 May 01;25(5):279-284. [doi: 10.1089/omi.2021.0038] [Medline: 33961517]
61. Gómez EJ, Harris J. Political repression, civil society and the politics of responding to AIDS in the BRICS nations. Health Policy Plan 2015 Apr 08;31(1):56-66. [doi: 10.1093/heapol/czv021]
62. DeJean D, Giacomini M, Simeonov D, Smith A. Finding Qualitative Research Evidence for Health Technology Assessment. Qual Health Res 2016 Aug 26;26(10):1307-1317. [doi: 10.1177/1049732316644429] [Medline: 27117960]
63. Teasdale E, Clarke H, Chen N, Everitt H. Online forum users' views and experiences of managing irritable bowel syndrome: a qualitative analysis of discussion content. BJGP Open 2020 Oct 13;4(5):bjgpopen20X101084. [doi: 10.3399/bjgpopen20x101084]
64. Jamison J, Sutton S, Mant J, De Simoni A. Online stroke forum as source of data for qualitative research: insights from a comparison with patients' interviews. BMJ Open 2018 Mar 30;8(3):e020133. [doi: 10.1136/bmjopen-2017-020133]
65. Franz D, Marsh HE, Chen JI, Teo AR. Using Facebook for Qualitative Research: A Brief Primer. J Med Internet Res 2019 Aug 13;21(8):e13544. [doi: 10.2196/13544]
66. Zheng K, George M, Roehlkepartain E, Santelli J, Bruzzese J, Smaldone A. Developmental Assets of Adolescents and Young Adults With Chronic Illness and Comorbid Depression: Qualitative Study Using YouTube. JMIR Ment Health 2021 Feb 16;8(2):e23960. [doi: 10.2196/23960]
67. Giménez-Llort L. An Ethnography Study of a Viral YouTube Educational Video in Ecuador: Dealing With Death and Grief in Times of COVID-19. Front Psychiatry 2021;12:648569 [FREE Full text] [doi: 10.3389/fpsyt.2021.648569] [Medline: 34305667]
68. Zhang S, Yang Y, Yan D, Yuan B, Jiang X, Song C. Internet videos and colorectal cancer in mainland China: a content analysis. BMC Med Inform Decis Mak 2018 Dec 04;18(1):129 [FREE Full text] [doi: 10.1186/s12911-018-0711-x] [Medline: 30514290]
69. Brown C, Lloyd K. Qualitative methods in psychiatric research. Adv. psychiatr. treat 2018 Jan 02;7(5):350-356. [doi: 10.1192/apt.7.5.350]
70. Teli M, Pisanu F, Hakken D. The internet as a library-of-people: for a cyberethnography of online groups. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research 2007 Sep 30;8(3):e. [doi: 10.17169/fqs-8.3.283]
71. Murphy F. Preparing for the field: developing competence as an ethnographic field worker. Nurse Researcher 2005 Jan;12(3):52-60. [doi: 10.7748/nr2005.01.12.3.52.c5948]
72. Mohd Hanim MF, Md Sabri BA, Yusof N. Online News Coverage of the Sugar-Sweetened Beverages Tax in Malaysia: Content Analysis. JMIR Public Health Surveill 2021 Aug 18;7(8):e24523. [doi: 10.2196/24523]
73. Willis R. Observations online: Finding the ethical boundaries of Facebook research. Research Ethics 2017 Nov 11;15(1):1-17. [doi: 10.1177/1747016117740176]
74. Hetland P, Mørch A. Ethnography for investigating the internet. Seminar.net 2016 Nov 06;12(1):E. [doi: 10.7577/seminar.2335]
75. Gatson STM, politics, ethics ORIOE. The methods, politics, and ethics of representation in online ethnography. Collecting and interpreting qualitative materials 2011:245-275.
76. Jackson A, Brownlow C, Manchin T. Internet research and the implications for ethical research practice. 2016 Presented at: Australasian Ethics Network Conference 2016; Nov 23-15, 2016; Adelaide, Australia.
77. Sveningsson M. Ethics in internet ethnography. In: Buchanan E, editor. Readings in Virtual Research Ethics: Issues and Controversies. Hershey, PA: IGI Global; 2004:45-61.
78. Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AYS, et al. Ethical Issues of Social Media Usage in Healthcare. Yearb Med Inform 2018 Mar 10;24(01):137-147. [doi: 10.15265/iy-2015-001]

Abbreviations

OpenAPS: Open Artificial Pancreas System
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

XSL•FO
RenderX

Edited by R Kukafka; submitted 12.03.22; peer-reviewed by E Rubenstein, S Yamane; comments to author 27.07.22; revised version received 05.09.22; accepted 01.11.22; published 29.11.22.

Please cite as:
Gao Y, Chen X, Zhang W, Wang Q, Liu J, Zhou L
Online Ethnography for People With Chronic Conditions: Scoping Review
J Med Internet Res 2022;24(11):e37941
URL: https://www.jmir.org/2022/11/e37941
doi:10.2196/37941
PMID:36445746

©Yajing Gao, Xuemei Chen, Wei Zhang, Qiuyi Wang, Jing Liu, Lanshou Zhou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Ethical and Methodological Considerations of Twitter Data for Public Health Research: Systematic Review

Courtney Takats[1], MPH; Amy Kwan[1], MPH, DrPH; Rachel Wormer[1], MPH; Dari Goldman[1], MPH; Heidi E Jones[1], MPH, PhD; Diana Romero[1], MA, MPhil, PhD

City University of New York School of Public Health, New York City, NY, United States

Corresponding Author:
Courtney Takats, MPH
City University of New York School of Public Health
55 W 125th St
New York City, NY, 10027
United States
Phone: 1 6313988664
Email: courttakats14@gmail.com

Abstract

Background: Much research is being carried out using publicly available Twitter data in the field of public health, but the types of research questions that these data are being used to answer and the extent to which these projects require ethical oversight are not clear.

Objective: This review describes the current state of public health research using Twitter data in terms of methods and research questions, geographic focus, and ethical considerations including obtaining informed consent from Twitter handlers.

Methods: We implemented a systematic review, following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, of articles published between January 2006 and October 31, 2019, using Twitter data in secondary analyses for public health research, which were found using standardized search criteria on SocINDEX, PsycINFO, and PubMed. Studies were excluded when using Twitter for primary data collection, such as for study recruitment or as part of a dissemination intervention.

Results: We identified 367 articles that met eligibility criteria. Infectious disease (n=80, 22%) and substance use (n=66, 18%) were the most common topics for these studies, and sentiment mining (n=227, 62%), surveillance (n=224, 61%), and thematic exploration (n=217, 59%) were the most common methodologies employed. Approximately one-third of articles had a global or worldwide geographic focus; another one-third focused on the United States. The majority (n=222, 60%) of articles used a native Twitter application programming interface, and a significant amount of the remainder (n=102, 28%) used a third-party application programming interface. Only one-third (n=119, 32%) of studies sought ethical approval from an institutional review board, while 17% of them (n=62) included identifying information on Twitter users or tweets and 36% of them (n=131) attempted to anonymize identifiers. Most studies (n=272, 79%) included a discussion on the validity of the measures and reliability of coding (70% for interreliability of human coding and 70% for computer algorithm checks), but less attention was paid to the sampling frame, and what underlying population the sample represented.

Conclusions: Twitter data may be useful in public health research, given its access to publicly available information. However, studies should exercise greater caution in considering the data sources, accession method, and external validity of the sampling frame. Further, an ethical framework is necessary to help guide future research in this area, especially when individual, identifiable Twitter users and tweets are shared and discussed.

Trial Registration: PROSPERO CRD42020148170; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=148170

(J Med Internet Res 2022;24(11):e40380) doi:10.2196/40380

KEYWORDS

systematic review; Twitter; social media; public health ethics; public health; ethics; ethical considerations; public health research; research topics; Twitter data; ethical framework; research ethics

XSL•FO
RenderX

Introduction

Since its launch in 2006, Twitter has become one of the most popular social media sites as a platform that allows users to post and interact with short messages known as tweets. According to a 2019 survey by Pew Research Center [1], 1 in 5 (23%) adults in the United States report using Twitter. While Twitter users are not representative of the general population (users tend to be younger, more educated, and located in urban or suburban areas) [2], the volume of publicly available tweets allows for research to be conducted on large data sets, eschewing a common perceived limitation of small samples.

Public health researchers have identified "big data" from Twitter as a new wellspring from which research can be conducted [3]. However, the utility of these data depends on the appropriateness of the research questions and the methodological approaches used in sampling and analyzing the data. Previous systematic reviews have explored how Twitter data have been used. A systematic review by Sinnenberg et al [4] of 137 articles using Twitter in health research between 2010 and 2015 found that the main research questions explored with Twitter data involved content analysis, surveillance, engagement, recruitment, intervention, and network analysis. Similarly, a scoping review from 2020 [5] found 92 articles that fell within 6 domains: surveillance, event detection, pharmacovigilance, forecasting, disease tracking, and geographic identification. Additional systematic reviews of social media, beyond Twitter alone, have examined specific domains, for instance, exploring how these data, including Twitter, are being used for public health surveillance [6-8] or pharmacovigilance [9-11].

While social media provides new opportunities for data sources in research, some unique obstacles are also present. For instance, the presence of spam and noisy data can make it difficult for researchers to identify a legitimate signal for the research topic in question [12]. To navigate this issue, researchers sometimes opt to employ traditional manual coding of content; however, this can be a nonideal solution given the size of the data sets and the time and effort required for these analyses [13]. Other teams have used natural language processing (NLP) or machine learning approaches, which present their own problems; one study [14] found that among the algorithms built to classify emotions, the highest performing model had an accuracy of 65%. The landscape of social media necessitates understanding of the mechanisms and limitations of the platforms, as well as adaptations to the requirements of this landscape.

In addition to the research questions and methodological approaches used with Twitter data, the extent to which social media data are in general considered public, and what this means for ethical research oversight are unclear. There is substantial literature discussing the ethics of using social media data for public health research, but clear ethical guidelines have not been established [15-24].

The need for these guidelines is increasingly pressing, as leveraging social media for public health research raises questions about privacy and anonymity; properly deidentifying user data requires the researchers to understand an "increasingly networked, pervasive, and ultimately searchable dataverse" [18]. Information shared on social media can often be intensely personal; hence, anonymity would be even more important for research involving sensitive data such as health conditions and disease [23]. This is particularly relevant for the field of public health, since the data collected and analyzed for public health research will often fall into these more sensitive categories.

Beyond the questions of user anonymity, when conducting research on more sensitive health information, traditional research protocols center the importance of informed consent among participants. However, there are currently no established guidelines for the expectation of consent when leveraging publicly available social media data. Some theorists in the realm of internet research ethics have proposed an assessment model that determines the need for consent based on possibility of pain or discomfort. They further suggest that this assessment should consider the vulnerability of the population being studied and the sensitivity of the topics [22].

In the systematic review by Sinnenberg et al [4], approximately one-third of the 137 articles included therein mentioned ethical board approval. Given that Twitter usage has changed dramatically in recent years [25], this systematic review is an updated examination of both ethical considerations and research questions or methodologies across all domains of public health research using Twitter.

We sought to investigate the methodological and ethical aspects of using Twitter data for public health research from 2006, when Twitter was launched, to 2019 [26]. Specifically, we describe the measures being used in Twitter research, the extent to which they are validated and reliable, and the extent to which ethical oversight is included in studies using publicly available tweets.

Methods

Design

This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [27,28] and was registered with PROSPERO (CRD42020148170).

Eligibility Criteria

The database search was limited to peer-reviewed public health studies originally written in English, which were published between January 2006 and October 31, 2019, and used social media data to explore a public health research question. The social media platforms included in the search were Twitter and Sina Weibo (China's version of Twitter), Facebook, Instagram, YouTube, Tumblr, or Reddit.

Studies were excluded if they were systematic or literature reviews, marketing or sales research, only investigated organizational-level tweets, investigated tweets from conferences in disciplines other than public health, or included primary data collection asking participants about their social media use. We excluded articles that focused on organizations disseminating information to the public (evaluation of social media dissemination and analysis of organizational- or institutional-level social media data) or testing interventions that used social media as a method (intervention study using social media), as our research question was not related to

XSL•FO
RenderX

interventions using social media platforms as a tool but rather explored how existing social media data are being used in secondary analyses in public health research.

Given the volume of studies identified, separate analyses were conducted on Facebook and YouTube; thus, this systematic review focuses solely on Twitter. Studies that included Twitter and other social media platforms were included, but only Twitter findings were extracted.

Information Sources

We searched PubMed, SocINDEX, and PsycINFO for articles about social media and public health after consulting with our institutional librarian on the best approaches to the search.

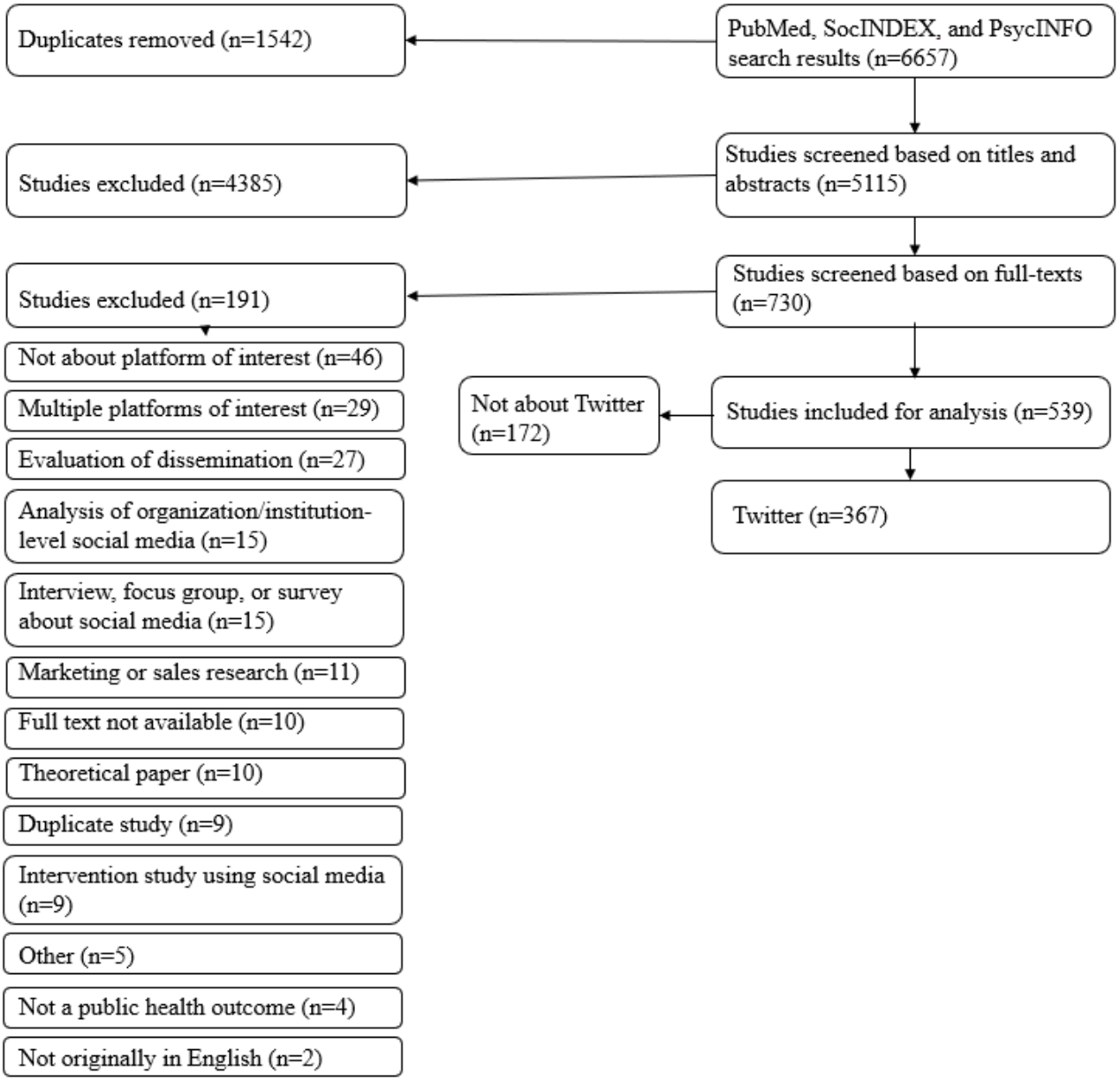
Search

The search strategy consisted of the Boolean search term: ((“Social media” OR twitter OR tweet* OR facebook OR instagram OR youtube OR tumblr OR reddit OR “web 2.0” OR “public comments” OR hashtag*) AND (“public health” OR “health research” OR “community health” OR “population health”)).

Study Selection

Three authors reviewed abstracts for eligibility in a 2-step process, with each abstract reviewed by 2 authors independently. A first screen was performed on the basis of the title and abstract; if deemed ineligible, the study was excluded from further screening. Disagreements were resolved through discussion and consensus. Full texts of the remaining articles were retrieved for the second screen and reasons for exclusion were coded and ranked by the priority of exclusion criteria for cases in which more than one exclusion criterion was applied (Figure 1). Disagreements about inclusion and exclusion criteria were resolved through discussion and consensus.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for systematic review of methodological approaches and ethical considerations for public health research using Twitter data, 2006-2019.



XSL•FO
RenderX

Data Collection Process

Data were extracted using a standardized data extraction spreadsheet, which was developed a priori and refined during the data extraction process. This refinement resulted in the removal of data elements; new data elements were not added. To establish consistency in extractions, 2 reviewers independently extracted data from the same 5 articles and compared the results. This process continued during weekly meetings, in which papers of varying complexity were discussed until consensus was reached. No studies were excluded on the basis of their quality.

Data Items

The data items in this review categorized information about the study within 4 domains: (1) study characteristics: public health topic, year, and country of publication; (2) study design and results: sample size, Twitter data extraction method, operationalization (ie, which data points were collected from social media posts and how researchers quantified these data), methodologic and analytic approaches, primary results, and descriptions of linking or account data; (3) ethical considerations: ethical approval, discussion of informed consent, and general discussion of ethical issues; and (4) risk of bias or methodological checks: quality assessment, validity, reliability, and accuracy checks implemented. We defined methodological approach as the overall objective of a research project coupled with the operationalization of methods to fulfill this objective.

Quality assessment metrics were adapted from existing quality assessment tools used for systematic reviews [29-31]. The specific quality assessment metrics were the following: whether the stated research question matches the data-defined research question, the presence of a clearly defined objective or hypothesis, validity of measures, reliability of measures, validation of computer algorithms, whether the data analysis is sufficiently grounded, whether findings logically flow from the analysis and address the research questions, and the presence of a clear description of limitations. A study was considered to have addressed validity if the measures used were based on validated measures, previous studies, or existing frameworks. A study addressed reliability if manual coding efforts incorporated checks or assessed intercoder reliability, descriptions of reliability were not expected for studies that only used machine learning. Accuracy checks were described if manual checks were performed by researchers or validation of computer algorithms used for studies using machine learning algorithms and NLP.

Summary Measures

The summary measures related to methods and study design include the following: the frequency of studies by topic, geographic focus, year of publication, analytic approach, sampling approach, and overall methodological approach or objective of the study (ie, surveillance, content exploration, sentiment mining, network science, and model development and testing). The summary measures related to ethical considerations include the frequency of studies that sought institutional review board (IRB) review or approval, included informed consent from Twitter handlers, discussed ethical considerations within the paper, and reported identifying results (ie, verbatim tweets). For quality assessment, we present information on the validity and reliability of measures used; a full summary of quality assessments is provided in Multimedia Appendix 1.

Results

Our search resulted in 6657 unique studies for review, of which 730 required full-text review (Figure 1). We identified 539 studies across all social media platforms; 367 used Twitter data forming the analytic sample for this review (Multimedia Appendix 2 for the full list of included articles with all data extraction fields; for readability of text, references are only included when details of specific articles are provided as contextual examples).

Study Characteristics

Public Health Research Topics

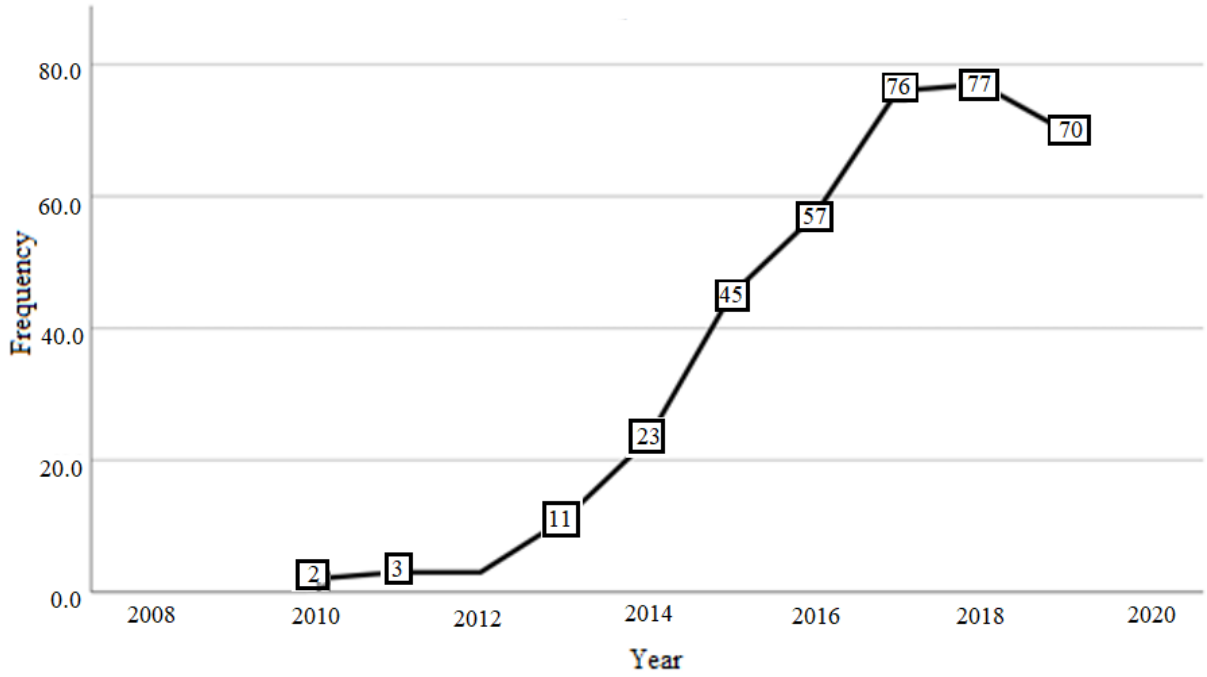
The most common public health topics among the articles reviewed were communicable diseases (eg, influenza, Ebola, and Zika; n=80, 22%), substance use (n=66, 18%), health promotion (n=63, 17%), chronic disease (eg, cancer; n=48, 13%), and environmental health (n=48, 13%; Multimedia Appendix 1).

Year of Publication

The year of publication for the articles in this review ranged from 2010 to 2019. A sharp increase in the number of Twitter articles was observed from 2012 to 2017 (Figure 2). Two preprint articles on October 31, 2019, were included in the count for 2019 [32,33].

XSL•FO
RenderX

Figure 2. Number of articles published by year for systematic review of methodological approaches and ethical considerations for public health research using Twitter data, 2006-2019.



Geographic Focus

Most studies analyzed tweets originating from the United States (n=158, 43%) or worldwide (n=134, 36%); only 75 (20%) of them focused on non-US regions or countries. Of the articles that had a global geographic focus, 23 (17%) of them collected geotags and reported on geospatial metrics within the body of the article. Despite having a worldwide focus, these 23 articles demonstrated a bias toward the United States, western Europe (namely the United Kingdom), Canada, and Australia; the majority of the data collected in these studies were posts originating in these countries, with a distinct minority representing other regions or countries.

Study Design and Results

Sample Size and Unit of Analysis

Of the 367 articles reviewed here, 355 (97%) used individual tweets as the unit of analysis and 11 (3%) used Twitter accounts (or "handles") as the unit of analysis. One article (0.3%) used keywords as the unit of analysis, as the study sought to identify keywords that would help researchers detect influenza epidemics via Twitter [34].

There was a wide range of sample sizes. For studies with tweets as the unit of analysis (n=353), the number of analyzed tweets ranged from 82 [35] to 2.77 billion [36] (median=74,000), with 90 papers having a sample size larger than 1 million. Similarly, for studies using Twitter handles as the unit of analysis (n=11), the sample size ranged from 18 [37] to 217,623 [32].

Methods for Accessing Data

To pull data from Twitter, most studies used application programming interfaces (APIs) that were developed by Twitter (eg, Gardenhose and Firehose) and could be integrated into statistical software packages. Third-party APIs (eg, Twitonomy and Radian6) were also used frequently, either through contracting with a commercial vendor, purchasing tweets that match specified criteria, or using software developed by an entity outside of Twitter. Most studies either mentioned that they used an API without indicating the specific type (37%) or did not mention their method of tweet accession (13%; Table 1). Of papers that identified the API used, purposive and random sampling were equally employed. However, only 22 (7%) articles explicitly mentioned whether the API used was purposive or random in its sampling technique; when the API was named (eg, decahose, search API, and Gardenhose) but the sampling type was not noted in the article, we looked up the sampling technique in use by the API.

We also found that the description of the sampling method was often not described. For instance, some Twitter APIs are purposive in nature (eg, Twitter Search API) and some are random (Twitter Firehose API) or systematic (some REST APIs). Many studies did not specify what type of sampling was used to extract tweets from Twitter or did not fully explain retrieval limitations (eg, how it might affect the sample population if only a certain number of tweets could be retrieved daily through an API).

XSL•FO
RenderX

Table 1. Frequency of studies by access method and data source from a systematic review of methodological approaches and ethical considerations for public health research using Twitter data, 2006-2019.

Method or source for Twitter data	Frequency (N=367), n (%)
Access method	
Unspecified application programming interface (API)	136 (37)
Purposive sampling[a]	88 (24)
Random sampling[a]	84 (23)
Existing database	10 (3)
Unspecified method of accession	49 (13)
Data source	
Native Twitter API/functionality	222 (60)
Third-party API	102 (28)
Unknown	34 (9)
In-house program	9 (3)

[a]Accession methods and sampling type are differentiated as random or purposive in accordance with reports from the articles' authors or Twitter.

Methodological Approach

As seen in Table 2, the most common methodological approaches were as follows: thematic exploration (eg, describing the themes of conversations about e-cigarettes on Twitter) [38], sentiment mining (eg, assessing if tweets about vaccines are positive, negative, or neutral) [39], and surveillance (eg, tracking the patterns of information spread about an Ebola outbreak) [40]. Less common methodological approaches were tool evaluation (eg, using Twitter data to predict population health indices) [41] and network science (eg, examining health information flows) [42]. Different methodological approaches tended to be pursued for different topics. For example, most infectious disease research was in the domain of surveillance, whereas research about mental health and experiences with the health care system was more conducive to thematic exploration and sentiment mining.

Across the 3 most common study methodological approaches (thematic exploration, sentiment mining, and surveillance), approximately one-third of the papers (36%) used machine learning (Table 2). Machine learning here is defined as an application of algorithms and statistical modeling to reveal patterns and relationships in data without explicit instruction (eg, to identify the patterns of dissemination related to Zika virus–related information on Twitter) [43]. This can be contrasted to NLP, which necessitates explicit instruction; often, NLP is used to identify and classify words or phrases from a predefined list in large data sets (eg, to identify the most common key topics used by Twitter users regarding the opioid epidemic) [44]. Of the articles reviewed, NLP was more prevalent in sentiment mining than in other types of methodological approaches.

XSL•FO
RenderX

Table 2. Frequency of studies by methodological approach and analytical technique from a systematic review of methodological approaches and ethical considerations for public health research using Twitter data, 2006-2019.

Methodological approach and analytical technique[a]	Frequency (N=367), n (%)
Sentiment mining	227 (62)
Natural language processing	145 (64)
Machine learning	66 (29)
Spatial analysis	12 (5)
Descriptive analyses or frequencies	4 (2)
Surveillance	224 (61)
Natural language processing	104 (46)
Machine learning	85 (38)
Spatial analysis	17 (8)
Descriptive analyses or frequencies	18 (8)
Thematic exploration	217 (59)
Natural language processing	114 (52)
Machine learning	81 (37)
Spatial analysis	13 (6)
Descriptive analyses or frequencies	9 (4)
Tool evaluation	61 (16)
Network science	36 (10)

[a]Multiple responses were allowed.

Ethical Considerations

Presence of Identifying Information

Just under half (n=174, 47%) of the articles reviewed did not contain any identifying information of Twitter accounts or tweets, 36% (n=131) of them contained anonymized account information or paraphrased tweets, and 17% (n=62) of them contained direct quotes of tweets or identifiable information such as Twitter handles or account names (Table 3). Of the 62 articles that included verbatim tweets or identifying information about the user, one-third (n=21, 34%) of them included a discussion of ethics in the paper (eg, Berry et al [45]).

Less than half of the articles (n=173, 47%) indicated that they did not use any of the metadata (eg, username, demographics, and geolocation) associated with the tweet (Multimedia Appendix 1). Approximately one-third of the articles (n=110, 30%) used geographic information associated with the tweet, and a much smaller number of articles (n=15, 4%) included photos associated with the account or health information (such as illness disclosure or mentions of medications taken). Of the articles analyzing tweets from either the United States or another specific region or country (n=233), 37% (n=86) of them used geotags of Twitter accounts to identify the location of the tweets; of the articles that did not specify a geographic region (n=134), 17% (n=23) of them used geotagging.

Though research on infectious disease and health promotion were most likely to include user metadata in their data analyses, linked health information was most often used in papers about infectious disease and mental health, often in the form of medical self-disclosures.

IRB Approval and Informed Consent

Just under one-third of the articles reviewed (n=119; 32%) explicitly stated that those studies sought and received IRB review or approval (Table 3). The majority (n=226, 61%) of them did not mention IRB approval, although many of these articles included statements about the nature of Twitter posts being publicly available. Only a small subset (n=23, 6%) of studies explicitly stated that IRB approval was not necessary.

Among those that sought IRB approval (n=119), over half (n=68, 57%) of them were granted exemptions; just under half (n=49, 41%) of them did not specify the type of approval received. Two studies [46,47] received full IRB approval. One of them [46] retrospectively examined existing public data about health beliefs regarding the human papillomavirus and was approved with a waiver of consent owing to its retrospective design. The other study [47] had 2 parts: study 1 consisted of a survey of self-reported stress following a school lockdown, and study 2 consisted of data mining of community-level rumor generation during the lockdown on Twitter. The survey necessitated informed consent as it involved human participants; hence, the full scope of the study (parts 1 and 2) had to undergo IRB review. None of the studies using only Twitter data sought informed consent, even when including identifying information from Twitter handlers or tweets. Over two-thirds of the articles (n=258, 70%) did not include a discussion of ethics or privacy concerns.

XSL•FO
RenderX

Additionally, 53 (49%) articles discussed the anonymization of data used in their study either by omitting usernames and Twitter handles [48] or by providing only paraphrased tweets to prevent exact-match searching [49]. Only 5 studies included specific and extensive discussions around the ethical implications of social media research and went beyond disclaimer statements about the publicly available nature of tweets. One study [50] described consulting guidelines for internet research from various organizations and researchers, while another [51] included a long "ethical considerations" section that described needing to "weigh threats to safety and privacy against benefits gained by using novel approaches to study suicide," and acknowledged vulnerable populations and risks of stigma and discrimination. Another study [52] raised the challenge of social media research given the lack of relevant ethical frameworks.

Table 3. Frequency of studies by ethics-related factors from a systematic review of methodological approaches and ethical considerations for public health research using Twitter data, 2006-2019.

Ethics-related factors	Frequency (N=367), n (%)
Level of identification	
No identifying information	174 (47)
Anonymized data and paraphrased tweets	131 (36)
Identifiable information and direct quotes	62 (17)
Institutional review board (IRB) approval obtained	
Yes	119 (32)
No	23 (6)
Not mentioned/unclear	225 (61)
Among those with IRB approval (n=119)	
Exempt	68 (57)
Nonexempt	2 (2)
Not specified (eg, "approved")	49 (41)
Informed consent of Twitter handler attempted	
Yes	0 (0)
No	119 (100)
Any discussion of ethical considerations, including disclaimers	
Yes[a]	109 (30)
Discussion of anonymization process	53 (49)
Extensive discussion[b]	5 (5)
Other discussion, including disclaimers	54 (49)
No	258 (70)

[a]Note that 3 articles included both an extensive discussion of ethics as well as details regarding their anonymization process.

[b]The denominator for the articles that discussed ethics is 109.

Risk of Bias in Individual Studies

We found that 270 (74%) articles included a clear description of the validity of measures; 21 (6%) articles were purely exploratory in nature and collected only counts of tweets, so we deemed them exempt from an assessment of validity of measures; 76 (21%) articles did not include efforts at establishing measurement validity. Further, of the 264 articles involving human coding, 184 (70%) included a description of intercoder reliability and quality assurance checks, while 80 (30%) did not. Similarly, 235 articles involved computer algorithms or automated coding, of which 165 (70%) explicitly described accuracy checks or validation of the algorithms, while 70 (39%) did not.

In addition to concerns about validity and reliability of measures, one of the main sources of bias was the sampling frame. The self-selection of Twitter users was discussed in most of the studies, with 85% (n=314) of them describing this as a potential limitation.

Discussion

Principal Findings

Summary Measures

We saw evidence of a steep increase in publications using Twitter data after 2012, which may be due to Twitter releasing its native standard (version 1.1) API in 2012, which made mining of its data much more accessible to the general public

XSL•FO
RenderX

without the need for complex coding capabilities [53]. The prevalence of research using "big data" from Twitter is increasing and will likely continue to do so in the coming years [50].

Infectious disease was the most common topic of the research papers, which may indicate a burgeoning interest in using social media to detect disease outbreaks. It is likely that a review of studies using Twitter data that picks up from where this study left off (ie, after October 31, 2019) would support this finding given the onset of the COVID-19 pandemic in late 2019.

There are some major considerations that this review highlights for the future of public health research using Twitter data. Most of the research focused on Twitter users in the United States; this includes the articles with a global focus that demonstrated a bias toward the anglophone world. Three articles appeared to genuinely have a representative global scope; interestingly, two of these were about the Zika virus. This indicates the data scraped from Twitter tends to be heavily focused on the United States and English-speaking settings.

Another major consideration is that of the accession method used to build a data set. Most of the studies examined in this review used APIs or variations thereof; only 10 studies used alternative accession methods. Those 10 studies used data either extracted from Twitter for previous studies or hosted in pre-existing databases. Of the remaining studies that used an API, only 22 studies explained whether the API used was purposive or random in nature. This is of interest because the sampling technique of APIs has been called into question in previous papers [54,55]. In particular, the Twitter Streaming API is considered to produce less representative samples and should be approached with caution; this API is susceptible to intentional or accidental bias based on inclusion and exclusion criteria selected for a particular study [56]. Owing to the "black box" nature (ie, lack of documentation of the sampling approach) of native Twitter APIs, it cannot be determined that data retrieved using Twitter APIs are truly random [57,58].

In addition to the aforementioned obstacles, there are questions about the accuracy of algorithms using machine learning and NLP. A little less than half of the papers reviewed for this systematic review involved surveillance and prediction, and approximately one-sixth of them evaluated new tools or frameworks in the realm of Twitter data. Machine learning was commonly used for these methodological approaches. However, a previous evaluation of the efficacy of using various machine learning algorithms to automatically identify emotions expressed on Twitter found that the highest performing algorithm achieved an accuracy rate of 65% [14]. Another recent article found that machine learning was not effective in making meaningful predictions about users' mental health from language use on social media; further, Twitter metadata and language use was not specific to any one mental health condition [59].

This raises concerns about the overall use of social media data for research, as data science in general and public health research in particular use data to make insights; these data "then get acted upon and the decisions impact people's lives" [20]. Hence, conscientious planning is advised when using publicly available social media data for the purpose of public health research.

Discussion of Ethics

Given that slightly over one-third of studies anonymized Tweets or Twitter users, many researchers seem to think that there are ethical considerations when using these data, even if they are publicly available. Nevertheless, the majority of projects did not seek IRB review or approval. This contradiction suggests an implicit understanding that while there are no international or place-specific ethical guidelines around research using social media data, there is something unique about the nature of this research that distinguishes it from truly public data.

International ethical standards for biomedical and public health research already exist, and these standards often continue to influence the national guidelines that develop within a given country [60-62]. Given the global scope of social media, it may be most prudent for guidelines to be established on an international scale and then adapted to place-specific committees and ethics boards. However, this is complicated by the ever-evolving landscape of social media use and data agreements. The field of research ethics has yet to fully address the introduction of new media as sources of data; even before a comprehensive international framework is introduced, it may be advisable for institutions and regions to enact their own interim frameworks to mitigate possible harm and preserve user privacy and anonymity to the extent possible.

Limitations

This systematic review has a number of limitations. Owing to the iterative nature of data extraction for a large number of articles included, it is possible that there were differences in how data were coded as we refined our process. However, we attempted to minimize this concern through weekly research team meetings during the extraction process. Another limitation is that because we only examined articles originally published in English, we may be underestimating the number of articles that were conducting research in a specific geographic area other than the United States. The influence of this underestimation should be minimal; however, as most leading journals for health research are published in English [63]. One final limitation is that the literature review spanned from 2010 to 2019, so we are not capturing changes since then, which may have taken place in the approach to ethics or methodology in research using social media data since then. This is an evolving field of research; hence, we anticipate that standards and norms may have also evolved.

Comparison With Prior Work

Similar to Sinnenberg et al's [4] review, this study examined whether ethics board approvals were sought when using social media data for public health research, finding equivalent proportions of articles that obtained IRB approval. Our study further explored whether there were other types of ethical considerations (eg, ethical discussion) present in the body of the articles. We also assessed the presence and use of identifiable information such as personal health information, verbatim Tweets, and user account metadata. In both this review and in that of Sinnenberg et al [4], many articles noted that the public nature of tweets allows researchers to observe the content. This presents a clear need for an ethical guideline framework for

XSL•FO
RenderX

researchers using Twitter, especially when including identifying information.

Conclusions

Twitter data appear to be an increasingly important source of data in public health research. However, attention needs to be paid to sampling constraints, ethical considerations involved in using these data, and the specific methodologies to be used to ensure the rigorous conduct of this research.

Acknowledgments

We would like to thank Sarah Pickering, MPH, Jessie Losch, MPH, and Rebecca Berger, MPH, graduate students at the City University of New York (CUNY) School of Public Health, who contributed to refinement of the data extraction forms, data extraction, and quality assessments. This study was partially funded by an anonymous private foundation. The foundation did not play any role in implementation of the systematic review or manuscript preparation. The authors have no financial disclosures to report.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 16 KB - jmir_v24i11e40380_app1.docx]

Multimedia Appendix 2
Full data extraction sheet.
[XLSX File (Microsoft Excel File), 165 KB - jmir_v24i11e40380_app2.xlsx]

References

1. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-11-02]
2. Wojcik S, Hughes A. Sizing up Twitter users. Pew Research Center. 2019 Apr 24. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2021-11-02]
3. Paul M, Dredze M. You are what you tweet: analyzing Twitter for public health. ICWSM 2021 Aug 03;5(1):265-272. [doi: 10.1609/icwsm.v5i1.14137]
4. 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]
5. Edo-Osagie O, De La Iglesia B, Lake I, Edeghere O. A scoping review of the use of Twitter for public health research. Comput Biol Med 2020 Jul;122:103770 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103770] [Medline: 32502758]
6. Jordan S, Hovet S, Fung I, Liang H, Fu K, Tse Z. Using Twitter for public health surveillance from monitoring and prediction to public response. Data 2018 Dec 29;4(1):6. [doi: 10.3390/data4010006]
7. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: a systematic review. J Biomed Inform 2020 Aug;108:103500 [FREE Full text] [doi: 10.1016/j.jbi.2020.103500] [Medline: 32622833]
8. Schwab-Reese LM, Hovdestad W, Tonmyr L, Fluke J. The potential use of social media and other internet-related data and communications for child maltreatment surveillance and epidemiological research: Scoping review and recommendations. Child Abuse Negl 2018 Nov;85:187-201 [FREE Full text] [doi: 10.1016/j.chiabu.2018.01.014] [Medline: 29366596]
9. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
10. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse drug reaction identification and extraction in social media: a scoping review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
11. Tricco AC, Zarin W, Lillie E, Jeblee S, Warren R, Khan PA, et al. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. BMC Med Inform Decis Mak 2018 Jun 14;18(1):38 [FREE Full text] [doi: 10.1186/s12911-018-0621-y] [Medline: 29898743]
12. Karpf D. Social science research methods in internet time. iCS 2012 Jun;15(5):639-661. [doi: 10.1080/1369118x.2012.665468]
13. Giglietto F, Rossi L, Bennato D. The open laboratory: limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. J Technol Hum Serv 2012 Jul;30(3-4):145-159. [doi: 10.1080/15228835.2012.743797]

XSL•FO
RenderX

14. Wang W, Chen L, Thirunarayan K, Sheth A. Harnessing twitter "big data" for automatic emotion identification. 2012 Presented at: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conferenece on Social Computing; September 3-5, 2012; Amsterdam. [doi: 10.1109/socialcom-passat.2012.119]
15. Zimmer M. “But the data is already public”: on the ethics of research in Facebook. Ethics Inf Technol 2010 Jun 4;12(4):313-325. [doi: 10.1007/s10676-010-9227-5]
16. Fiesler C, Proferes N. “Participant” perceptions of Twitter research ethics. SM+S 2018 Mar 10;4(1):205630511876336. [doi: 10.1177/2056305118763366]
17. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. J Med Internet Res 2017 Jun 06;19(6):e195 [FREE Full text] [doi: 10.2196/jmir.7082] [Medline: 28588006]
18. Hunter RF, Gough A, O’Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/ajph.2017.304249]
19. Rivers CM, Lewis BL. Ethical research standards in a world of big data. F1000Res 2014 Aug 21;3:38. [doi: 10.12688/f1000research.3-38.v2]
20. MacPherson Y, Pham K. Ethics in health data science. In: Celi LA, Majumder MS, Ordóñez P, Osorio JS, Paik KE, Somai M, editors. Leveraging data science for global health. Cham: Springer; 2020:365-372.
21. Vitak J, Proferes N, Shilton K, Ashktorab Z. Ethics regulation in social computing research: examining the role of institutional review boards. J Empir Res Hum Res Ethics 2017 Dec 23;12(5):372-382. [doi: 10.1177/1556264617725200] [Medline: 28831844]
22. Elgesem D, Ess C, Larsson A, Lüders M, Prabhu R, Segadal KU, et al. In: Fossheim H, Ingierd H, editors. Internet research ethics. Oslo: Cappelen Damm Akademisk; 2016.
23. Taylor J, Pagliari C. Mining social media data: how are research sponsors and researchers addressing the ethical challenges? Research Ethics 2017 Oct 26;14(2):1-39. [doi: 10.1177/1747016117738559]
24. Zook M, Barocas S, Boyd D, Crawford K, Keller E, Gangadharan SP, et al. Ten simple rules for responsible big data research. PLoS Comput Biol 2017 Mar 30;13(3):e1005399 [FREE Full text] [doi: 10.1371/journal.pcbi.1005399] [Medline: 28358831]
25. Iqbal M. Twitter Revenue and Usage Statistics (2022). Business of Apps. URL: https://www.businessofapps.com/data/twitter-statistics/ [accessed 2022-02-19]
26. History.com Editors. Twitter launches. A&E Television Networks. 2019. URL: https://www.history.com/this-day-in-history/twitter-launches [accessed 2022-02-19]
27. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
28. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
29. Wong WC, Cheung CS, Hart GJ. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. Emerg Themes Epidemiol 2008 Nov 17;5(1):23 [FREE Full text] [doi: 10.1186/1742-7622-5-23] [Medline: 19014686]
30. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
31. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun 01;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
32. Talbot C, O'Dwyer S, Clare L, Heaton J, Anderson J. Identifying people with dementia on Twitter. Dementia (London) 2020 May 06;19(4):965-974. [doi: 10.1177/1471301218792122] [Medline: 30081665]
33. Oh HJ, Kim CH, Jeon JG. Public sense of water fluoridation as reflected on Twitter 2009-2017. J Dent Res 2020 Jan;99(1):11-17. [doi: 10.1177/0022034519885610] [Medline: 31682777]
34. Woo H, Sung Cho H, Shim E, Lee JK, Lee K, Song G, et al. Identification of keywords from Twitter and web blog posts to detect influenza epidemics in Korea. Disaster Med Public Health Prep 2018 Jun 31;12(3):352-359. [doi: 10.1017/dmp.2017.84] [Medline: 28756796]
35. Squiers LB, Holden DJ, Dolina SE, Kim AE, Bann CM, Renaud JM. The public's response to the U.S. Preventive Services Task Force's 2009 recommendations on mammography screening. Am J Prev Med 2011 May;40(5):497-504. [doi: 10.1016/j.amepre.2010.12.027] [Medline: 21496748]
36. Qiu R, Hadzikadic M, Yu S, Yao L. Estimating disease burden using Internet data. Health Informatics J 2019 Dec 29;25(4):1863-1877 [FREE Full text] [doi: 10.1177/1460458218810743] [Medline: 30488754]
37. Rocheleau M, Sadasivam RS, Baquis K, Stahl H, Kinney RL, Pagoto SL, et al. An observational study of social and emotional support in smoking cessation Twitter accounts: content analysis of tweets. J Med Internet Res 2015 Jan 14;17(1):e18 [FREE Full text] [doi: 10.2196/jmir.3768] [Medline: 25589009]

XSL•FO
RenderX

38. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-Cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
39. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. Vaccine 2017 Jun 22;35(29):3621-3638 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.052] [Medline: 28554500]
40. Liang H, Fung IC, Tse ZTH, Yin J, Chan C, Pechta LE, et al. How did Ebola information spread on twitter: broadcasting or viral spreading? BMC Public Health 2019 Apr 25;19(1):438 [FREE Full text] [doi: 10.1186/s12889-019-6747-8] [Medline: 31023299]
41. Nguyen T, Larsen ME, O'Dea B, Nguyen DT, Yearwood J, Phung D, et al. Kernel-based features for predicting population health indices from geocoded social media data. Decis Support Syst 2017 Oct;102:22-31. [doi: 10.1016/j.dss.2017.06.010]
42. Cumbraos-Sánchez MJ, Hermoso R, Iñiguez D, Paño-Pardo JR, Allende Bandres M, Latorre Martinez MP. Qualitative and quantitative evaluation of the use of Twitter as a tool of antimicrobial stewardship. Int J Med Inform 2019 Nov;131:103955. [doi: 10.1016/j.ijmedinf.2019.103955] [Medline: 31487575]
43. Gui X, Wang Y, Kou Y, Reynolds TL, Chen Y, Mei Q, et al. Understanding the patterns of health information dissemination on social media during the Zika outbreak. AMIA Annu Symp Proc 2017;2017:820-829 [FREE Full text] [Medline: 29854148]
44. Glowacki EM, Glowacki JB, Wilcox GB. A text-mining analysis of the public's reactions to the opioid crisis. Subst Abus 2018 Sep 20;39(2):129-133. [doi: 10.1080/08897077.2017.1356795] [Medline: 28723265]
45. 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]
46. Du J, Cunningham RM, Xiang Y, Li F, Jia Y, Boom JA, et al. Leveraging deep learning to understand health beliefs about the Human Papillomavirus vaccine from social media. NPJ Digit Med 2019;2:27 [FREE Full text] [doi: 10.1038/s41746-019-0102-4] [Medline: 31304374]
47. Jones NM, Thompson RR, Dunkel Schetter C, Silver RC. Distress and rumor exposure on social media during a campus lockdown. Proc Natl Acad Sci U S A 2017 Oct 31;114(44):11663-11668 [FREE Full text] [doi: 10.1073/pnas.1708518114] [Medline: 29042513]
48. Donelle L, Booth R. Health Tweets: an exploration of health promotion on Twitter. Online J Issues Nurs 2012 Sep 30;17(3). [doi: 10.3912/ojin.vol17no03man04]
49. Waring M, Baker K, Peluso A, May CN, Pagoto SL. Content analysis of Twitter chatter about indoor tanning. Transl Behav Med 2019 Jan 01;9(1):41-47 [FREE Full text] [doi: 10.1093/tbm/iby011] [Medline: 29474700]
50. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
51. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring online discussions about suicide among Twitter users with schizophrenia: exploratory study. JMIR Ment Health 2018 Dec 13;5(4):e11483 [FREE Full text] [doi: 10.2196/11483] [Medline: 30545811]
52. Tian X, He F, Batterham P, Wang Z, Yu G. An analysis of anxiety-related postings on Sina Weibo. Int J Environ Res Public Health 2017 Jul 13;14(7):775 [FREE Full text] [doi: 10.3390/ijerph14070775] [Medline: 28703773]
53. Twitter API. Twitter. URL: https://developer.twitter.com/en/docs/twitter-api [accessed 2021-11-02]
54. Morstatter F, Pfeffer J, Liu H, Carley K. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. ICWSM 2021 Aug 03;7(1):400-408. [doi: 10.1609/icwsm.v7i1.14401]
55. Pfeffer J, Mayer K, Morstatter F. Tampering with Twitter's sample API. EPJ Data Sci 2018 Dec 19;7(1). [doi: 10.1140/epjds/s13688-018-0178-0]
56. Tromble R, Storz A, Stockmann D. We don't know what we don't know: when and how the use of Twitter's public APIs biases scientific inference. SSRN J 2017. [doi: 10.2139/ssrn.3079927]
57. Driscoll K, Walker S. Big data, big questions| Working within a black box: transparency in the collection and production of big Twitter data. Int J Commun 2014;8.
58. Wang Y, Callan J, Zheng B. Should we use the sample? Analyzing datasets sampled from Twitter's stream API. ACM Trans Web 2015 Jun 20;9(3):1-23. [doi: 10.1145/2746366]
59. Kelley SW, Mhaonaigh CN, Burke L, Whelan R, Gillan CM. Machine learning of language use on Twitter reveals weak and non-specific predictions. NPJ Digit Med 2022 Mar 25;5(1):35 [FREE Full text] [doi: 10.1038/s41746-022-00576-y] [Medline: 35338248]
60. Standards and operational guidance for ethics review of health-related research with human participants. World Health Organization. 2011. URL: https://www.who.int/publications/i/item/9789241502948 [accessed 2022-11-16]
61. International ethical guidelines for health-related research involving humans. Council for International Organizations of Medical Sciences. 2017. URL: https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf [accessed 2022-11-16]
62. WMA Declaration of Helsinki–ethical principles for medical research involving human subjects. World Medical Association. URL: https://tinyurl.com/bda43r5k [accessed 2022-11-21]

XSL•FO
RenderX

63. Yeung AWK, Kletecka-Pulker M, Eibensteiner F, Plunger P, Völkl-Kernstock S, Willschke H, et al. Implications of Twitter in health-related research: a landscape analysis of the scientific literature. Front Public Health 2021;9:654481 [FREE Full text] [doi: 10.3389/fpubh.2021.654481] [Medline: 34307273]

Abbreviations

API: application programming interface
IRB: institutional review board
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by R Kukafka; submitted 17.06.22; peer-reviewed by N Yiannakoulias, S Hiremath; comments to author 25.10.22; revised version received 08.11.22; accepted 13.11.22; published 29.11.22.

Please cite as:
Takats C, Kwan A, Wormer R, Goldman D, Jones HE, Romero D
Ethical and Methodological Considerations of Twitter Data for Public Health Research: Systematic Review
J Med Internet Res 2022;24(11):e40380
URL: https://www.jmir.org/2022/11/e40380
doi:10.2196/40380
PMID:36445739

©Courtney Takats, Amy Kwan, Rachel Wormer, Dari Goldman, Heidi E Jones, Diana Romero. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

The Patient Role in a Federal National-Scale Health Information Exchange

Joshua C Mandel[1,2], MD; J P Pollak[3], PhD; Kenneth D Mandl[4], MPH, MD

[1]Microsoft Healthcare, Redmond, WA, United States

[2]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

[3]Cornell Tech, New York, NY, United States

[4]Computational Health Information Program, Boston Children's Hospital, Boston, MA, United States

Corresponding Author:
Kenneth D Mandl, MPH, MD
Computational Health Information Program
Boston Children's Hospital
401 Park Drive
LM5506, Mailstop BCH3187
Boston, MA, 02215
United States
Phone: 1 617 355 4145
Email: kenneth_mandl@harvard.edu

Abstract

The federal Trusted Exchange Framework and Common Agreement (TEFCA) aims to reduce fragmentation of patient records by expanding query-based health information exchange with nationwide connectivity for diverse purposes. TEFCA provides a common agreement and security framework allowing clinicians, and possibly insurance company staff, public health officials, and other authorized users, to query for health information about hundreds of millions of patients. TEFCA presents an opportunity to weave information exchange into the fabric of our national health information economy. We define 3 principles to promote patient autonomy and control within TEFCA: (1) patients can query for data about themselves, (2) patients can know when their data are queried and shared, and (3) patients can configure what is shared about them. We believe TEFCA should address these principles by the time it launches. While health information exchange already occurs on a large scale today, the launch of TEFCA introduces a major, new, and cohesive component of 21st-century US health care information infrastructure. We strongly advocate for a substantive role for the patient in TEFCA, one that will be a model for other systems and policies.

(J Med Internet Res 2022;24(11):e41750) doi:10.2196/41750

KEYWORDS

health information exchange; patient control; Health Insurance Portability and Accountability Act; HIPAA; patient record; information exchange; information sharing; health record; privacy; security; public health; health policy; health information; federal trusted exchange; insurance company; patient data

Introduction

Since medical records are usually stored where they are produced, when patients traverse sites of care their information often becomes fragmented. The 21st Century Cures Act called for a Trusted Exchange Framework and Common Agreement (TEFCA) to enable turnkey access to medical histories across organizations. TEFCA, which is expected to be implemented starting in 2023, provides a framework for participating organizations to exchange patient data and also anticipates patients retrieving their own records.

After an overview of TEFCA and the history leading to it, we define 3 principles to promote patient autonomy and control as "rules of the road" for national-scale health information exchange (HIE).

Health Information Exchange: Past and Present

Query-based exchanges emerged in the 1990s. Successful examples, such as the Indianapolis Network for Patient Care and Research [1], led to community health information networks

XSL•FO
RenderX

[2], then to regional health information organizations, later renamed HIEs.

Query-based exchanges have primarily supported treatment, and most were deployed in limited contexts or vendor-defined boundaries [3]. Though myriad exchange organizations [4] have proved financially unsustainable [5], many today are sustainable, such as the government-supported Massachusetts Health Information Highway and the nonprofit Manifest Medex. Carequality, a membership-based nonprofit, underpins exchange among HIEs, electronic health record (EHR) vendors, and others. Epic Systems, leveraging extensive market share, enables its customers to participate in exchange. The Commonwell Health Alliance manages HIE for non-Epic members. Prior to TEFCA, common agreements for cross-organizational data exchange have been proposed within individual networks, by the Carequality Interoperability Framework [6] and the Markle Foundation [7].

Trusted Exchange Framework and Common Agreement

Entities, including HIEs, will apply to become Qualified Health Information Networks (QHINs), committing to standardize a technical framework and implement the Common Agreement.

The number of accessible patient records is anticipated to grow to over 200 million with nationwide reach [5]. TEFCA expands the purposes of use; information may be exchanged for treatment, operations, payment, public health, government benefits determination, and individual access.

Importantly, TEFCA expands the number of users authorized to query. Authorization is handled within a *hierarchical trust model*; a small number (~10) of QHINs will offer connectivity for on the order of 10,000 organizational participants including EHR vendors and health systems. These, in turn, authorize many users, likely on the order of millions of clinicians, insurance company staff, public health officials, and others under the current proposals. Responses to queries would be obligatory in the context of treatment or individual access, and permitted in other cases, except where prohibited by applicable laws [8]. Responsibilities for security enforcement are delegated. For example, a QHIN would trust a hospital to provision accounts and maintain credentials for its authorized users. In turn, the hospital would trust its users to comply with policies and laws.

General Challenges for TEFCA and Information Exchange

As is evident in public comments [9], TEFCA's goal of broad access brings challenges around privacy, security, and autonomy. As there is no consistent approach to verifying patient identity in health care, matching is probabilistic, based on demographics. Each query result might represent a "true positive" (correctly returning data), "true negative" (correctly returning no data), "false positive" (returning data from the wrong patient), or "false negative" (not returning existing data). Systems to verify user identity and match records across care sites can be expensive and raise privacy concerns because they sometimes aggregate large amounts of identifiable data, including biometrics.

Though queries are audited, any authorized user may look up data about any patient with an expectation of automated, immediate responses. One organization's security lapse or a user's compromised credentials could allow a malicious actor to find information about any patient, a risk that grows as networks expand.

Principles for Designing the Patient Role

Overview

Because sharing protected health information for treatment is exempt from HIPAA's (Health Insurance Portability and Accountability Act) requirement for patient-facing accounting of disclosures, patients have little visibility into when or with whom their health data are shared through HIE. A health care provider sharing a patient's information need not obtain consent. HIPAA does not compel a provider to heed patient requests to restrict sharing. Notifying patients during front-desk registration through signed "consent for treatment" yields patients no opportunity to negotiate. Patient concerns about privacy breaches and misuse of their information in exchanges have been well documented [10,11].

As implementers of TEFCA continue to address challenges through technical and business controls leading up to the network's launch, we propose 3 principles for meeting core objectives while recognizing and supporting patient autonomy and control, even if only a subset of patients are unsatisfied with default sharing permissions.

Principle 1. Patients Can Query for Data About Themselves

Today, patients face substantial challenges in assembling their records across sites of care [12,13], and as a result, uptake of individual access to patient records has been slow [14]. Exercising "individual access" using query-based exchange under TEFCA affords transparency about what records exist and allows patients to determine where their records are stored, identify errors [15], and correct missing information from failed matching. A single point of access to one's entire history of care may lead more individuals to seek out digital copies of their records [14].

To manage patient access, TEFCA anticipates that third-party "individual access providers" will verify a patient's identity, execute queries, and share results with the patient. We prefer a design affording individual access as a first-class feature of all QHINs rather than adding the technical, security, and organizational complexity of third-party coordination. This design would enable QHINs to absorb the costs of patient identity verification, rather than outsourcing them to a new category of businesses that must establish revenue streams to offset these costs.

XSL•FO
RenderX

Principle 2. Patients Can Know When Their Data Are Queried and Shared

The ability to see how one's own data are being queried can serve as a check that the system is working as intended or as a leading indicator that something has gone wrong. Patients are well positioned to notice unexpected queries or to detect the absence of an expected query. Under TEFCA's current policies, such details would be invisible to patients.

Principle 3. Patients Can Configure What Is Shared About Them

It is not yet established whether the widespread availability of data for care always improves outcomes, and there may be unintended consequences. Research is needed to see whether unfettered access to prior opinions and diagnoses improves care and whether restricted access introduces risks or degrades care. Given concerns about insurability, legal consequences, and stigmatization, patients may even avoid care to prevent widely accessible documentation. Additionally, national-scale data availability raises concerns about access and disclosure by political, journalistic, or adversarial actors. The option to configure what is shared may help establish a new patient-doctor relationship if a soured prior relationship is apparent in the chart. During a diagnostic odyssey, sharing less may reduce second-opinion clinicians from becoming prejudiced by previous specialists' assessments. Control over sharing may also help patients restrict queries about pregnancy-related care.

TEFCA currently does not provide control to patients. We recognize that permissive default settings that maximize access might satisfy a majority [16] and jumpstart network growth. We propose that patient concerns could be addressed pragmatically, starting with an all-or-nothing ability to opt out of exchange. More sophisticated controls could include (1) the ability to approve individual requests as queries are submitted and potentially (2) enabling access to a subset of encounters. For aspects of the record where a full picture is critical, purpose-built registries (eg, prescription drug–monitoring programs) provide accurate information irrespective of TEFCA.

If TEFCA-based exchange proves to become a data source for research and public health, patient autonomy to opt out of sharing may need to be balanced with requirements for unbiased data sets [17].

Conclusions

Launching query exchange capability on a national scale is a vast and worthy undertaking. While details are in flux and there is a TEFCA roadmap for future improvements, we believe these principles enforcing patient rights to autonomy and control should be addressed in policy and technology from the initial TEFCA launch. This will increase the likelihood of programmatic success by preemptively addressing legitimate concerns by advocacy groups. Though HIE is widespread today, and generally without a well-defined and protected patient role, TEFCA could serve as a model to underpin a 21st-century, patient-centered health information economy.

Conflicts of Interest

JCM is employed by Microsoft Corporation. KDM reports that Boston Children's Hospital receives corporate philanthropic support for his laboratory from SMART Advisory Committee members, which include the American Medical Association, the BMJ Group, Eli Lilly and Company, Google Cloud, Hospital Corporation of America, Microsoft Corporation, Cambia Health Solutions, Humana, and Quest Diagnostics. KDM previously served as an advisor to Ciox. JPP reports no conflicts.

References

1. Overhage JM, Tierney WM, McDonald CJ. Design and implementation of the Indianapolis Network for Patient Care and Research. Bull Med Libr Assoc 1995 Jan;83(1):48-56 [FREE Full text] [Medline: 7703939]
2. Dullabh P, Moiduddin A, Nye C, Virost L. The evolution of the state health information exchange cooperative agreement program: state plans to enable robust HIE. NORC at the University of Chicago. 2011 Aug. URL: https://www.healthit.gov/sites/default/files/pdf/state-health-info-exchange-program-evolution.pdf [accessed 2022-05-30]
3. Holmgren AJ, Adler-Milstein J. Health information exchange in US hospitals: the current landscape and a path to improved information sharing. J Hosp Med 2017 Mar;12(3):193-198. [doi: 10.12788/jhm.2704] [Medline: 28272599]
4. Overhage LM, Covich-Bordenick J, Li X, Overhage JM. Factors influencing progress of health information exchange organizations in the United States. Stud Health Technol Inform 2017;245:1236. [Medline: 29295323]
5. Adler-Milstein J, Garg A, Zhao W, Patel V. A survey of health information exchange organizations in advance of a nationwide connectivity framework. Health Aff (Millwood) 2021 May;40(5):736-744. [doi: 10.1377/hlthaff.2020.01497] [Medline: 33939510]
6. Carequality interoperability framework. Carequality. 2018. URL: https://carequality.org/resources/ [accessed 2022-09-26]
7. Markle common framework. Markle. 2022. URL: https://www.markle.org/health/markle-common-framework/connecting-professionals/ [accessed 2022-09-26]
8. Standard operating procedure (SOP): exchange purposes. The Sequoia Project. 2022. URL: https://rce.sequoiaproject.org/wp-content/uploads/2022/06/SOP-Exchange-Purposes.pdf [accessed 2022-10-21]
9. TEFCA comments summary: public comments on the Trusted Exchange Framework and Common Agreement draft 2. HealthIT.gov. URL: https://www.healthit.gov/tefcacomments [accessed 2022-08-07]

XSL•FO
RenderX

10. Simon SR, Evans JS, Benjamin A, Delano D, Bates DW. Patients' attitudes toward electronic health information exchange: qualitative study. J Med Internet Res 2009 Aug 06;11(3):e30 [FREE Full text] [doi: 10.2196/jmir.1164] [Medline: 19674960]
11. Wen K, Kreps G, Zhu F, Miller S. Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey. J Med Internet Res 2010 Dec 18;12(4):e73 [FREE Full text] [doi: 10.2196/jmir.1668] [Medline: 21169163]
12. Mandl KD, Szolovits P, Kohane IS. Public standards and patients' control: how to keep electronic medical records accessible but private. BMJ 2001 Feb 03;322(7281):283-287 [FREE Full text] [doi: 10.1136/bmj.322.7281.283] [Medline: 11157533]
13. Mandl KD, Kohane IS. Time for a patient-driven health information economy? N Engl J Med 2016 Jan 21;374(3):205-208. [doi: 10.1056/nejmp1512142]
14. Adler-Milstein J, Longhurst C. Assessment of patient use of a new approach to access health record data among 12 US health systems. JAMA Netw Open 2019 Aug 02;2(8):e199544 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9544] [Medline: 31411708]
15. Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. JAMA Netw Open 2020 Jun 01;3(6):e205867 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5867] [Medline: 32515797]
16. Cimino JJ, Frisse ME, Halamka J, Sweeney L, Yasnoff W. Consumer-mediated health information exchanges: the 2012 ACMI debate. J Biomed Inform 2014 Apr;48:5-15 [FREE Full text] [doi: 10.1016/j.jbi.2014.02.009] [Medline: 24561078]
17. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. NPJ Digit Med 2021 Jan 04;4(1):2 [FREE Full text] [doi: 10.1038/s41746-020-00362-8] [Medline: 33398052]

Abbreviations

EHR: electronic health record
HIE: health information exchange
HIPAA: Health Insurance Portability and Accountability Act
QHIN: Qualified Health Information Network
TEFCA: Trusted Exchange Framework and Common Agreement

Edited by G Eysenbach; submitted 15.08.22; peer-reviewed by A Malec, I Mircheva, K Taira; comments to author 05.09.22; revised version received 26.09.22; accepted 07.10.22; published 04.11.22.

Please cite as:
Mandel JC, Pollak JP, Mandl KD
The Patient Role in a Federal National-Scale Health Information Exchange
J Med Internet Res 2022;24(11):e41750
URL: https://www.jmir.org/2022/11/e41750
doi:10.2196/41750
PMID:36331535

©Joshua C Mandel, J P Pollak, Kenneth D Mandl. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Omnichannel Communication to Boost Patient Engagement and Behavioral Change With Digital Health Interventions

Agata Blasiak[1,2,3,4*], PhD; Yoann Sapanel[1], MS; Dana Leitman[5], MBA; Wei Ying Ng[1,2], BSocSci; Raffaele De Nicola[5], MBA, PhD; V Vien Lee[1,2*], PhD; Atanas Todorov[5*], MD, PhD; Dean Ho[1,2,3,4,6*], PhD

[1]The Institute for Digital Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]The N.1 Institute for Health, National University of Singapore, Singapore, Singapore

[3]Department of Biomedical Engineering, College of Design and Engineering, National University of Singapore, Singapore, Singapore

[4]Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[5]Arcondis Pte Ltd, Singapore, Singapore

[6]The Bia-Echo Asia Centre for Reproductive Longevity and Equality, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Agata Blasiak, PhD
The Institute for Digital Medicine
Yong Loo Lin School of Medicine
National University of Singapore
28 Medical Drive
#05-COR Centre for Life Sciences
Singapore, 117456
Singapore
Phone: 65 6601 7766
Email: agata.blasiak@nus.edu.sg

Abstract

Digital health interventions are being increasingly incorporated into health care workflows to improve the efficiency of patient care. In turn, sustained patient engagement with digital health interventions can maximize their benefits toward health care outcomes. In this viewpoint, we outline a dynamic patient engagement by using various communication channels and the potential use of omnichannel engagement to integrate these channels. We conceptualize a novel patient care journey where multiple web-based and offline communication channels are integrated through a "digital twin." The principles of implementing omnichannel engagement for digital health interventions and digital twins are also broadly covered. Omnichannel engagement in digital health interventions implies a flexibility for personalization, which can enhance and sustain patient engagement with digital health interventions, and ultimately, patient quality of care and outcomes. We believe that the novel concept of omnichannel engagement in health care can be greatly beneficial to patients and the system once it is successfully realized to its full potential.

(J Med Internet Res 2022;24(11):e41463) doi:10.2196/41463

KEYWORDS

digital health intervention; omnichannel engagement; behavioral change; communication channels; personalized engagement; health care; patient care; health care outcome; patient engagement; digital twin; DHI; digital health; eHealth; framework; development

Introduction

Digital health is an emerging field that involves the use of information technology for a range of health-related applications—from general wellness to medical devices [1]. A wide range of digital health technologies across the spectrum of mobile health, wearable devices, telehealth, and personalized medicine solutions is being explored. Digital health interventions (DHIs) are digital tools that help modify an individual's health behavior through direct interaction. These approaches can potentially improve the quality, accessibility, and affordability of health care worldwide [2]. Moreover, due to the availability and convenience of technology access and its immense potential for development, DHIs successfully implemented in a clinical

XSL•FO
RenderX

practice can have the capability of driving patients' health behavioral change in response to real-time tracking of intervention progress.

The COVID-19 pandemic has accelerated the adoption of digital health technologies on a global scale, with large programs already deployed in clinical practice [3,4]. With rapid digitalization of health care coupled with recent efforts to reimburse DHIs (eg, Digitale Gesundheitsanwendungen, also translated as digital health applications in Germany, the mHealthBELGIUM platform in Belgium, National Health Service Scotland) and the emergence of new digital interaction modalities, for example, Metaverse [5], as potential interfaces for DHIs in the future [6], guidance on how to best improve an individual's health with DHIs is required, including effective patient engagement strategies.

In this viewpoint study, we aimed to identify, in broad strokes, how DHIs could be potentially integrated into future health care. We focus on patient engagement as the critical component to effectively change the behavior of the patient via DHIs [7]. As summarized in a review by Barello et al [8] on eHealth, patient engagement encompasses adherence to treatments and medications and thus directly impacts all relevant outcomes of health interventions [9]. Among others, treatment adherence has been the most recently discussed as a key explanation for the observed reduction in cardiovascular mortality in the polypill trials [10]. By targeting patient engagement, it is possible to reshape behavioral intentions, motivations, and attitudes, as well as provide direct and continuous feedback on the effect of the changed behavior, which can further reinforce and sustain appropriate change [11,12]. Despite its significant role in digital health tool efficiency, patient engagement in the context of "last-foot delivery" of DHIs has not yet been explored to its full potential.

Communication Channels for Engagement in DHIs and Behavioral Change

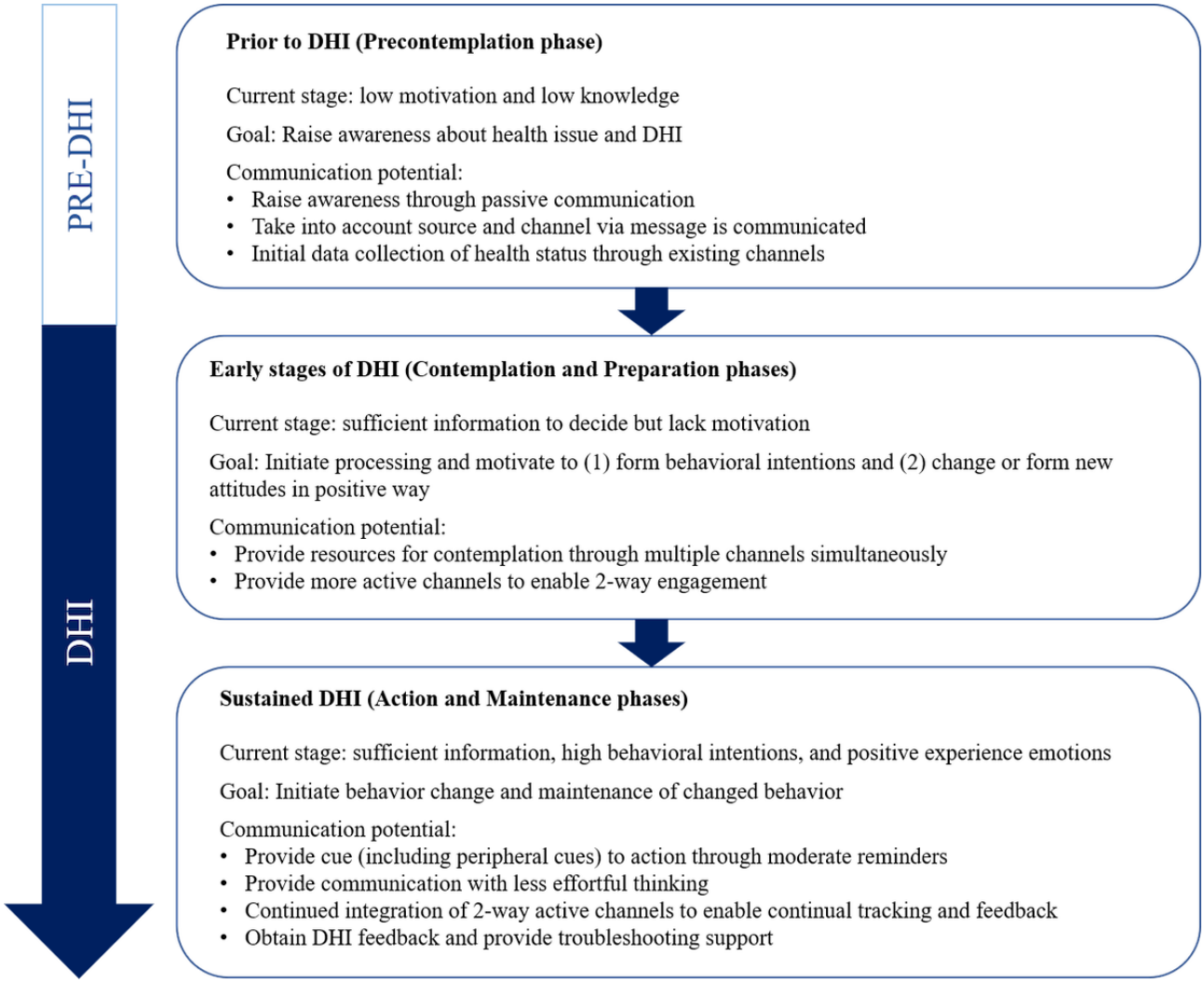
Patient engagement is highly interlinked with communication [13] and is a natural extension of the standard clinical patient education [14]. During a DHI, communication represents a dynamic process, which can begin passively (eg, hearing about the intervention through peers or doctors or advertisements), followed by more active interactions during initial and continued usage [15]. According to an integrated conceptual framework for health marketing communications [12], the communication potential and goals vary through different stages of patient interaction with the DHI (Figure 1). Given the variety and adaptability of communication channels [16], deploying the right communication channels and their dynamic adaptation throughout the intervention period can be leveraged as an effective means to build and sustain patient engagement [12]. This consideration precedes the introduction of other epiphenomena of communication, such as gamification [17].

First, patients need to form the motivation for using DHIs by means of prior information and involvement [12]. As such, it is crucial to raise awareness about the health issue and DHIs through physical channels such as health care providers and traditional mass media and virtual channels such as social media. As each patient's awareness about DHIs is increased, communication channels can evolve to become more personalized and engage patients more actively. During this stage, promoting the motivation to change and forming behavioral intentions and new positive attitudes are essential [12]. Increasing autonomous motivation has been shown to lead to positive health behavioral changes [18], and simultaneous use of multiple communication channels has the potential to employ a variety of motivation and behavior change techniques [19]. Eventually, as patients progress through the DHI process with high motivation and engagement, the goals of communication channels should shift toward initiating and maintaining changed behavior through less effortful thinking. At this stage, continued cue to action through moderate reminders and the use of peripheral cues such as recommendations from doctors and discussion with other patients can be beneficial [12]. Different communication channels can serve different roles and have varying effectiveness on engagement (see Dahl et al [20] for the list of omnichannel touchpoints and perceived effectiveness). Table 1 explores the potential application of communication channels and probable patient adoption considerations as inferred from existing literature on implementing these channels into DHIs. Of note, given the fast-paced nature of technology development, inventories of communication channels will continue to develop as new channels are established and older channels are phased out. With a diverse list of communication possibilities, patient engagement will remain dynamic as informed by the users' changing communication preferences, the stage of their DHI journey, and the information collected through the longitudinal intervention monitoring.

For optimal personalization and seamless integration, the DHI needs to adapt to the user communication preferences, which vary with demographics, psychographics, and the specific health condition [35,36]. Broadly, young adults prefer solutions that blend in with their current usage—accessible through existing hardware and software, have a familiar syntax (eg, use of emoticons), and are aligned with their existing habits [37]. Older adults, often with limited digital skills, also prefer familiar or easy-to-use communication channels for DHIs [35,36]. Despite the stated importance of a seamless and personalized experience in health care across physical and virtual channels [38], many of the current DHIs are unable to integrate seamlessly with each other and the rest of the health care system to personalize and improve engagement. The lack of integration of DHIs in the patient journey can exacerbate the existing health care fragmentation, which is particularly evident in the gaps in the continuity of care between the hospital and outpatient treatments. The effect of this discontinuity in user experience, among others, shows in the attrition rates with digital health apps. In one of many examples, the usage analysis of 57 mental health apps showed that the percentage of users who opened the app dropped from 69.4% to 3.9% within the first 15 days [39], potentially rendering them ineffective. This rapid disengagement is a particularly detrimental issue for DHIs that target chronic diseases, where continuity of care is crucial.

XSL•FO
RenderX

Figure 1. Communication potential of digital health interventions: the potential role of communication toward behavioral change before and during the digital health intervention, as guided by the framework proposed by Manika and Gregory-Smith [12]. DHI: digital health intervention.



XSL•FO
RenderX

Table 1. Examples of possible communication channels, their potential applications, and some of the patient adoption considerations. Any of such lists should be dynamically updated to match the evolving changes in the individual's digital presence and preferences.

Patient/caregiver interface	Examples of potential applications	Patient adoption considerations	
		Preintervention	Intervention and postintervention
In-person visit	Consultation with a physician, a rehabilitation session in a center, monitoring and drug titration with specialized nurses	Synchronous communication, requires initiative and time on the patient's part	Active communication channel
Email	Email consultations, communication of test results, e-referral	Asynchronous communication, both convenient and nondisruptive in the day-to-day lives of patients	Passive communication channel
Video consultation	Teleconsultation, e-prescription, e-referral	Synchronous communication, initial adoption can be affected by ease of use and perceived usefulness [21]	Active communication channel can be used to replace in-person visits for higher engagement but patient satisfaction is essential for continued usage [22]. Common issues include problems with internet connectivity and accessibility [23].
Mobile apps	Wellness apps, screening and monitoring solutions, or digital therapeutic interventions, with virtual reality–based therapy, personalized nudges, and real-world evidence collection	Initial adoption can be affected by ease of usage and user interface. Involvement of health care providers and public health authorities can benefit adoption [24].	Ease of use in long term enables disease monitoring and symptom tracking. Usage can be sustained through generating electronic satisfaction [25].
Digital social environment	Social media and metaverse that can be used for educating, supporting, and influencing behavior change as well as individualized lifestyle improvement	Adoption might be affected by digital divide and privacy concerns.	Enables patients' engagement with the ability to share, comment on, and react but also social norms intervention, especially for the young adults [26]
Web-based platforms	Web-based learning/educational platforms such as internet-based diabetes management platform or digital patient-reported outcome measure and electronic patient-reported outcome platforms	Adoption determinants include reliability of provider, level of awareness toward web-based platform, and accessibility [27].	Enables self-management. Continued engagement relies on confirmation of user's initial expectations and perceived usefulness [28].
Messaging solutions	SMS-based communication, chatbot, virtual health assistant, messaging apps for just-in-time interventions, triaging	Familiar platform and increased ease of usage to encourage adoption. Limited with regards to nature and amount of information shared. Narrative persuasion through first-person point of view can impact persuasion of health messages [29].	Enables reminders for chronic disease management. Limited in terms of clinical data sharing—quality of responsiveness and variability can help improve conversation quality [30].
Safe messaging apps	Direct messaging to health care providers, appointment booking, patient reminders, data sharing between health care providers, and patient data storage	Adoption is influenced by convenience and the integration of safe messaging apps to electronic health record systems [31]	Continued usage relies on patient's preference for enhanced security features and ease of use [32]
Voice-based assistant	Smart home and mobile-based digital assistant that enable proactive reminders, remote care access, and the ability to capture patients' responses to artificial intelligence–powered voice biomarkers could also be used to screen for and diagnose a wide range of diseases as well as triaging.	The relative simplicity lowers traditional technological barriers and provides the benefits of hands-free and eyes-free engagement mode. The disadvantages are that some patients may have difficulty in formulating a structured sentence for a command as well as may have privacy concerns associated with having a voice assistant "always-on" [33].	Potential for passive symptom tracking and just-in-time interventions
Wearables and sensors	Global positioning system tracking devices, Bluetooth beacon technology, body vitals sensors	Initial adoption can be affected by novelty of wearable technology—for example, design aesthetic is a prominent factor when influencing behavioral intention [34]	Enables passive symptom tracking and generates long-term data for feedback and analysis. Continued perceived usefulness and positive attitude need to be maintained [34].

XSL•FO
RenderX

Omnichannel Engagement for DHI-Led Behavioral Change

Encouraged by its success in retail [40], we identified omnichannel engagement (OCE) as a highly promising and untapped approach for DHI patient engagement. OCE is a strategy to integrate and interconnect multiple communication channels in a synchronized operating model, which leverages data and digital tools to deliver a seamless, consistent, and personalized experience for the user [41]. The key features of OCE are (1) integrated and interconnected communication using multiple available channels, (2) synchronized/coordinated activity, (3) personalized and consistent experience for the user, and (4) continuous improvement via data and digital tools.

The correct management of the following 6 enablers is crucial for achieving a successful OCE: people, platform, service and content, channels, legal framework, and data and measurement (Table 2). Several consumer technology companies have maximized their profits by executing long-term omnichannel strategies based on a thorough management of these enablers. The consumer technology omnichannel approach is therefore being considered the state-of-the-art and a reference for other industries [42]. Regarding the patient digital behavior, studies show that individuals who seek and are capable of utilizing eHealth information are also more open to use health services provided digitally [43]. Accordingly, health care companies and public sector utilize new ways of doing marketing, where medical science liaisons—services for patients and health care professionals—use increasingly more digital means besides the traditional in-person communication channel [44]. For example, social media posts on Twitter and Facebook have been used by the largest hospital chain in Turkey [45] as well as by the Portuguese national health service to improve health literacy on disease prevention and well-being together with other nondigital channels [46]. The electronic word-of-mouth–based strategy on social media for pharma marketing on over-the-counter medications has also been explored [47]. It is not surprising that the health care market has been targeted in recent years by the consumer technology companies familiar with OCE, which have introduced fitness apps, wearables, and activity trackers [48].

We stipulate that OCE can play a significant role as an integral element of DHIs and not only as an adjacent marketing stream. We envision a patient journey where DHIs are a crucial part of a stand-alone or auxiliary treatment and behavioral change and leverage OCE to boost its efficiency. In our vision, DHIs are distributed into several services (eg, dedicated disease service, tracking service), which are integrated with various communication channels as well as wearables/sensors and the digital presence of the patient. The information flow is bidirectional—the patient receives the digital intervention through OCE and provides information for the DHI operations. All available relevant data, for example, medical data, health data, connections, digital preferences, and data for health can be centrally collected and integrated to form a digital twin [49,50]. Digital twin is a digital representation of the individual whose data it is based on. It can offer insights about that individual, which, in turn, facilitate the design and monitoring of an intervention. The digital twin in our vision may not need to encompass all available data but merely may need access to multiple sources such as the personal health record to gain sufficient insights to be actionable. For example, to understand the phenotypic response and to recommend an optimal pharmacological dosage with sufficient safety and accuracy, only 3-6 data pairs may be necessary [51,52]. Artificial intelligence could prove crucial in managing and extracting relevant features from the multidimensional, noisy, and incomplete data as well as in following up with a variety of appropriate adaptations and actions [53,54]. It will be required to find the right balance of complexity and size of the data model versus the quality and rapid generation of actionable insights, according to the general premise of efficient use of artificial intelligence in health care [55].

OCE serves the purpose of linking the patient to a network of health services through various integrated communication channels revolving around the patient's health-seeking experience. In already existing examples, instant messaging, be it stand-alone (eg, WhatsApp) or platforms linked to social media such (eg, Facebook Messenger), can be used for triaging in-person and web-based consultations [56]. Wearables enable long-term symptom management, and emails and teleconsultations provide a seamless and convenient method for follow-ups and communication of clinical data. These services can be combined with the aligned nudges for lifestyle and nutritional choices delivered according to the patient's digital presence, for example, specific food suggestions when shopping online.

OCE also creates opportunities for obtaining additional insights. An analysis of the interconnection between communication channels may allow for analysis of the network of communication and support, which can be a powerful ally for behavioral change. In 1 example, obesity is known to spread through social ties through a diffusion of unhealthy behaviors [57], including physical inactivity [58]. Accordingly, social connections become informal channels of information with real impact on health decisions for self and others, as observed in parenting [59]. Indeed, rapid diffusion of information has been demonstrated within social networks [60]. Such insights could lead to broader understanding of one's behavior motivators and their environment to launch a precise DHI and form beneficial habits.

XSL•FO
RenderX

Table 2. Key enablers for effective omnichannel strategies.

Enablers	Definition
People	Clear roles and responsibilities around governance, omnichannel engagement expertise, skills, and know-how
Platforms	Foundational technology stack
Services and content	Engagement plans of defined customer-centric services, with modular and reusable content
Channels	Integrated operations across all customer-facing channels
Legal framework	Multimedia consent, cookie and privacy policies, medical and regulatory review policies
Data and measurement	Customer identifier, data management strategies for optimal analytics and insights

Discussion

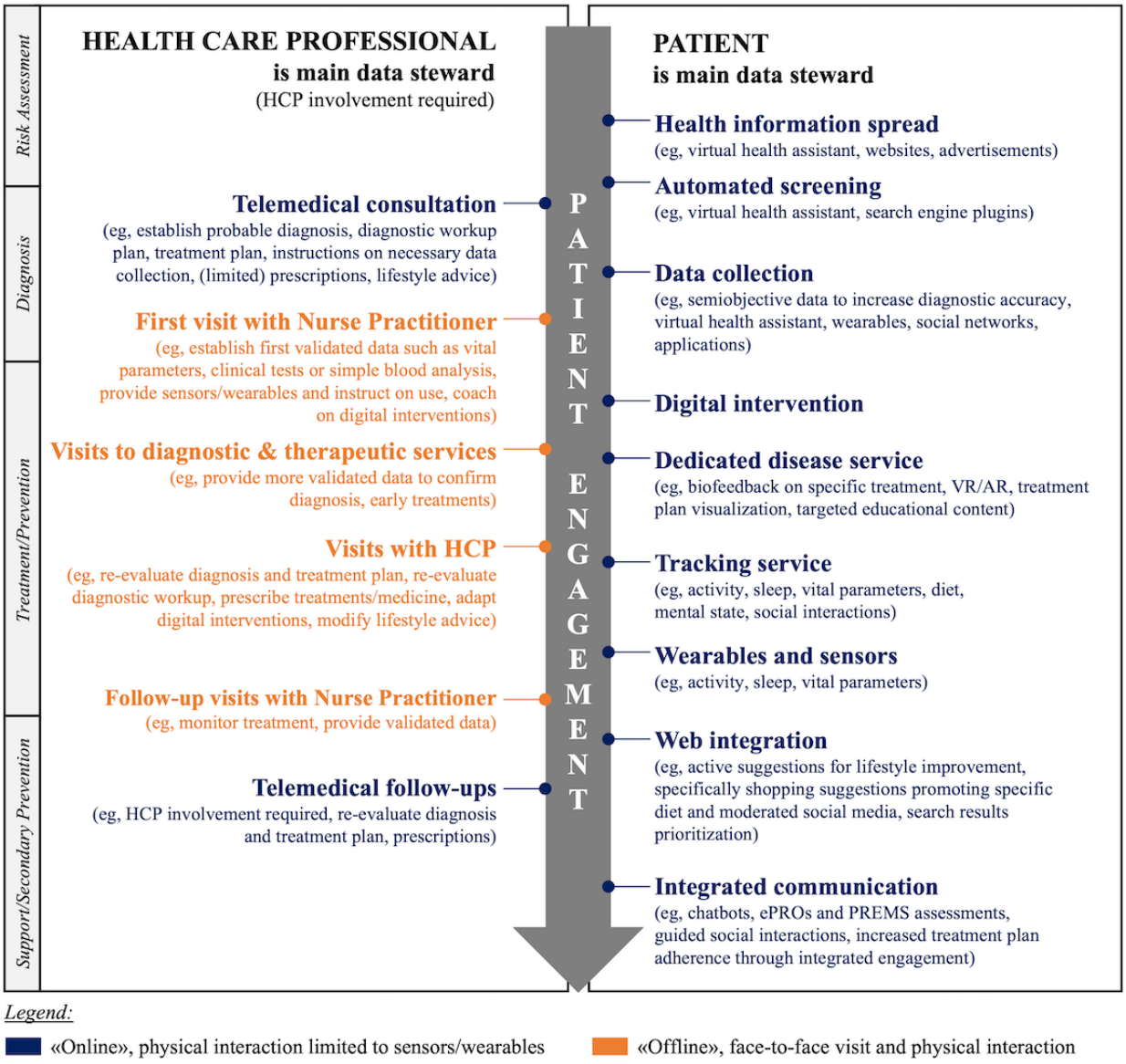
Supporting the Patient Journey With DHIs and OCE

In this viewpoint paper, we outline how DHIs supported by OCE may become an impactful health intervention. Below, we describe how a patient journey benefiting from such a tool could look like (Figure 2). At the beginning of the patient journey, individuals at risk of a disease receive broadly targeted medical information through passive communication, with which they can decide to follow up. Interested individuals are supported by an automatic screening procedure in deciding whether contact with a health care professional is necessary. If this is the case, the initial contact is a telemedicine consultation. This allows for a low-cost and rapid way of establishing an initial differential diagnosis, diagnostic workup, and treatment plan. The patient can be instructed on the necessary data to be collected, for example, symptom diaries, wearables, or other tracking platforms, to improve the positive predictive value of the diagnostic procedures. Once sufficient data have been collected, an in-person visit with a health care practitioner follows, during which more clinical data are gathered and the treatment plan is adapted accordingly. Ideally, necessary additional wearables and the correct instructions for new or already used ones are provided at this appointment, together with a recommendation/prescription of an appropriate DHI. Following this are 2 parallel tracks: while the DHI is applied to the patient to change behavior and monitor progress/treatment compliance, the patient also follows up on visits to diagnostic services (eg, imaging services), therapeutic services (eg, physical therapy, nutritional specialist), visits to the doctor's office, and further visits to the health care practitioner as necessary. We envision that, thanks to the adaptive engagement with DHIs through OCE, patient adherence to both web-based and offline components is high, leading to a rapid and sustained shift in behavior and ultimately better outcomes while reducing the most resource-intense engagements, that is, face-to-face visits. Once treatment is completed, DHIs continue to support the habituation of behavior and thus provide secondary prevention. In patients with chronic diseases, telemedical visits continue as necessary for treatment plan adaptation, while DHIs provide a much closer view of the disease state and offer an asynchronous way for the health care practitioner to stay up-to-date on the patient status.

To allow the flexibility necessary for this approach, it is crucial that health care professionals are agnostic about the effective engagement channel and that their primary systems used to document clinical data support the integration with the patient's digital twin. Moreover, health care professionals may be required to assess individuals who will most likely need assistance with the familiarization and use of DHIs [61]. Although patients are increasingly open to incorporating technology as part of their care delivery [62], continual efforts through personalization and understanding the resources and capabilities of targeted patient populations should be made to promote greater technology access and literacy. As digital health technologies continue to augment health care professionals' roles, new skills may become necessary and a new role of a digital health specialist may emerge as a primary health care provider in the future.

XSL•FO
RenderX

Figure 2. Patient journey with a digital health intervention through omnichannel engagement. The section on the right side of the arrow presents selected communication channels with direct interface to the patient, where patient is the main data steward. The section on the left side of the arrow provides selected communication channels, where the patient interacts with health care professionals who are the main data stewards. The engagement approach with patients is constantly adapted depending on the preferred patient channels and based on the approach that results in an effective behavioral change response. Health care professionals should therefore be channel-agnostic. The data are owned by the patient with stewardship granted to the main contributors (where appropriate). Data actively create a digital twin that enables actionable health predictions and sustained behavioral change. AR: augmented reality; ePRO: electronic patient-reported outcome; HCP: health care professional; PREM: patient-reported experience measure; VR: virtual reality.



Challenges to Implementing DHIs With OCE

The presented approach, incorporating a digital twin of the patient as well as OCE, needs to align with the collectively accepted principles and values such as “privacy, equity, fairness, patient safety, and patient autonomy over health-care decisions” [63]. Although DHIs may aim to pragmatically increase patient autonomy, the data they use and generate can be seen as a commercial asset, giving rise to risks such as the breach of data security or privacy as well as data governance issues. Adding to this complexity, DHIs operating in multiple countries will need to comply with local, national, regional, and international standards.

Data privacy need to be ensured while using available communication channels and tracking patient preferences and managing the intervention via the digital twin. DHIs have to comply with the prevailing regulations, for example, the general data protection regulation (GDPR) in the European Union as well as with further specifications of each of the European Union member states where a DHI is deployed [64]. The requirements pertain to the physical location of data, the way in which patient consent is gathered, and the necessity for consent before sharing data with specific third parties such as insurers. A specific example in the context of DHIs with OCE is the integration and use of WhatsApp [31]. WhatsApp is a popular messaging app with great potential as a seamless DHI communication channel; however, it fails to comply with GDPR regulations, as messages

XSL•FO
RenderX

are stored in servers outside the European Union. In fact, a vast majority of health apps fall outside the remit of the regulations in most nations [65]. A range of secure messaging apps with increased data privacy are being developed and marketed [66-70]. However, their compliance with GDPR varies greatly and questions remain on how they can ensure a seamless experience and sustained use to be potentially adapted for future integration with DHIs.

Patients should be empowered with increased information and awareness of their own health care journey as authorities establish the balance between data protection and data exchange and as new governance models arise. Patients should be given the opportunity to participate actively in their own health decisions and treatment by fully utilizing their health data rights. Among others, an open question remains on the mechanisms to guarantee patient-controlled access to their own data and broadly speaking to the digital twin. The digital twin represents an actionable collection of information designed specifically to be able to "connect the dots" and both predict, monitor, and influence a patient's behavior. Conceptually, new decentralized technologies such as blockchain may support the control over access and secure information flow in the digital twin [71,72]; however, those are not yet broadly adopted, and beyond technical challenges, a regulatory shift and user education are required to minimize the risk of the digital twin being used as a means to influence the autonomy of the health care decisions.

Some regulatory bodies have issued laws attempting to realize publicly controlled digital twin–like patient profiles (eg, the National Electronic Patient Record in Germany, the National Electronic Health Record in Singapore, the Electronic Patient Dossier in Switzerland). The key benefit of such data exchange systems is to overcome the challenges encountered with historical system interoperability issues and to enable data exchange between health care professionals. While demonstrating significant developments for integrated patient profiles, those efforts have not yet reached a comprehensive patient-specific, actionable digital twin, and the vision of the patient journey illustrated in Figure 2 is not yet facilitated.

Conclusions

In this viewpoint paper, we have outlined the principles and challenges of DHIs combined with patient engagement through OCE. The evidence stemming from trials and real-world data will verify the true value of such an approach. One strategy to rapidly pilot potential OCE for DHIs could be the combination of a wellness/lifestyle intervention with a health care intervention. This could provide initial evidence for increased patient engagement while testing the concept of continuous improvement with data generated through the engagement. Further integration, for example, with a personal health record could then be added in the second phase. However, we do not see this as a bypass for adhering to the appropriate health care and data privacy regulations. Recently, there have been worrisome reports of the breaches of data privacy and data security by wellness mobile apps [73-75]. In addition to data breaches, other concerns include the aggregation of multisource data and generating digital twins without appropriate access control for the individual represented by the digital twin.

The global strategy on digital health 2020-2025 states that digital health is a means of promoting equitable, affordable, and universal access to health for all, including the special needs of groups that are vulnerable in the context of digital health [76]. To truly fulfil that role, DHIs need proof of effectiveness, accessibility, feasibility, sustainable resource use, and the observation of equity and rights [2]. The extent of success of DHIs in changing health outcomes depends on the level of patient engagement during the intervention and subsequent sustained change of the intended health behavior [77]. Capitalizing on the adaptability of OCE, there is potential for personalizing the user experience to each patient and the flexibility to adapt to the needs and regulations of different geographical regions. Unanswered questions remain with regards to the emergence of new communication channels and behaviors in a digital space as well as adoption of health care models and health care professionals' roles. In addition, careful management of the commercial versus health care interests in the integration of the communication channels and the digital twin is required. However, we believe that DHIs focused on patient engagement through a dynamic adaptable OCE approach and built on a well-formed compliant digital twin have the potential to disrupt unsustainable health care systems and greatly improve the health of many individuals.

Acknowledgments

DH gratefully acknowledges the funding of National Research Foundation Singapore under its Artificial Intelligence Singapore Program (award AISG-GC-2019-002); the Singapore Ministry of Health's National Medical Research Council under its Open Fund Large Collaborative Grant (grant MOH-OFLCG18May-0028); Institute for Digital Medicine Translational Research Program (grant R-719-000-037-733) at the Yong Loo Lin School of Medicine, National University of Singapore; Ministry of Education Tier 1 Faculty Research Committee Grant (grant R-397-000-333-114), Micron Foundation; and Sun Life Singapore.

Authors' Contributions

AB, DL, AT, and DH conceived the investigation. AB, YS, DL, VVL, AT, and DH outlined the investigation. AB, YS, DL, WYN, RDN, VVL, AT, and DH contributed to scientific discussion, critical review, drafting, and editing of the manuscript and approved the final manuscript. AB, VVL, AT, and DH supervised the investigation.

XSL•FO
RenderX

Conflicts of Interest

DL, RDN, and AT are employees of Arcondis (AG and Pte Ltd). AB and DH are coinventors of previously filed pending patents on artificial intelligence–based therapy development. DH is a cofounder and shareholder of KYAN Therapeutics, which has licensed intellectual property pertaining to artificial intelligence–based oncology drug development. The other authors declare no competing interests.

References

1. What is digital health? 2020. US Food and Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2022-05-12]
2. Recommendations on digital interventions for health system strengthening: evidence and recommendations. World Health Organization. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2022-05-12]
3. Horgan D, Hackett J, Westphalen C, Kalra D, Richer E, Romao M, et al. Digitalisation and COVID-19: the perfect storm. Biomed Hub 2020 Sep 17;5(3):1341-1363 [FREE Full text] [doi: 10.1159/000511232] [Medline: 33564668]
4. Patel AM, Bhardwaj A, Basch E, Hudson KE, Escudier SM, Books H, et al. Evaluating mass implementation of digital health solutions to improve quality and reduce disparities in a large multisite community oncology practice. JCO 2022 Jun 01;40(16_suppl):1507-1507. [doi: 10.1200/jco.2022.40.16_suppl.1507]
5. Riva G, Wiederhold B. What the Metaverse is (really) and why we need to know about it. Cyberpsychol Behav Soc Netw 2022 Jun 01;25(6):355-359. [doi: 10.31234/osf.io/dvcmw]
6. Plechatá A, Makransky G, Böhm R. Can extended reality in the metaverse revolutionise health communication? NPJ Digit Med 2022 Sep 02;5(1):132 [FREE Full text] [doi: 10.1038/s41746-022-00682-x] [Medline: 36056245]
7. Sucala M, Ezeanochie N, Cole-Lewis H, Turgiss J. An iterative, interdisciplinary, collaborative framework for developing and evaluating digital behavior change interventions. Transl Behav Med 2020 Dec 31;10(6):1538-1548 [FREE Full text] [doi: 10.1093/tbm/ibz109] [Medline: 31328775]
8. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for Patient Engagement: A Systematic Review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
9. Shortell SM, Poon BY, Ramsay PP, Rodriguez HP, Ivey SL, Huber T, et al. A Multilevel Analysis of Patient Engagement and Patient-Reported Outcomes in Primary Care Practices of Accountable Care Organizations. J Gen Intern Med 2017 Jun;32(6):640-647 [FREE Full text] [doi: 10.1007/s11606-016-3980-z] [Medline: 28160187]
10. Rao S, Jamal Siddiqi T, Khan MS, Michos ED, Navar AM, Wang TJ, et al. Association of polypill therapy with cardiovascular outcomes, mortality, and adherence: A systematic review and meta-analysis of randomized controlled trials. Prog Cardiovasc Dis 2022;73:48-55. [doi: 10.1016/j.pcad.2022.01.005] [Medline: 35114251]
11. Ogden J, Karim L, Choudry A, Brown K. Understanding successful behaviour change: the role of intentions, attitudes to the target and motivations and the example of diet. Health Educ Res 2007 Jun;22(3):397-405. [doi: 10.1093/her/cyl090] [Medline: 16971672]
12. Manika D, Gregory-Smith D. Health marketing communications: An integrated conceptual framework of key determinants of health behaviour across the stages of change. Journal of Marketing Communications 2014 Aug 30;23(1):22-72. [doi: 10.1080/13527266.2014.946436]
13. Gallivan J, Kovacs BK, Bellows M, Eigenseher C. The many faces of patient engagement. J Participat Med. 2012. URL: https://participatorymedicine.org/journal/evidence/research/2012/12/26/the-many-faces-of-patient-engagement/ [accessed 2022-11-04]
14. Gruman J, Rovner MH, French ME, Jeffress D, Sofaer S, Shaller D, et al. From patient education to patient engagement: implications for the field of patient education. Patient Educ Couns 2010 Mar;78(3):350-356. [doi: 10.1016/j.pec.2010.02.002] [Medline: 20202780]
15. 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]
16. Weaver B, Lindsay B, Gitelman B. Communication Technology and Social Media: Opportunities and Implications for Healthcare Systems. Online J Issues Nurs 2012 Sep 30;17(3):3. [doi: 10.3912/ojin.vol17no03man03]
17. Seaborn K, Fels DI. Gamification in theory and action: A survey. International Journal of Human-Computer Studies 2015 Feb;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
18. Sheeran P, Wright CE, Avishai A, Villegas ME, Rothman AJ, Klein WMP. Does increasing autonomous motivation or perceived competence lead to health behavior change? A meta-analysis. Health Psychol 2021 Oct;40(10):706-716. [doi: 10.1037/hea0001111] [Medline: 34881939]
19. Teixeira PJ, Marques MM, Silva MN, Brunet J, Duda JL, Haerens L, et al. A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts. Motivation Science 2020 Dec;6(4):438-455. [doi: 10.1037/mot0000172]
20. Dahl AJ, D'Alessandro AM, Peltier JW, Swan EL. Differential effects of omni-channel touchpoints and digital behaviors on digital natives' social cause engagement. JRIM 2018 May 31;12(3):258-273. [doi: 10.1108/jrim-04-2018-0051]

XSL•FO
RenderX

21. Anderson JT, Bouchacourt LM, Sussman KL, Bright LF, Wilcox GB. Telehealth adoption during the COVID-19 pandemic: A social media textual and network analysis. Digit Health 2022;8:20552076221090041 [FREE Full text] [doi: 10.1177/20552076221090041] [Medline: 35392254]
22. Drerup B, Espenschied J, Wiedemer J, Hamilton L. Reduced No-Show Rates and Sustained Patient Satisfaction of Telehealth During the COVID-19 Pandemic. Telemed J E Health 2021 Dec;27(12):1409-1415. [doi: 10.1089/tmj.2021.0002] [Medline: 33661708]
23. Hoi KK, Brazina SA, Kolar-Anderson R, Zopf DA, Bohm LA. A Longitudinal Comparison of Telemedicine Versus In-Person Otolaryngology Clinic Efficiency and Patient Satisfaction During COVID-19. Ann Otol Rhinol Laryngol 2021 Dec 03:34894211055349. [doi: 10.1177/00034894211055349] [Medline: 34861781]
24. Aamir J, Ali SM, Kamel Boulos MN, Anjum N, Ishaq M. Enablers and inhibitors: A review of the situation regarding mHealth adoption in low- and middle-income countries. Health Policy and Technology 2018 Mar;7(1):88-97. [doi: 10.1016/j.hlpt.2017.11.005]
25. Wu P, Zhang R, Zhu X, Liu M. Factors Influencing Continued Usage Behavior on Mobile Health Applications. Healthcare (Basel) 2022 Jan 21;10(2):2 [FREE Full text] [doi: 10.3390/healthcare10020208] [Medline: 35206823]
26. Friedman VJ, Wright CJC, Molenaar A, McCaffrey T, Brennan L, Lim MSC. The Use of Social Media as a Persuasive Platform to Facilitate Nutrition and Health Behavior Change in Young Adults: Web-Based Conversation Study. J Med Internet Res 2022 May 18;24(5):e28063 [FREE Full text] [doi: 10.2196/28063] [Medline: 35583920]
27. Ramtohul I. The adoption of e-health services: Comprehensive analysis of the adoption setting from the user s perspective. Health Policy and Technology 2015 Sep;4(3):286-293. [doi: 10.1016/j.hlpt.2015.04.007]
28. Leung L, Chen C. E-health/m-health adoption and lifestyle improvements: Exploring the roles of technology readiness, the expectation-confirmation model, and health-related information activities. Telecommunications Policy 2019 Jul;43(6):563-575. [doi: 10.1016/j.telpol.2019.01.005]
29. Chen M, Bell RA. A meta-analysis of the impact of point of view on narrative processing and persuasion in health messaging. Psychol Health 2022 May;37(5):545-562. [doi: 10.1080/08870446.2021.1894331] [Medline: 33678078]
30. 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]
31. Masoni M, Guelfi MR. WhatsApp and other messaging apps in medicine: opportunities and risks. Intern Emerg Med 2020 Mar;15(2):171-173. [doi: 10.1007/s11739-020-02292-5] [Medline: 32062746]
32. Fonner VA, Kennedy S, Desai R, Eichberg C, Martin L, Meissner EG. Patient-Provider Text Messaging and Video Calling Among Case-Managed Patients Living With HIV: Formative Acceptability and Feasibility Study. JMIR Form Res 2021 May 27;5(5):e22513 [FREE Full text] [doi: 10.2196/22513] [Medline: 34042596]
33. Kim S. Exploring How Older Adults Use a Smart Speaker-Based Voice Assistant in Their First Interactions: Qualitative Study. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e20427 [FREE Full text] [doi: 10.2196/20427] [Medline: 33439130]
34. Niknejad N, Ismail WB, Mardani A, Liao H, Ghani I. A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges. Engineering Applications of Artificial Intelligence 2020 Apr;90:103529. [doi: 10.1016/j.engappai.2020.103529]
35. Yeshua-Katz D, Shapira S, Aharonson-Daniel L, Clarfield AM, Sarid O. Matching Digital Intervention Affordances with Tasks: The Case of a Zoom and WhatsApp Mental Health Intervention for Seniors during the COVID-19 Pandemic. Health Commun 2021 Jul 30:1-13. [doi: 10.1080/10410236.2021.1956071] [Medline: 34325581]
36. Jiang J, Zhu Q, Zheng Y, Zhu Y, Li Y, Huo Y. Perceptions and Acceptance of mHealth in Patients With Cardiovascular Diseases: A Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e10117 [FREE Full text] [doi: 10.2196/10117] [Medline: 30714942]
37. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End User and Implementer Experiences of mHealth Technologies for Noncommunicable Chronic Disease Management in Young Adults: Systematic Review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
38. Mahoney K, Wilson-Maher A. How Penn medicine is advancing care delivery to meet more patients' needs. Front Health Serv Manage 2022 Apr;38(3):4-9. [doi: 10.1097/hap.0000000000000135]
39. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
40. Simone A, Sabbadin E. The New Paradigm of the Omnichannel Retailing: Key Drivers, New Challenges and Potential Outcomes Resulting from the Adoption of an Omnichannel Approach. IJBM 2017 Dec 18;13(1):85. [doi: 10.5539/ijbm.v13n1p85]
41. Vasconcelos A, Casaca F. Systematic selection and prioritization of communication channels in the healthcare sector. 2021 Presented at: International Conference on Enterprise Information Systems; April 26-28; Virtual p. 668-676. [doi: 10.5220/0010441706680676]
42. Yrjölä M, Saarijärvi H, Nummela H. The value propositions of multi-, cross-, and omni-channel retailing. IJRDM 2018 Dec 10;46(11/12):1133-1152. [doi: 10.1108/ijrdm-08-2017-0167]

XSL•FO
RenderX

43. Özen H, Yaşin B. A pathway to sustainable health during COVID-19 pandemic: digital health services. In: Çalıyurt KT, editor. New Approaches to CSR, Sustainability, Accountability, Volume II. Singapore: Springer; 2021:55-76.
44. Bishop M. Healthcare social media for consumer informatics. In: Indmunds M, Hass C, Holve E, editors. Consumer Informatics and Digital Health, Solutions for Health and Health Care. Cham: Springer; 2019:61-86.
45. Çağlıyor S, Tosun P, Uray N. Communicating value in healthcare marketing from a social media perspective. In: Topcu YI, Önsel Ekici Ş, Kabak Ö, Aktas E, Özaydın Ö, editors. New Perspectives in Operations Research and Management Science. Cham: Springer; 2022:134-170.
46. Goiana-da-Silva F, Cruz-e-Silva D, Morais NA, Carriço M, Costa F, Miraldo M. Disrupting the landscape: how the Portuguese National Health Service built an omnichannel communication platform. WHO Regional Publications. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/327063/php-5-2-3-314-323-eng.pdf?sequence=1&isAllowed=y [accessed 2022-11-02]
47. Proença M. The application of social media for marketing strategies in pharma healthcare. Iscte - University Institute of Lisbon. 2021. URL: https://repositorio.iscte-iul.pt/bitstream/10071/25816/1/phd_maria_dias_proenca.pdf [accessed 2022-11-02]
48. Bongaerts R, Henn H, Khare A. The Pac-Man principle in the healthcare market. In: Khare A, Stewart B, Schatz R, editors. Phantom Ex Machina. Cham: Springer; 2017:175-184.
49. Schwartz SM, Wildenhaus K, Bucher A, Byrd B. Digital Twins and the Emerging Science of Self: Implications for Digital Health Experience Design and "Small" Data. Front Comput Sci 2020 Oct 15;2:2. [doi: 10.3389/fcomp.2020.00031]
50. Venkatesh KP, Raza MM, Kvedar JC. Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation. NPJ Digit Med 2022 Sep 22;5(1):150 [FREE Full text] [doi: 10.1038/s41746-022-00694-7] [Medline: 36138125]
51. Blasiak A, Khong J, Kee T. CURATE.AI: Optimizing Personalized Medicine with Artificial Intelligence. SLAS Technol 2020 Apr;25(2):95-105. [doi: 10.1177/2472630319890316] [Medline: 31771394]
52. Blasiak A, Truong A, Tan WJL, Kumar KS, Tan SB, Teo CB, et al. PRECISE CURATE.AI: A prospective feasibility trial to dynamically modulate personalized chemotherapy dose with artificial intelligence. JCO 2022 Jun 01;40(16_suppl):1574-1574. [doi: 10.1200/jco.2022.40.16_suppl.1574]
53. Impedovo D, Pirlo G. eHealth and Artificial Intelligence. Information 2019 Mar 19;10(3):117. [doi: 10.3390/info10030117]
54. Massaro A, Maritati V, Savino N, Galiano A, Convertini D, De Fonte E, et al. A Study of a Health Resources Management Platform Integrating Neural Networks and DSS Telemedicine for Homecare Assistance. Information 2018 Jul 19;9(7):176. [doi: 10.3390/info9070176]
55. Volovici V, Syn NL, Ercole A, Zhao JJ, Liu N. Steps to avoid overuse and misuse of machine learning in clinical research. Nat Med 2022 Oct;28(10):1996-1999. [doi: 10.1038/s41591-022-01961-6] [Medline: 36097217]
56. Thomas K. Wanted: a WhatsApp alternative for clinicians. BMJ 2018 Feb 12;360:k622. [doi: 10.1136/bmj.k622] [Medline: 29440047]
57. Christakis NA, Fowler JH. The Spread of Obesity in a Large Social Network over 32 Years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/nejmsa066082]
58. Huang H, Yan Z, Chen Y, Liu F. A social contagious model of the obesity epidemic. Sci Rep 2016 Nov 28;6:37961 [FREE Full text] [doi: 10.1038/srep37961] [Medline: 27892501]
59. Cornwell B, Yan X, Carlin RF, Fu L, Wang J, Moon RY. Social network influences on new mothers' infant sleep adjustments. Soc Sci Med 2021 Jan;269:113585 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113585] [Medline: 33333376]
60. D'Agostino G, D'Antonio F, De Nicola A, Tucci S. Interests diffusion in social networks. Physica A: Statistical Mechanics and its Applications 2015 Oct;436:443-461. [doi: 10.1016/j.physa.2015.05.062]
61. Hughes H, Canino R, Sisson S, Hasselfeld B. A simple way to identify patients who need tech support for telemedicine. Harvard Business Review. 2021 Aug 10. URL: https://hbr.org/2021/08/a-simple-way-to-identify-patients-who-need-tech-support-for-telemedicine [accessed 2022-09-10]
62. Swan EL, Dahl AJ, Peltier JW. Health-care marketing in an omni-channel environment. JRIM 2019 Nov 20;13(4):602-618. [doi: 10.1108/jrim-03-2019-0039]
63. Kickbusch I, Piselli D, Agrawal A, Balicer R, Banner O, Adelhardt M, et al. The Lancet and Financial Times Commission on governing health futures 2030: growing up in a digital world. The Lancet 2021 Nov;398(10312):1727-1776. [doi: 10.1016/s0140-6736(21)01824-9]
64. Wolford B. What is GDPR, the EU's new data protection law? GDPR EU. URL: https://gdpr.eu/what-is-gdpr/ [accessed 2022-06-22]
65. Magrabi F, Habli I, Sujan M, Wong D, Thimbleby H, Baker M, et al. Why is it so difficult to govern mobile apps in healthcare? BMJ Health Care Inform 2019 Nov;26(1):e100006 [FREE Full text] [doi: 10.1136/bmjhci-2019-100006] [Medline: 31744843]
66. The free secure messaging apps for healthcare professionals. Siilo. URL: https://www.siilo.com/ [accessed 2022-04-20]
67. Trillian is modern and secure instant messaging for people, businesses and healthcare. Trillian. URL: https://trillian.im/ [accessed 2022-04-20]
68. Private practice your way. Simple Practice. URL: https://www.simplepractice.com/ [accessed 2022-04-20]

XSL•FO
RenderX

69. How care teams connect and collaborate. Tiger Connect. URL: https://tigerconnect.com/ [accessed 2022-04-20]
70. OnCall Health. URL: https://www.oncallhealth.com/ca/ [accessed 2022-04-20]
71. Lee S, Kim J, Kwon Y, Kim T, Cho S. Privacy Preservation in Patient Information Exchange Systems Based on Blockchain: System Design Study. J Med Internet Res 2022 Mar 22;24(3):e29108 [FREE Full text] [doi: 10.2196/29108] [Medline: 35315778]
72. Sengupta A, Subramanian H. User Control of Personal mHealth Data Using a Mobile Blockchain App: Design Science Perspective. JMIR Mhealth Uhealth 2022 Jan 20;10(1):e32104 [FREE Full text] [doi: 10.2196/32104] [Medline: 35049504]
73. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, Data Sharing, and Data Security Policies of Women's mHealth Apps: Scoping Review and Content Analysis. JMIR Mhealth Uhealth 2022 May 06;10(5):e33735 [FREE Full text] [doi: 10.2196/33735] [Medline: 35522465]
74. Parker L, Halter V, Karliychuk T, Grundy Q. How private is your mental health app data? An empirical study of mental health app privacy policies and practices. Int J Law Psychiatry 2019;64:198-204. [doi: 10.1016/j.ijlp.2019.04.002] [Medline: 31122630]
75. Developer of popular women's fertility-tracking app settles FTC allegations that it misled consumers about the disclosure of their health data. Federal Trade Commission. URL: https://www.ftc.gov/news-events/news/press-releases/2021/01/developer-popular-womens-fertility-tracking-app-settles-ftc-allegations-it-misled-consumers-about [accessed 2022-06-20]
76. Global strategy on digital health 2020-2025. World Health Organization. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2022-06-20]
77. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding Health Behavior Technology Engagement: Pathway to Measuring Digital Behavior Change Interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]

Abbreviations

DHI: digital health intervention
GDPR: general data protection regulation
OCE: omnichannel engagement

Edited by A Mavragani; submitted 26.07.22; peer-reviewed by MS Aslam, M Lewis, S Madanian; comments to author 07.09.22; revised version received 27.09.22; accepted 13.10.22; published 16.11.22.

Please cite as:
Blasiak A, Sapanel Y, Leitman D, Ng WY, De Nicola R, Lee VV, Todorov A, Ho D
Omnichannel Communication to Boost Patient Engagement and Behavioral Change With Digital Health Interventions
J Med Internet Res 2022;24(11):e41463
URL: https://www.jmir.org/2022/11/e41463
doi:10.2196/41463
PMID:36383427

©Agata Blasiak, Yoann Sapanel, Dana Leitman, Wei Ying Ng, Raffaele De Nicola, V Vien Lee, Atanas Todorov, Dean Ho. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

A Wolf in Sheep's Clothing: Reuse of Routinely Obtained Laboratory Data in Research

L Malin Overmars[1*], MSc; Michael S A Niemantsverdriet[1,2*], MSc; T Katrien J Groenhof[3], MD, PhD; Mark C H De Groot[1], PhD; Cornelia A R Hulsbergen-Veelken[1], PhD; Wouter W Van Solinge[1], PhD; Ruben E A Musson[1], PhD; Maarten J Ten Berg[1], PhD; Imo E Hoefer[1], MD, PhD; Saskia Haitjema[1], MD, PhD

[1]Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[2]SkylineDx, Rotterdam, Netherlands

[3]Department of Obstetrics and Gynecology, St Antonius Hospital, Utrecht, Netherlands

[*]these authors contributed equally

Corresponding Author:
Saskia Haitjema, MD, PhD
Central Diagnostic Laboratory
University Medical Center Utrecht
Utrecht University
Heidelberglaan 100
Utrecht, 3584 CX
Netherlands
Phone: 31 887550759
Email: s.haitjema@umcutrecht.nl

Abstract

Electronic health records (EHRs) contain valuable data for reuse in science, quality evaluations, and clinical decision support. Because routinely obtained laboratory data are abundantly present, often numeric, generated by certified laboratories, and stored in a structured way, one may assume that they are immediately fit for (re)use in research. However, behind each test result lies an extensive context of choices and considerations, made by both humans and machines, that introduces hidden patterns in the data. If they are unaware, researchers reusing routine laboratory data may eventually draw incorrect conclusions. In this paper, after discussing health care system characteristics on both the macro and micro level, we introduce the reader to hidden aspects of generating structured routine laboratory data in 4 steps (ordering, preanalysis, analysis, and postanalysis) and explain how each of these steps may interfere with the reuse of routine laboratory data. As researchers reusing these data, we underline the importance of domain knowledge of the health care professional, laboratory specialist, data manager, and patient to turn routine laboratory data into meaningful data sets to help obtain relevant insights that create value for clinical care.

(J Med Internet Res 2022;24(11):e40516) doi:10.2196/40516

KEYWORDS

laboratory data; electronic health records; preprocessing; applied data science; laboratory; data; clinical; decision support; decision; research; analysis; patient; value; clinical care

Introduction

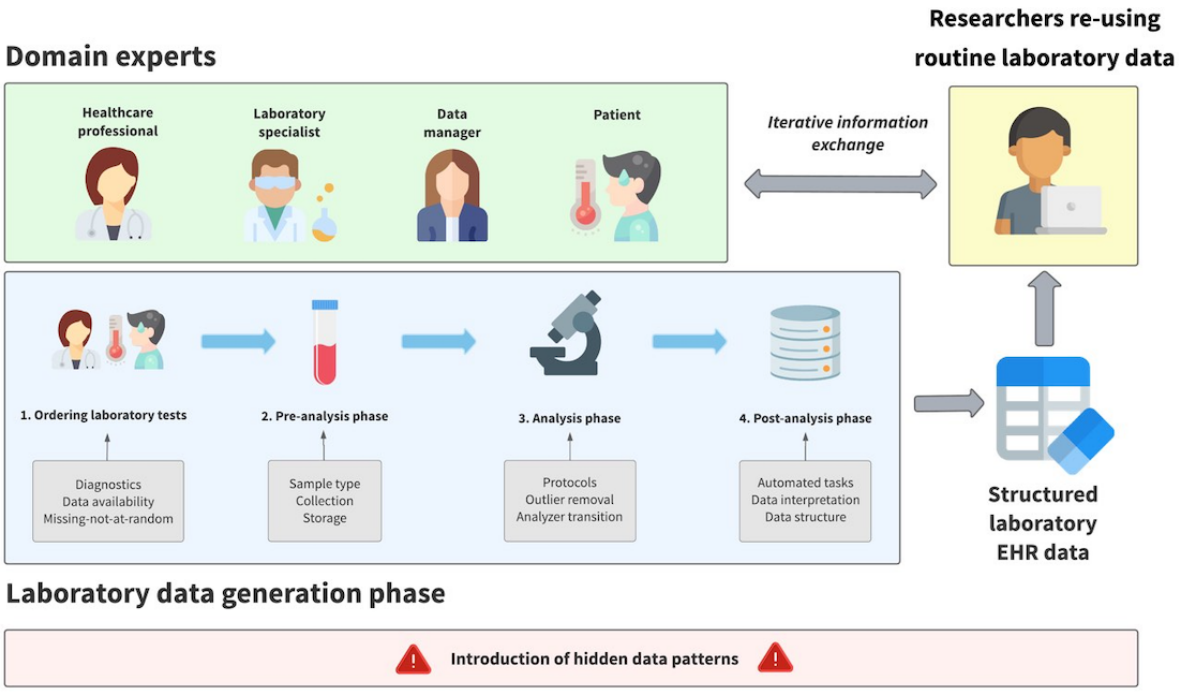
The availability of routinely collected laboratory data in electronic health records (EHRs) provides valuable information for medical diagnostics and decision-making in routine care [1]. Data extracted from EHR databases are often reused for deduction of knowledge to perform health care quality evaluations, conduct clinical and epidemiological studies, build clinical decision support systems, and facilitate disease understanding [2]. Moreover, the continuous increase of computing power and the introduction of new machine learning methods will likely further increase the reuse of the vast amounts of routine laboratory data to further personalize care [3]. As laboratory data are abundantly present, often numeric, generated by certified laboratories, and stored in a structured way, one may assume that the data are immediately suitable for (re)use. Technically, this may be true, for such data do not require additional "translation" steps as is the case for unstructured data (eg, written text). However, behind each laboratory test result lies an extensive context of choices and considerations, made by both humans and machines, that introduces hidden patterns in the data. Researchers reusing routine laboratory data, such

XSL•FO
RenderX

as epidemiologists and data scientists, who are oblivious to this "world behind the numbers" may either apply or omit (un)necessary preprocessing steps important for the creation of a clinically meaningful data set that may lead to false conclusions. Understanding different steps involved in the generation of routine laboratory data and the insights of domain experts responsible for these steps enables appropriate multidisciplinary preprocessing in medical research.

We will introduce the reader to these hidden aspects of laboratory data originating from the EHR database with examples drawn from our own experience. In 2005, the Central Diagnostic Laboratory of the University Medical Center Utrecht (UMCU) began collecting EHRs, including raw data from our International Organization for Standardization (ISO)-15189 certified laboratory, and compiling them in the Utrecht Patient Orientated Database (UPOD) [4]. With data from the UPOD, our department performed numerous studies in various disciplines over the past 20 years that taught us the necessity of multidisciplinary teams and preprocessing of routine laboratory data to make them suitable for (re)use. We describe the generation of these data in 4 steps and explain how each step may interfere with reuse of the data (Figure 1). First, we provide insights into how laboratory diagnostics is used in clinical care, which largely explains data availability and meaning. To fully understand why a test was ordered in the first place may be difficult (if not impossible) to answer, yet circumstantial evidence can often be found in data patterns and metadata. We also discuss the collection and analysis of specimens from patients in our ISO-15198 controlled laboratory setting. Finally, we elaborate on data processing storage by the laboratory information system (LIS) and EHR database, including automatic calculations and data corrections. We show how every step is reflected in the data, how it can hamper analysis, and then provide possible multidisciplinary preprocessing directions. Increasing the awareness of challenges in preprocessing routine laboratory data and the need for domain expertise may help turn raw laboratory data into meaningful laboratory data that can be reused for translation into value for clinical care.

Figure 1. The 4 steps involved in generating structured routine laboratory data and how domain experts and researchers should collaborate in establishing meaningful data sets for reuse.



Setting the Scene

Before diving into the ordering, (pre)analysis, and storage of routine laboratory data, one needs to be aware of large- and small-scale system characteristics that may affect the data at hand. For example, on a macrolevel, in many health care systems, first-line care is characterized by a more protocolized type of care aimed at the screening and exclusion of severe conditions; therefore, first-line care data carry a high number of "normal" values of a restricted set of laboratory tests. In contrast, specialized physicians in academic hospitals who look for less common diseases may require measurements of more "exotic" parameters, since standard examinations have not yet provided a definite diagnosis, resulting in a high number of abnormal values of a wide range of tests. An interesting development in this regard is the upcoming use of home testing. These tests are performed in uncontrollable collection circumstances and without connection to widely used laboratory information systems, circumventing the auditable data generating, approval, and release procedures we will discuss later in this paper. Limited interoperability between laboratory information systems and home devices may result in duplicate procedures for patients referred to different health care professionals or data storage in different systems that may be available to the treating clinicians but not to the researcher. Moreover, some health care systems reflect a more defensive way of practicing medicine because of medicolegal issues, which

XSL•FO
RenderX

means more laboratory data are generated as a result of routine screening procedures, attenuating the meaning of the results.

More specific system characteristics may include the 24/7 availability of resources such as expert laboratory staff, where the absence of laboratory values during the night may simply reflect a closed facility or the other way around—the presence of laboratory values may indicate acute disease during the night and laboratory testing could not wait until the morning. Available metadata in the form of time stamps is indeed collected in the laboratory process and may be used to determine the consecutive steps of the diagnostic workflow. Yet, some of these time stamps outside the lab may be less accurate, as they are written down by hand or indicate preferred collection times. Inside the lab, most time stamps are generated by analyzers and analysis tracks, leading to more precise metadata. Moreover, standard workflow protocols either following international guidelines or local policy may lead to the use of predefined sets of laboratory tests, reflex testing according to specific reference values, or point-of-care testing (POCT), in which the central diagnostic facility is not used due to warranted rapid analysis, preferably near the bedside.

According to the specific research question, patients, physicians, and laboratory specialists can help to provide insights into the local implications of health care systems and indicate workarounds and applicable current and historic clinical and laboratory guidelines that may affect data availability and meaning.

Step 1. Ordering Laboratory Tests

Ordering laboratory analyses is part of the process in which the patient and the health care professional interact to limit the number of possible differential diagnoses to a minimum. Medical tests meant to confirm or at least discard a potential diagnosis should always be ordered with a clear intent. For instance, a “shotgun approach” (ie, ordering many tests in the hope of being guided by their outcomes) may be a good strategy in the identification of new biomarkers. In modern medicine, however, such an approach will likely only result in high costs with a high chance of spurious findings [5]. The latter is a logical consequence of reference value or reference range definition. In laboratory medicine, reference ranges are defined as the 95% confidence interval. In other words, 95 out of 100 tests of any given parameter will fall within that reference range, and the remaining 5 will be considered an outlier. Statistically, assuming that all parameters to be completely independently regulated, testing for 5 parameters would result in 23% chance of at least 1 outlier, testing for 10 would result in a 40% chance, and testing for 20 would result in a 64% chance. While multiple comparisons in studies can be managed by applying correction factors, this is not done in clinical practice. Fortunately, diagnostic means such as laboratory tests are ordered with this knowledge in mind. Conversely, *not* testing for a specific parameter also needs to be regarded as meaningful, as the said parameter was most likely not considered to contribute to clinical decision-making. This is referred to as missing not at random (MNAR) patterns (ie, the choice for or against a test itself already contains meaning) [6].

In other words, many factors affect data availability. This has major implications for handling missing values. In general, it is not recommended to carelessly impute routine laboratory data, given the many different possible underlying reasons for missingness (Table 1). Unfortunately, the preceding deliberations that climax in the test order are usually not captured by order forms, in contrast to radiology or pathology requests, where the clinical question is an integral part (“pneumonia?” or “metastasis?”) [7]. A single laboratory test order can also be used to rule out or confirm more than 1 clinical question. Accordingly, the interpretation of a laboratory test result may differ between different clinical questions so that an identical result can have completely different meanings. This may be obvious in many cases. For example, a normal hemoglobin level in an elite athlete can be expected. In a patient with chronic fatigue, this may mean that the fatigue is probably not caused by anemia, whereas a normal hemoglobin level in a polytrauma patient with severe blood loss most likely only reflects successful blood transfusions. To make it more complicated, the interpretation of laboratory test results depends on the acuteness of the disease. For example, acute versus chronic anemia can acutely result from subacute severe hemorrhage but can also follow chronic blood loss in small amounts, iron deficiency, or bone marrow suppression.

The process of clinical reasoning is not part of laboratory standard operating procedures, so unfortunately, from a reuse point of view, it is not captured in any (meta)data. Because it is influenced by multiple factors, it can introduce patterns in data. Therefore, clinical characteristics influence availability and meaning in this very first step of data generation and are among the hardest to discover and, consequently, the hardest to model in projects reusing laboratory data. Physicians and patients as domain experts may help shed some light in this step, for example on how to distinguish between the different meanings of missing data (see Table 1 as a conversation starter), yet the specific choices that affect data availability may vary.

XSL•FO
RenderX

Table 1. Examples of physician-initiated and system-initiated processes that may affect the presence of laboratory values.

Reasons (not) to perform laboratory diagnostics	Example	Meaningful missingness pattern
Physician-initiated		
No laboratory measurements needed because there is no sign the of the disease	No CRP[a] measurement was ordered because patient did not have a fever	CRP missing: most likely means a *normal* value
No laboratory measurements needed because the disease is very obvious	No herpes zoster antibodies were ordered because the patient displayed shingles	Antibodies missing: most likely means an *abnormal* value
To confirm a diagnosis	Hemoglobin ordered for anemia	Hemoglobin missing: may either mean *normal* value or *abnormal* value
To exclude diagnosis	Metanephrines ordered for pheochromocytoma	Metanephrines missing: may either mean *normal* value or *abnormal* value
To determine treatment	Assess kidney function to titrate dosage of antibiotics	Kidney function missing: probably means a *normal* value.
Preprocedural risk assessment	Determination of INR[b] before thrombolysis	INR missing: probably means a *normal* value.
To monitor disease activity	Bacterial infection followed up by a lactate measurement	Lactate missing: most likely means a *normal* value
To monitor treatment effect	CA125 ovarian tumor marker ordered	CA125 missing: most likely means a *normal* value
Drug adherence	Low density lipoprotein cholesterol ordered	LDL[c] missing: probably means a *normal* value
Physician uncertainty or the reassurance of a patient	Lyme disease screening ordered	Lyme disease screening missing: most likely means a *normal* value
Screening	Fecal occult blood test for colorectal cancer	Fecal occult blood test missing: probably means a *normal* value.
Periodic health check-ups	Fasting glucose test for diabetes	Fasting glucose missing: most likely means a *normal* value
System-initiated		
Standard protocol within EHR[d]	Specific triaging protocol in the emergency department for suspected sepsis	Missingness describes the patient population: sepsis not suspected
Standard combinations of parameters that cannot be ordered separately	Hemoglobin and hematocrit	Missingness of either may indicate the measurement has failed
Reflex tests	Free T4 is measured only when TSH[e] is outside reference range	Missingness of free T4 if TSH is available means that free T4 is most likely *normal*
Automatic alerts in laboratory analyzers	Immature cells detected in blood by hematology analyzer	Missingness of immature blood cells means there are none detected

[a]CRP: c-reactive protein.

[b]INR: international normalized ratio.

[c]LDL: low-density lipoprotein.

[d]EHR: electronic health record.

[e]TSH: thyroid-stimulating hormone.

Step 2. Before the Laboratory: Preanalysis Phase

Depending on the considerations by the health care professional and their patient, the clinical presentation, and the ordered test, the next choice to be made concerns the collection of biological material. Blood and urine collection are widely known, but modern laboratories run diagnostic tests on many more body fluids and materials, such as spinal fluid, serosal fluids, or feces. The number of possibilities further increases by (1) the method of sample collection (eg, venous vs capillary blood sampling) and (2) the collection material (eg, ethylenediaminetetraacetic acid vs citrate-buffered blood collection tubes). These 2 factors may seem minor at first glance but have a major impact on the "test menu" (ie, which analyses can be performed as well as result reliability). For example, in the UMCU, there are over 2900 different tests. Some tests can only be performed using specific collection tubes, and while some can be performed in minute amounts, others may require sample volumes that preclude capillary blood sampling. Finally, time is relevant. For example, glucose levels in whole blood samples significantly decrease over time due to their consumption by blood cells [8]—not to mention circadian variations of numerous analytes. Laboratory specialists are trained to support the treating physician in assessing the possibilities and making the right decision to avoid diagnostic and treatment delay, unnecessary

XSL•FO
RenderX

additional testing, and associated costs. Altogether, this phase is usually referred to as "preanalysis."

The following example illustrates the often-underestimated impact of the preanalytic choices: A "simple" glucose level can be measured in various body fluids, such as blood or liquor (spinal fluid) in this case. Glucose levels in liquor are usually 30% to 40% lower than in blood [9]. Another well-known example is the concentration difference of creatinine in blood and urine. On an individual level, such factors may be easy to look up and adjudicate. In large data sets, this becomes impossible. A glucose level of 50 mg/dL may mean that it was measured in liquor and can be considered "normal," whereas when measured in blood, it may indicate a hypoglycemia. Therefore, is it essential to account for specimen types in available metadata and perform the preprocessing accordingly by either filtering out measurements or annotating them as separate variables to avoid false interpretation. Standards for laboratory test identification and naming have been proposed but are unfortunately not widely adopted [10]. Hence, laboratory specialists and data managers must cooperate when interpreting available data sources and deduce specimen type in the absence of unique identifiers that account for the available combinations of the aforementioned factors.

In addition to intrinsic circadian variations, patient behavior and actions before and/or during specimen collection (eg, medication or food intake) can affect test results. A well-known example is the variability of blood glucose levels in relation to recent food intake. Therefore, blood glucose concentrations are usually measured after at least 8 hours of fasting. In practice, not all patients faithfully report previous eating or drinking (other than water, unsweetened tea, or coffee) before sample collection. Therefore, even controlled and lege artis performed sample collection and analysis cannot guarantee a valid reflection of patient status. As such, accurate annotation (fasting/nonfasting), where possible, is important to either remove or include these measurements depending on the research question. Patient compliance to drugs or instructions varies significantly between treatments or patient groups. Physicians can help identify or even point to specific patients or patient groups that might distort the data set, thereby providing additional meaning to the data.

Step 3. In the Laboratory: Analysis Phase

Clinical laboratories use different types of analyzers to perform tests on the obtained sample. New analyzers, reagents, biomarkers, and software packages are introduced and updated frequently, and tests can be added, removed, or changed. Analyzers are regularly calibrated, but both distribution shifts and continuity imperfections within calibration ranges may occur over time and become apparent in large data sets. Although in most cases analytes are named identically across different laboratories, their actual level may differ significantly even when the same sample is measured. This can be due to different types of analyzers, different assay types, or different analyzer/assay manufacturers [11]. It is critical to note that the resulting differences in reported analyte concentrations can be substantial when comparing assays from different providers. In a nutshell, analyte X from provider Y is not the same as analyte X from provider Z. Harmonization efforts are ongoing but not yet widely implemented [12].

Within a given laboratory, the usual lifespan of analyzers is around 10 years. In other words, even in 1 hospital, analyzers need to be replaced every now and then. Accordingly, after transition, reported analyte concentrations can change from one day to the next. In certified laboratories, transition between analyzers is accompanied by an intermediate calibration phase in which the new analyzer is aligned. These calibration data are generated on specific test samples and do not usually appear in clinical data. However, when data are analyzed over a longer period, the analyzer transition can still be visible in the data. As an example, in 2013, our laboratory recalibrated the serum creatinine analysis; as a result, mean serum creatinine values were lower in the following years. Not having this background knowledge could easily result in the misinterpretation laboratory values over time and is therefore an essential step in quality control. Metadata analysis can help with identifying such changes and performing adequate correction actions.

Population-specific reference ranges (eg, age or sex dependent) are a form of clinical decision support, as they guide the physician in interpreting test results in clinical practice. Researchers reusing routine laboratory data can use these reference ranges to establish whether a test result deviates from a healthy population, assuming said reference indeed reflects health status. Reference ranges correspond to a set of values covering 95% of the results from testing a reference, such as a healthy population. Such populations are carefully selected in the approval process of any assay and analyzer based on defined criteria, such as sex, age, and pregnancy. These ranges are usually known along with the test result but may differ between laboratories, limiting the interoperability of findings. Grossly increased values that significantly deviate from reference values (eg, creatinine kinase in myocardial infarction and human chorionic gonadotropin in pregnancy) may seem outliers from a statistical point of view, but they may actually reflect the state of the human body and need to be interpreted on a log scale, as they are almost never the result of analytical error.

Initial outlier detection takes place in the laboratory as part of quality assurance and validation, where measurement errors are typically identified, signified, and even corrected. Therefore, applying rigorous outlier detection and removal by data reusers may remove highly informative test results and should be carefully deployed. Moreover, reference ranges can be modified over time. This can be the case when analyzers or assays change, as mentioned earlier in this paper, but can also be the case when new evidence becomes available. In cardiovascular medicine, for example, sex-specific reference thresholds are currently applied for cardiac troponin levels, a marker of cardiomyocyte damage or injury [13].

The LIS is the interface between analyzers and the EHR database. In addition to its role in test order management, it performs a variety of automated tasks essential for routine clinical care. It plays a central role in identifying aberrant measurements and performing simple and complex calculations such as low-density lipoprotein (LDL) cholesterol calculation

XSL•FO
RenderX

with the Friedewald formula [14]. These alerts and formulas may change over time, potentially resulting in inconsistent data or breaks in trends. Kidney function is usually monitored using an estimation of the glomerular filtration rate (GFR) instead of performing a "real" measurement, including urine collection, for 24 hours. Several formulas have been established and are applied clinically to this end. Our laboratory replaced the Modification of Diet in Renal Disease (MDRD) with the Chronic Kidney Disease Epidemiology Collaboration (CDK-EPI) formula in 2013. Consequently, the distribution of GFR measurements changed significantly over time due to considerable variations between the resulting minimum and maximum results. Depending on the research question at hand, such ambiguities may be resolved by recomputing the GFR for all patients with the same formula to obtain a single distribution when studying physiological processes. However, when studying health care processes, GFR values should not be recomputed, as health care professionals were only provided with the value calculated at that time and made their decisions accordingly. Thus, the choice whether using reported or recomputed data largely depends on the research question.

Laboratory values are only released into the LIS and EHR database after they have been approved by a laboratory specialist or a prespecified protocol. Values that surpass the preset assay specification (ie, lower limit of detection and upper limit of quantification) may be reported as "less than" or "greater than," respectively, and these modifiers may be found as a text variable in a different data location. The "greater than" and "less than" signs mean that the results fall outside the measurement range of the assay. However, this does not mean that they are incorrect, as laboratory specialists review these values and check their plausibility before approving them for reporting in the EHR. Some analyses may be of such complexity that they cannot be done by an automated analyzer and thus require manual interpretation by a laboratory specialist before release (eg, blood smears, bone marrow morphology, or specific blood type analyses). These are often accompanied by a free text comment from the laboratory specialist to aid the clinician with interpretation and may require natural language processing before being used further.

Step 4. After the Laboratory: Postanalysis Phase

After having been checked and approved, test results are reported in the LIS together with metadata describing analyzer configurations, sampling time, assay time, and more. The clinical laboratory is responsible for the correct laboratory value provision in clinical practice. However, for some analytes, only the result of the test ordered by the physician, accompanying comments of the laboratory specialist are eventually stored in the EHR, and raw data and metadata may be only available in the LIS. For specific research questions, such metadata can be indispensable for correct data analysis and interpretation and should therefore be retrieved in collaboration between data managers and laboratory specialists.

As previously exemplified in this paper, data continuity can be infringed at various levels. Analyzer updates may lead to different parameter names. Switching to a different analyzer, LIS, or EHR system can cause significant data variation. A parameter name change may seem minor for a human, like the addition of a single letter. In database terms, this may mean that a data column is added at a point in time, while no new data are added to the column reflecting the "old" name. This can be overcome with specialist knowledge input to merge data before and after the name was changed, provided that the change was indeed limited to the name and that the distribution of results has not changed over time. Doing so leads to a strong reduction in missing data. A variable name change can also be accompanied by a change in the definition of that parameter, resulting in aberrant distributions. Quantile-quantile and density plots can be used to demonstrate shifts after which they can be substantiated with, for example, a Kolmogorov-Smirnov test. Laboratory specialists are needed to help pointing out the cause of the change and help decide on preprocessing protocols and if data can be merged. In the future, data standards such as Logical Observation Identifiers Names and Codes (LOINC) or Systematized Nomenclature of Medicine Clinical Terms (SNOMED) may partly resolve some of these issues.

Conclusion

Routine laboratory data are a potential information gold mine for reuse beyond their purpose for care, as they are abundantly present, most often numeric, generated by certified laboratories, and stored in a structured manner. However, they can be a wolf in sheep's clothing. To fully exploit their potential, one should be aware of their caveats because statistical methods are generally not designed to incorporate their intricate clinical context, particularly not in the absence of metadata. Laboratory values are only part of the full clinical picture that health care professionals use to diagnose and treat patients. Moreover, this picture is merely reflected and not fully captured by routine health care data. Therefore, as researchers reusing routine laboratory data, we underline the prerequisite of collaboration between the health care professional, laboratory specialist, data manager, and patient to accurately recreate a valuable proxy of the clinical state of the patient by addressing and resolving all imperfections in electronic health record data, especially laboratory data [15]. Instead of massive computing power, this collaborative human effort may well be the only way to leverage the potential of artificial intelligence by turning routine data into meaningful data sets to help obtain relevant insights that create value for clinical care.

Acknowledgments

We would like to thank all patients, health care professionals, laboratory specialists, data managers, researchers, data scientists, and other domain experts that shared their valuable insights in collaborative efforts using the Utrecht Patient Oriented Database (UPOD) over the past 20 years for thus helping shape this paper.

XSL•FO
RenderX

Authors' Contributions

Conception and design: LMO, MSAN, and SH were responsible for the conception and design of this paper. All authors were involved in writing the manuscript and approving its final version.

Conflicts of Interest

MSAN is supported by a PhD fellowship from SkylineDx, Rotterdam. SH is supported by an Abbott Diagnostics Fellowship. All others have no competing financial interests to declare.

References

1. Blumenthal D, Glaser JP. Information technology comes to medicine. N Engl J Med 2007 Jun 14;356(24):2527-2534. [doi: 10.1056/nejmhpr066212] [Medline: 17568035]
2. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical data reuse or secondary use: current status and potential future progress. Yearb Med Inform 2017 Sep 11;26(01):38-52. [doi: 10.15265/iy-2017-007]
3. 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]
4. Ten Berg BM, Huisman A, Van den Bemt PMLA, Schobben AFAM, Egberts ACG, Van Solinge WW. Linking laboratory and medication data: New opportunities for pharmacoepidemiological research. CCLM 2007;45(1):9. [doi: 10.1515/cclm.2007.009]
5. Bindraban RS, van Beneden ML, Kramer MH, van Solinge WW, Neppelenbroek SI, van Wijnen M, et al. A multicenter before-after study on reducing unnecessary diagnostics by changing the attitude of caregivers: protocol for the RODEO Project. JMIR Res Protoc 2018 Aug 21;7(8):e10473 [FREE Full text] [doi: 10.2196/10473] [Medline: 30131316]
6. Groenwold RHH. Informative missingness in electronic health record systems: the curse of knowing. Diagn Progn Res 2020 Jul 02;4(1). [doi: 10.1186/s41512-020-00077-0]
7. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. BMJ 2018 Apr 30;361:k1479 [FREE Full text] [doi: 10.1136/bmj.k1479] [Medline: 29712648]
8. Fischer M, Hannemann A, Winter T, Schäfer C, Petersmann A, Nauck M. Relative efficacy of different strategies for inhibition of in vitro glycolysis. Clin Chem 2021 Jul 06;67(7):1032-1034. [doi: 10.1093/clinchem/hvab071] [Medline: 34104943]
9. Mundt L. Graff's Textbook of Urinalysis and Body Fluids. Burlington, MA: Jones & Bartlett; 2015.
10. Bodenreider O, Cornet R, Vreeman D. Recent Developments in Clinical Terminologies - SNOMED CT, LOINC, and RxNorm. Yearb Med Inform 2018 Aug 29;27(1):129-139 [FREE Full text] [doi: 10.1055/s-0038-1667077] [Medline: 30157516]
11. Allardet-Servent J, Lebsir M, Dubroca C, Fabrigoule M, Jordana S, Signouret T, et al. Point-of-care versus central laboratory measurements of hemoglobin, hematocrit, glucose, bicarbonate and electrolytes: a prospective observational study in critically ill patients. PLoS One 2017 Jan 10;12(1):e0169593 [FREE Full text] [doi: 10.1371/journal.pone.0169593] [Medline: 28072822]
12. Plebani M. Harmonization in laboratory medicine: more than clinical chemistry? Clin Chem Lab Med 2018 Sep 25;56(10):1579-1586 [FREE Full text] [doi: 10.1515/cclm-2017-0865] [Medline: 29095697]
13. Lee KK, Ferry AV, Anand A, Strachan FE, Chapman AR, Kimenai DM, High-STEACS Investigators. Sex-specific thresholds of high-sensitivity troponin in patients with suspected acute coronary syndrome. J Am Coll Cardiol 2019 Oct 22;74(16):2032-2043 [FREE Full text] [doi: 10.1016/j.jacc.2019.07.082] [Medline: 31623760]
14. Friedewald W, Levy R, Fredrickson D. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972;18(6):502. [doi: 10.1093/clinchem/18.6.499]
15. Bezemer T, de Groot MC, Blasse E, Ten Berg MJ, Kappen TH, Bredenoord AL, et al. A human(e) factor in clinical decision support systems. J Med Internet Res 2019 Mar 19;21(3):e11732 [FREE Full text] [doi: 10.2196/11732] [Medline: 30888324]

Abbreviations

CDK-EPI: Chronic Kidney Disease Epidemiology Collaboration
EHR: electronic health record
GFR: glomerular filtration rate
ISO: International Organization for Standardization
LDL: low-density lipoprotein
LIS: laboratory information system
LOINC: Logical Observation Identifiers Names and Codes
MDRD: Modification of Diet in Renal Disease
MNAR: missing not at random
POCT: point-of-care testing

XSL•FO
RenderX

SNOMED: Systematized Nomenclature of Medicine Clinical Terms
UMCU: University Medical Center Utrecht
UPOD: Utrecht Patient Orientated Database

Edited by G Eysenbach; submitted 24.06.22; peer-reviewed by CMJ Wong, C Schäfer; comments to author 30.08.22; revised version received 19.09.22; accepted 13.10.22; published 18.11.22.

Please cite as:
Overmars LM, Niemantsverdriet MSA, Groenhof TKJ, De Groot MCH, Hulsbergen-Veelken CAR, Van Solinge WW, Musson REA, Ten Berg MJ, Hoefer IE, Haitjema S
A Wolf in Sheep's Clothing: Reuse of Routinely Obtained Laboratory Data in Research
J Med Internet Res 2022;24(11):e40516
URL: https://www.jmir.org/2022/11/e40516
doi:10.2196/40516
PMID:36399373

©L Malin Overmars, Michael S A Niemantsverdriet, T Katrien J Groenhof, Mark C H De Groot, Cornelia A R Hulsbergen-Veelken, Wouter W Van Solinge, Ruben E A Musson, Maarten J Ten Berg, Imo E Hoefer, Saskia Haitjema. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Cost-effectiveness Analysis of a Mobile Phone–Based Integrated HIV-Prevention Intervention Among Men Who Have Sex With Men in China: Economic Evaluation

Ke Yun[1], PhD; Jiaming Yu[2], PhD, MD; Changyang Liu[3], MD; Xinxin Zhang[3], PhD

[1]Department of Laboratory Medicine, The First Affiliated Hospital of China Medical University, Shenyang, China

[2]Department of Hospital Infection Management, The Fourth Affiliated Hospital of China Medical University, Shenyang, China

[3]Department of Ophthalmology Laboratory, The Fourth Affiliated Hospital of China Medical University, Shenyang, China

Corresponding Author:
Xinxin Zhang, PhD
Department of Ophthalmology Laboratory
The Fourth Affiliated Hospital of China Medical University
No. 11 Xinhua Road
Shenyang, 110006
China
Phone: 86 024 62035015
Email: zhangxinxin@cmu.edu.cn

Abstract

Background: Mobile phone–based digital interventions have been shown to be a promising strategy for HIV prevention among men who have sex with men (MSM).

Objective: This study aimed to evaluate the cost-effectiveness of a mobile phone–based digital intervention for HIV prevention among MSM in China from the perspective of a public health provider.

Methods: The cost-effectiveness of the mobile phone–based digital intervention was estimated for a hypothetical cohort of 10,000 HIV-negative MSM who were followed for 1 year. A model was developed with China-specific data to project the clinical impact and cost-effectiveness of two mobile phone–based digital strategies for HIV prevention among MSM. The intervention group received an integrated behavioral intervention that included 1) individualized HIV infection risk assessment, 2) recommendation of centers testing for HIV and other STIs, 3) free online order of condoms and HIV and syphilis self-test kits and 4) educational materials about HIV/AIDS. The control group was only given educational materials about HIV/AIDS. Outcomes of interest were the number of HIV infections among MSM averted by the intervention, intervention costs, cost per HIV infection averted by the mobile phone–based digital intervention, and quality-adjusted life-years (QALYs). Univariate and multivariate sensitivity analyses were also conducted to examine the robustness of the results.

Results: It is estimated that the intervention can prevent 48 MSM from becoming infected with HIV and can save 480 QALYs. The cost of preventing 1 case of HIV infection was US $2599.87, and the cost-utility ratio was less than 0. Sensitivity analysis showed that the cost-effectiveness of the mobile phone–based digital intervention was mainly impacted by the average number of sexual behaviors with each sexual partner. Additionally, the higher the HIV prevalence among MSM, the greater the benefit of the intervention.

Conclusions: Mobile phone–based digital interventions are a cost-effective HIV-prevention strategy for MSM and could be considered for promotion and application among high-risk MSM subgroups.

(J Med Internet Res 2022;24(11):e38855) doi:10.2196/38855

KEYWORDS

cost-effectiveness; digital intervention; men who have sex with men; mathematical model

XSL•FO
RenderX

Introduction

HIV infection confers a heavy public health burden worldwide [1,2]. In 2011, the Joint United Nations Programme on HIV/AIDS set a goal to increase financing to US $22 billion to $24 billion in order to prevent and control AIDS by 2015 [3]. At that time, the prevention and control of AIDS were put on a fast track to advance the 90-90-90 goal: 90% of all people living with HIV will know their HIV status, 90% of all people with diagnosed HIV infection will receive sustained antiretroviral therapy, and 90% of all people receiving antiretroviral therapy will have viral suppression. To achieve this goal, US $35 billion needed to be added to prevention-effort funding per year up to 2020 [3,4].

When policy makers for HIV prevention and control projects plan and assess AIDS prevention schemes, they need to balance the cost and effectiveness of various preventive efforts from behavioral studies that have transformed into community practices. There is an urgent unmet need for development of effective and cost-effective intervention strategies to prevent and control HIV in MSM. There have been many explorations on economic assessments of HIV behavioral interventions all over the world. For example, one study concluded that a video-based outpatient intervention effectively reduced the incidence of sexually transmitted infections (STIs) [5]. It was estimated that, on average, US $447,005 per year was spent offering 10,000 STI clinic patients the intervention, which equates to a cost per patient of US $43.30; the intervention prevented an average of 27.69 patients from acquiring HIV infection and saved US $5,544,408 in medical expenses [5]. Another study found that safer-sex lectures and skills training for MSM significantly increased condom use rates, and the incremental cost of the intervention was less than the cost-effectiveness threshold (ie, US $13,000), indicating cost savings [6]. In China, some studies have also demonstrated the cost-effectiveness of different HIV-prevention strategies (ie, advocacy interventions, peer education, outreach services, and intervention staff training), including a risk-related behavioral intervention among female sex workers [7], active HIV voluntary counseling and testing among the general population [8], and oral pre-exposure prophylaxis and expanded antiretroviral therapy among MSM [9].

Recently, the internet is becoming an essential platform for health behavior interventions and provides new opportunities to deliver information about HIV prevention for MSM [10]. In previous studies, we developed an HIV risk–prediction model for MSM [11] and comprehensive behavioral intervention strategies based on this model [12]. The comprehensive intervention strategies were shown to reduce the number of male sexual partners and increase the rate of condom use with casual partners among MSM through a randomized controlled trial [12]. However, participation in online projects is often not well controlled, resulting from the broad coverage of internet-based interventions, and can easily lead to cost overruns. Additionally, health economic evaluations regarding scaling up mobile phone–based digital interventions for risk reduction among MSM are lacking. Therefore, the aim of this study was to evaluate the health economic value of mobile phone–based digital intervention strategies among a larger number of MSM from the perspective of a public health provider.

Methods

Model Structure and Parameters

A widely used international model based on probability was used to project the number of HIV infections averted as well as the cost-effectiveness of two different intervention strategies among MSM [13]. The mathematical formula is defined as follows:

$$P_{HIV} = 1 - \{P[1 - R(1 - FE)]^N + (1 - P)\}^M$$

where P represents the HIV prevalence among MSM, R is the risk of HIV infection during unprotected sex, and E is the protective effect of condoms. These three parameters were obtained from literature searches. In the formula, F is the rate of condom use during sexual behaviors, N represents each partner's average number of sexual behaviors, and M represents the average number of sexual partners. These three parameters were obtained from the questionnaires completed during the clinical trial. The estimated value of each parameter was the annual average. The time interval studied and analyzed in this section was 1 year. A total of 10,000 MSM were included in the simulation.

Epidemiological Data

The epidemiologic and behavioral parameters of HIV-infected patients are shown in Table 1 [12,14-20]. All the parameters above were estimated over an interval of 1 year. Critical model-fitting parameters, including HIV infection rate, HIV infectivity of one homosexual behavior, condom effectiveness, quality-adjusted life-years (QALYs) saved per HIV infection averted, lifetime medical costs after HIV infection, and the discount rate, were obtained by consulting the literature. According to clinical trial data, the average numbers of sexual partners in the intervention and control groups were 6.01 and 3.51, respectively. Condom use rates among MSM and their random sexual partners in the intervention and control groups were 0.86 and 0.70, respectively. According to the mathematical formula listed above, the probabilities of infection in the intervention and control groups were 0.002 and 0.007, respectively. The costs associated with the intervention and control groups were nearly equal. Moreover, one-third of the total costs were made up of labor and facility charges, which were the largest expenses for both groups.

XSL•FO
RenderX

Table 1. Model input parameters and costs.

Parameter	Value	Range	Data source
Both groups			
HIV infection rate	0.073	__[a]	Qin et al [14]
HIV infectivity rate of one homosexual behavior	0.008	0.002-0.02	Vittinghoff et al [15]
Rate of condom effectiveness	0.95	0.69-0.99	Pinkerton and Abramson [16]
The average number of sexual behaviors with each sexual partner, n	6	1.2-13.2	Zhang et al [17]
QALYs[b] saved per HIV infection averted, n	10	—	Hu et al [18]
Lifetime medical costs after HIV infection (US $)	25,803.41	—	Wang et al [20]
Discount rate, %	3	—	Wilde and Parke [19]
Control group			
Epidemiology			
Number of sexual partners, mean	6.01	—	Yun et al [12]
Use rate of condoms for each sexual behavior	0.7	—	Yun et al [12]
Infection probability	0.007	—	Calculated
Costs (US $)			
Recruitment spending	156.25	—	Observed
Condoms and lubricant	—	—	Observed
HIV self-test kits and supporting supplies	—	—	Observed
HIV risk–assessment software development	—	—	Observed
Human capital cost	843.75	—	Observed
Communication fees and electricity charges	93.75	—	Observed
Computers	1250	—	Observed
Online survey and intervention system	—	—	Observed
Office supplies (eg, paper and pens)	15.63	—	Observed
Intervention group			
Epidemiology			
Number of sexual partners, mean	3.51	—	Yun et al [12]
Use rate of condoms for each sexual behavior	0.86	—	Yun et al [12]
Infection probability	0.0022	—	Calculated
Costs (US $)			
Recruitment spending	156.25	—	Observed
Condoms and lubricant	70.31	—	Observed
HIV self-test kits and supporting supplies	93.75	—	Observed
HIV risk–assessment software development	93.75	—	Observed
Human capital cost	843.75	—	Observed
Communication fees and electricity charges	93.75	—	Observed
Computers	1250	—	Observed
Online survey and intervention system	87.5	—	Observed
Office supplies (eg, paper and pens)	15.63	—	Observed

[a]Not reported.

[b]QALY: quality-adjusted life-year.

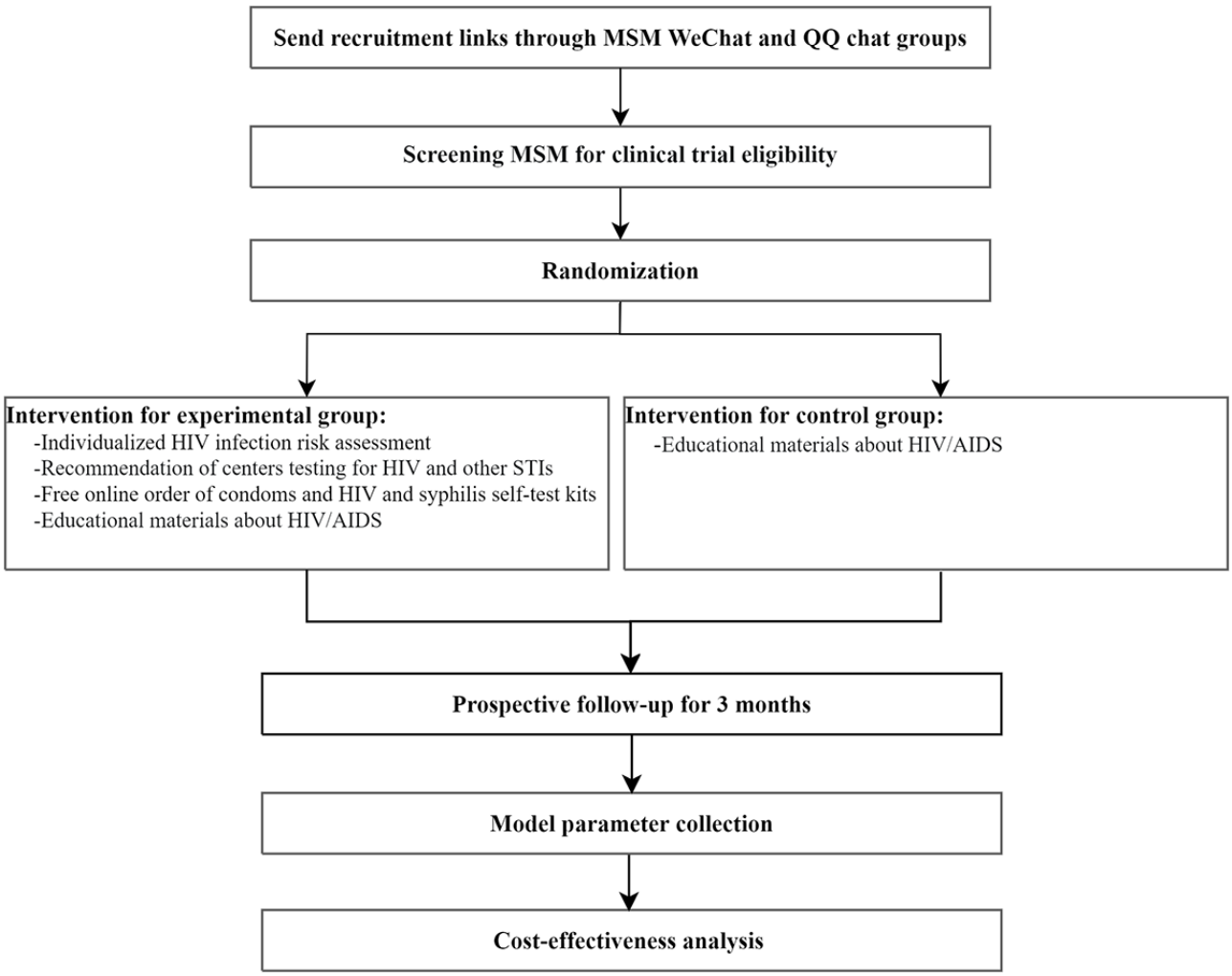
XSL•FO
RenderX

Intervention

These parameters were collected from a randomized, double-blind clinical trial, and the trial process is listed in Figure 1. With the assistance of a community-based organization, MSM were recruited from October 2017 to March 2018 through an advertisement in Tencent's WeChat and QQ chat groups, which are popular virtual dating environments for MSM in China. The changes in HIV-related behavioral indicators in the intervention and control groups of MSM in the clinical trial were published previously [12]. Briefly, we prospectively evaluated the efficacy of two mobile phone–based digital strategies targeting MSM in China.

The intervention group received an integrated behavioral intervention promoted through WeChat, which included 1) individualized HIV infection risk assessment, 2) recommendation of centers testing for HIV and other STIs, 3) free online order of condoms and HIV and syphilis self-test kits and 4) educational materials about HIV/AIDS. The control group was only given educational materials about HIV/AIDS.

Figure 1. Cost-effectiveness analysis flowchart of the mobile phone–based digital intervention. MSM: men who have sex with men; STI: sexually transmitted infection.



Measurement of Costs

Costs were estimated from the perspective of a public health provider; only the direct costs used for the intervention were evaluated. The costs, which were estimated according to the clinical trial, included human capital costs, nongovernmental organization recruitment and referral fees, communication fees, electricity charges, condom and lubricant costs, HIV self-test kit costs, postage fees, computer costs, fees related to the online questionnaire system, compensation for completing questionnaires, HIV risk–assessment software deployment costs, and office supplies (eg, paper and pens).

Estimation of QALYs Saved

The difference between the estimated values of behavioral parameters in the intervention and control groups was used to calculate the number of HIV-1 infections averted as a result of the intervention. QALYs represent the life-years obtained by the model multiplied by the definite utility value, shown as $QALY = \sum WY$, where W is the weight value and Y is the length of life. Quality of life–adjusted weights between 0 and 1 were used to calculate QALYs, also called health utility; this was the criterion used to assess health status and it reflected the severity of poor health. The total number of QALYs saved as a result of the intervention was calculated by the number of QALYs saved per HIV infection averted multiplied by the number of HIV infections averted. The cost-effectiveness ratio (CER) was

XSL•FO
RenderX

one of the indicators evaluating economics that was used in this study. A lower ratio indicated that a lower cost was needed to gain one indicator effect unit. In the cost-utility ratio (CUR) analysis, an effect indicator combining life-years and health-related quality of life was performed. CUR, the ratio between the net cost of the intervention and the total QALYs saved, reflected the cost of a beneficial intervention. The incremental CER (ICER), whose units are US dollars per QALY, was also calculated to reflect the incremental cost associated with one additional unit of the effect measure.

Sensitivity Analysis

As the values of many parameters were uncertain in the model, those of other parameters were controlled for consistency and to test the influence of parameter uncertainty on the final estimated results, which would help determine the essential factors affecting the cost effectiveness of the intervention. In this study, the one-way sensitivity analysis was mainly conducted to assess the average number of sexual behaviors with each sexual partner, HIV infectivity rate, QALYs saved per infection averted, HIV infection rate, rate of condom effectiveness, cost of offering intervention to 1 MSM (US $). Multiway sensitivity analysis was performed on the cost-effectiveness of the mobile phone–based digital intervention for HIV prevention as a function of intervention coverage and HIV prevalence among MSM.

Judging Criteria

The ICER represents the incremental cost associated with 1 QALY saved. This measurement was used to determine which strategy was better. The criterion applied by the World Health Organization to determine whether interventions are cost-effective suggests that strategies evaluated with an ICER of less than 0 are both effective and cost-effective, interventions with an ICER of less than the average per capita gross domestic product (GDP) for a given country or region are considered very cost-effective, interventions with an ICER of less than 3-fold average per capita GDP are still considered cost-effective, and those that exceed this level are considered not cost-effective [21]. The average per capita GDP for China in 2017 was US $9293.49, which was used as the threshold to indicate whether the interventions were cost-effective in this study. A negative CUR suggests that the cost of a strategy in the intervention group is less than the cost of a strategy in the control group, which defines a cost-saving intervention strategy. Simply, the lower the CUR, the fewer health resources that are consumed.

Ethics Approval

The Ethics Review Committee of the First Affiliated Hospital of China Medical University in Shenyang, China, approved this study (approval No. 2018-175-2).

Results

Cost-effectiveness Analysis

The average annual cost to offer the intervention to 10,000 MSM was estimated to be US $124,837.50, with a cost per MSM of US $12.48. The model estimated that 70 of the 10,000 MSM would be infected with HIV without the intervention (ie, control group parameters). The model also estimated that 22 of the 10,000 MSM would be infected with HIV if they received the intervention (ie, intervention group parameters were applied). HIV infections could be averted in 48 persons per year with the intervention, with an average cost of US $2599.87 per infection averted. The intervention saved 480 QALYs per year, equaling a savings of US $980,529.61. Each QALY saved was accompanied by an incremental cost of about US $260.20, and the CUR was negative, indicating cost-effectiveness. Details are shown in Table 2.

XSL•FO
RenderX

Table 2. The 1-year cost-effectiveness analysis of a mobile phone–based digital intervention among 10,000 men who have sex with men.

Parameters	Value	Data source
Intervention costs (US $)		
Recruiting advertisement fees	1562.50	Estimates from clinical trial data
Condoms and lubricant	28,125.00	Estimates from clinical trial data
HIV self-test kits and supporting supplies	37,500.00	Estimates from clinical trial data
Postage fees	37,500.00	Estimates from clinical trial data
Human capital costs	16,875.00	Estimates from clinical trial data
Communication fees and electricity charges	1875.00	Estimates from clinical trial data
Computers	1250.00	Estimates from clinical trial data
Online survey system	87.50	Estimates from clinical trial data
Office supplies (eg, paper and pens)	62.50	Estimates from clinical trial data
Annual cost	124,837.50	Estimates from clinical trial data
Cost of offering intervention to 1 MSM[a]	12.48	Estimates from clinical trial data
Intervention effect		
HIV infections averted, n	48	Calculated
QALYs[b] saved by the intervention, n	480	Calculated
Medical costs saved by the intervention (US $)	980,529.61	Calculated
CER[c]	260.20	Calculated
Cost associated with 1 QALY saved (US $)	260.20	Calculated
CUR[d]	<0	Calculated

[a]MSM: men who have sex with men.

[b]QALY: quality-adjusted life-year.

[c]CER: cost-effectiveness ratio.

[d]CUR: cost-utility ratio.

Sensitivity Analysis

Many parameters in the model were based on hypotheses and adjusted observed data, which led to various estimated outcomes under different situations; therefore, we performed a sensitivity analysis to test the effect of changes to the input parameters on the estimated outcomes of the model. The range of parameter fluctuations estimated according to the reported studies was adjusted to account for fluctuations of 33%. One-way sensitivity analyses showed that the number of infections averted was not sensitive to changes in most sexual behavior and epidemiologic parameters. Still, the average number of sexual behaviors with each sexual partner influenced the model's outcomes to a certain extent. When the average number of sexual behaviors with each sexual partner was small, the cost associated with 1 QALY saved was US $1560.01. However, when the average number of sexual behaviors with each sexual partner was large, the cost associated with 1 QALY saved was US $121.17. These results indicated that the mobile phone–based digital intervention was more cost-effective among MSM who had more high-risk sexual behaviors (Figure 2).

Multiway sensitivity analyses are shown in Figure 3, which showed ICER ranges for MSM as a function of intervention coverage and HIV prevalence. The cost-effectiveness of the mobile phone–based digital intervention depends on intervention coverage and HIV prevalence among MSM. A two-way sensitivity analysis showed that as MSM coverage increased, so did the cost, although the mobile phone–based digital intervention was still cost-effective in all simulated scenarios (HIV infection rate ranged from 0.02 to 0.14; intervention coverage ranged from 0.3 to 0.7). Additionally, the higher the HIV prevalence, the greater the benefit of the intervention.

XSL•FO
RenderX

Figure 2. Tornado diagram of the one-way sensitivity analysis summarizing the effect of parameters on the ICER. The extent of the fluctuations in the parameters obtained from the literature is estimated by the 33% increase or decrease in the parameters reported in the literature. ICER: incremental cost-effectiveness ratio; MSM: men who have sex with men; QALY: quality-adjusted life-year.

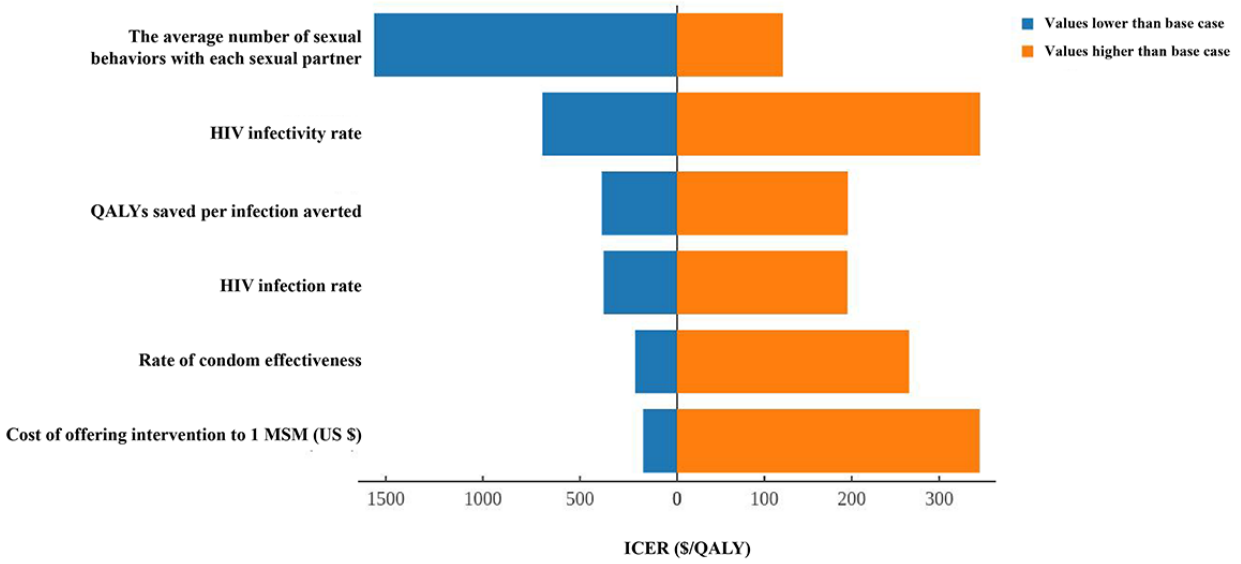
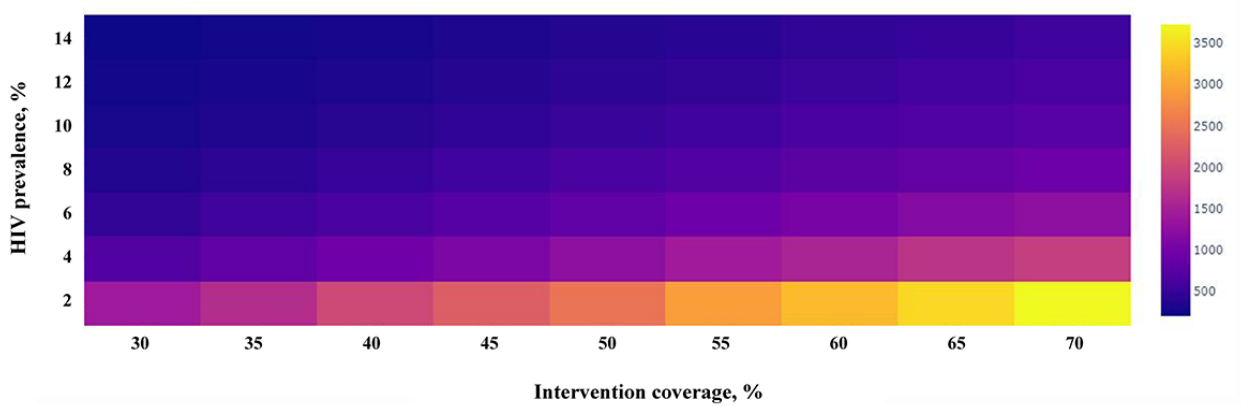


Figure 3. Multiway sensitivity analysis on the cost-effectiveness of the digital intervention for HIV prevention as a function of the intervention coverage and HIV prevalence among men who have sex with men. The colors signify the incremental cost-effectiveness ratios.



Discussion

Principal Findings

HIV infection leads to substantial health, economic, and social consequences [22]. This study was the first to perform an economic evaluation of a mobile phone–based digital intervention in China. Overall, the intervention had high cost-effectiveness; the average number of sexual behaviors with each sexual partner and the intervention coverage affected the model's outcomes. Since the number of people who avoided HIV infection as a result of the intervention cannot be directly measured, our study uses disease model prediction–based methods to estimate the number of people who avoided HIV infection as a result of the intervention. The cost-utility analysis used QALYs to consider both the length and quality of life, making it possible to compare different intervention strategies across the board. These findings provide a theoretical basis for decision-making regarding the government's intervention policies.

This study found that although the effect of the online, personalized, risk-reduction intervention was slight, many cases of HIV infection can be avoided at the population level to achieve the goal of saving QALYs. Similar findings have been confirmed by other studies. For example, Zhang et al [23] have used the HIV phylogenetic tree to reveal the influence of individual HIV infection risk fluctuations on HIV transmission dynamics. They found that the change in the individual risk of HIV infection could change not only the proportion of HIV transmission among MSM but also the endemic equilibrium of HIV infection. These findings also suggest that interventions to reduce the risk of HIV infection at the individual level are important for controlling the HIV epidemic, and that appropriate targeted intervention strategies should be developed as soon as possible.

Compared with the study that evaluated the health economics of facility-based interventions among MSM in China [20], our intervention was more cost-effective. The difference may be explained by differences in intervention patterns. Compared with facility-based intervention strategies, internet-based interventions have the obvious advantages of high recruitment

XSL•FO
RenderX

efficiency and low cost [10]. The essence of the intervention in this study was to convert offline consulting and testing into online risk self-assessment and self-testing (ie, facility-based testing or self-testing using testing strips). One of the reasons that this type of intervention strategy was cost-effective was that our study recommended individual intervention strategies and offered suggestions according to individual risk levels, leading to more precise resource allocation. In addition, internet-based interventions benefit hidden populations of MSM, especially young MSM [24,25], which is favorable for recruitment of large numbers of MSM. As a result, MSM benefit from improved access to health resources, expanded coverage of interventions, and improvements in the effectiveness of interventions [26]. In the future, we may consider building a real-time, interactive, big data intervention platform based on "Internet+," which will continue to enrich the content and form of the intervention. At the same time, this will increase the publicity and education efforts of the intervention through the internet development platform in order to narrow the gap between the MSM population and HIV-prevention information. This will promote a more efficient flow of prevention resources in order to avoid redundancy and waste of health resources and will allow more MSM to take the initiative in participating in AIDS prevention and control.

A cost-utility analysis, believed to be the gold standard of economic evaluation, was conducted that took both length and quality of life into account regarding QALYs to make horizontal comparisons of different intervention strategies possible. Compared with other primary prevention strategies for HIV, the cost per HIV infection averted in our study (US $3337) was lower than that of a community-level HIV risk reduction intervention for women (US $65,000) [27], a video-based intervention model in STI clinics for men (US $21,486) [5], HIV-prevention skills training for MSM (US $4150.14) [6], and an internet-based video intervention (US $14,926). The reason why this intervention model is superior to the above strategies may be that with other intervention strategies, such as video education, nonvideo education, and condom distribution, there are significant differences in HIV risk behavior across these subgroups, and the emphasis on the use of health resources is also different. Additionally, the sensitivity analysis in our study showed that the average number of sexual behaviors with each sexual partner influenced the cost associated with 1 QALY saved, which meant that interventions aimed at people with more high-risk sexual behaviors gained more health economic benefits. This suggested that the intervention strategy of this study is suitable for high-risk MSM in China, which could lead to accurate allocation of prevention resources based on risk assessment. Meanwhile, if the one-size-fits-all intervention model does not match the risk level of the target population, it will easily lead to a waste of resources. Another important reason why this intervention strategy is cost-effective might be because the intervention was internet based, which could save on housing and office facility costs, among others. Therefore, the risk assessment–based integrated online intervention is not only low cost but also likely to reach MSM who may not have the time, resources, or motivation to seek prevention services, which would improve the convenience and accessibility of prevention resources.

Limitations

There are several limitations in this study. First, the number of infections averted was estimated based on a model, and the model parameters were average values from the reported studies, which may have ignored the heterogeneity of data—the probability of HIV infection varies with the types of sexual partners—and may have impacted the estimated number of HIV infections averted. Further studies should distinguish the probability of HIV infection resulting from different subgroups, sexual roles, and sexual behavior patterns of MSM in order to accurately estimate the number of infections averted. Second, the number of sexual behaviors with each sexual partner was obtained from the literature and was not verified by questionnaires. Therefore, specific questionnaires need to be designed to investigate the number of sexual behaviors in order to test the reliability of input parameters.

Conclusions

Our analysis demonstrates that a mobile phone–based digital intervention to prevent HIV infection among MSM is very cost-effective. These findings can inform public health officials' and policy makers' decisions about the selection of, and recommendations for, intervention measures. These findings can also help in the exploration of innovative HIV intervention strategies in order to more effectively allocate resources to the prevention of HIV among MSM.

Acknowledgments

We thank the study participants who participated in the clinical trial. This research received funding from the Liaoning Natural Science Foundation Project (2020-BS-091).

Conflicts of Interest

None declared.

References

1. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]

XSL•FO
RenderX

2. Zhang W, Xu J, Zou H, Zhang J, Wang N, Shang H. HIV incidence and associated risk factors in men who have sex with men in Mainland China: An updated systematic review and meta-analysis. Sex Health 2016;13(4):373. [doi: 10.1071/sh16001]
3. Schwartländer B, Stover J, Hallett T, Atun R, Avila C, Gouws E, Investment Framework Study Group. Towards an improved investment approach for an effective response to HIV/AIDS. Lancet 2011 Jun 11;377(9782):2031-2041. [doi: 10.1016/S0140-6736(11)60702-2] [Medline: 21641026]
4. De Lay PR, Benzaken A, Karim QA, Aliyu S, Amole C, Ayala G, et al. Ending AIDS as a public health threat by 2030: Time to reset targets for 2025. PLoS Med 2021 Jun;18(6):e1003649 [FREE Full text] [doi: 10.1371/journal.pmed.1003649] [Medline: 34101728]
5. Sweat M, O'Donnell C, O'Donnell L. Cost-effectiveness of a brief video-based HIV intervention for African American and Latino sexually transmitted disease clinic clients. AIDS 2001 Apr 13;15(6):781-787. [doi: 10.1097/00002030-200104130-00014] [Medline: 11371693]
6. Pinkerton SD, Holtgrave DR, Valdiserri RO. Cost-effectiveness of HIV-prevention skills training for men who have sex with men. AIDS 1997 Mar;11(3):347-357. [doi: 10.1097/00002030-199703110-00013] [Medline: 9147427]
7. Han D, Luo Y, Jiang W, He Q, Wu X, Zhang J. Cost-effectiveness analysis of different AIDS interventions for female sex workers in entertainment service places in Chengdu. Chin J Dis Control 2012;16(3):216-218.
8. Zeng G. Evaluation of health economics of active AIDS testing and counseling strategies in medical institutions. China Center for Disease Control and Prevention. 2012. URL: https://cdmd.cnki.com.cn/Article/CDMD-84501-1016282229.htm [accessed 2022-10-17]
9. Jin X, Shi L, Wang C, Qiu T, Yin Y, Shen M, et al. Cost-effectiveness of oral pre-exposure prophylaxis and expanded antiretroviral therapy for preventing HIV infections in the presence of drug resistance among men who have sex with men in China: A mathematical modelling study. Lancet Reg Health West Pac 2022 Jun;23(4):100462-100633 [FREE Full text] [doi: 10.1016/j.lanwpc.2022.100462] [Medline: 35542892]
10. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: A systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
11. Yun K, Xu J, Leuba S, Zhu Y, Zhang J, Chu Z, et al. Development and validation of a personalized social media platform-based HIV incidence risk assessment tool for men who have sex with men in China. J Med Internet Res 2019 Jun 18;21(6):e13475 [FREE Full text] [doi: 10.2196/13475] [Medline: 31215509]
12. 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]
13. Sweat M, Gregorich S, Sangiwa G, Furlonge C, Balmer D, Kamenga C, et al. Cost-effectiveness of voluntary HIV-1 counselling and testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania. Lancet 2000 Jul 08;356(9224):113-121. [doi: 10.1016/S0140-6736(00)02447-8] [Medline: 10963247]
14. Qin Q, Tang W, Ge L, Li D, Mahapatra T, Wang L, et al. Changing trend of HIV, syphilis and hepatitis C among men who have sex with men in China. Sci Rep 2016 Aug 18;6:31081 [FREE Full text] [doi: 10.1038/srep31081] [Medline: 27535092]
15. Vittinghoff E, Douglas J, Judson F, McKirnan D, MacQueen K, Buchbinder SP. Per-contact risk of human immunodeficiency virus transmission between male sexual partners. Am J Epidemiol 1999 Aug 01;150(3):306-311. [doi: 10.1093/oxfordjournals.aje.a010003] [Medline: 10430236]
16. Pinkerton SD, Abramson PR. Effectiveness of condoms in preventing HIV transmission. Soc Sci Med 1997 May;44(9):1303-1312. [doi: 10.1016/s0277-9536(96)00258-4]
17. Zhang K, Li D, Li H, Beck EJ. Changing sexual attitudes and behaviour in China: Implications for the spread of HIV and other sexually transmitted diseases. AIDS Care 1999 Oct;11(5):581-589. [doi: 10.1080/09540129947730] [Medline: 10755033]
18. Hu Z, Ma Y, Qin X, Jiang Q, Zhang X. Application of cost-utility analysis in evaluation of AIDS behavioral intervention. China Health Econ 2004;23(6):31-33.
19. Wilde PE. Cost Effectiveness in Health and Medicine. In: Journal of Nutrition Education and Behavior. New York: Oxford University Press; Sep 2017:707-708.
20. Wang S, Moss JR, Hiller JE. The cost-effectiveness of HIV voluntary counseling and testing in China. Asia Pac J Public Health 2011 Jul;23(4):620-633. [doi: 10.1177/1010539511412576] [Medline: 21727082]
21. Griffiths M, Maruszczak M, Kusel J. The WHO-CHOICE cost-effectiveness threshold: A country-level analysis of changes over time. Value Health 2015 May;18(3):PA88. [doi: 10.1016/j.jval.2015.03.517]
22. Shah M, Risher K, Berry SA, Dowdy DW. The epidemiologic and economic impact of improving HIV testing, linkage, and retention in care in the United States. Clin Infect Dis 2016 Jan 15;62(2):220-229 [FREE Full text] [doi: 10.1093/cid/civ801] [Medline: 26362321]
23. Zhang X. Exploring Effect of Individual Risk Behavior Volatility on Model Inference of HIV Transmission Dynamics and HIV Phylogenetic Tree. 2017. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/140946/xinyuz_1.pdf?isAllowed=y&sequence=1 [accessed 2022-10-17]

XSL•FO
RenderX

24. Saxton P, Dickson N, Hughes A. Who is omitted from repeated offline HIV behavioural surveillance among MSM? Implications for interpreting trends. AIDS Behav 2013 Nov;17(9):3133-3144. [doi: 10.1007/s10461-013-0485-1] [Medline: 23605157]
25. Sanchez T, Smith A, Denson D, Dinenno E, Lansky A. Internet-based methods may reach higher-risk men who have sex with men not reached through venue-based sampling. Open AIDS J 2012;6:83-89 [FREE Full text] [doi: 10.2174/1874613601206010083] [Medline: 23049657]
26. Zhang A, Xia M, Liu Y, Wang X, Zhu D, Long F, et al. [Research progress in mobile phone software for health interventions in men who have sex with men]. Zhonghua Liu Xing Bing Xue Za Zhi 2015 Nov;36(11):1315-1318. [Medline: 26850258]
27. Pinkerton SD, Holtgrave DR, DiFranceisco WJ, Stevenson LY, Kelly JA. Cost-effectiveness of a community-level HIV risk reduction intervention. Am J Public Health 1998 Aug;88(8):1239-1242. [doi: 10.2105/ajph.88.8.1239] [Medline: 9702159]

Abbreviations

CER: cost-effectiveness ratio
CUR: cost-utility ratio
GDP: gross domestic product
ICER: incremental cost-effectiveness ratio
MSM: men who have sex with men
QALY: quality-adjusted life-year
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 20.04.22; peer-reviewed by W Hui; comments to author 09.06.22; revised version received 27.06.22; accepted 01.10.22; published 02.11.22.

Please cite as:
Yun K, Yu J, Liu C, Zhang X
A Cost-effectiveness Analysis of a Mobile Phone–Based Integrated HIV-Prevention Intervention Among Men Who Have Sex With Men in China: Economic Evaluation
J Med Internet Res 2022;24(11):e38855
URL: https://www.jmir.org/2022/11/e38855
doi:10.2196/38855
PMID:36322123

©Ke Yun, Jiaming Yu, Changyang Liu, Xinxin Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Codevelopment of "My Kidneys & Me": A Digital Self-management Program for People With Chronic Kidney Disease

Courtney J Lightfoot[1,2], PhD; Thomas J Wilkinson[1,3], PhD; Michelle Hadjiconstantinou[4], PhD, CPsychol; Matthew Graham-Brown[5,6], PhD; Jonathan Barratt[5,6], PhD; Christopher Brough[7], BSc; James O Burton[5,6,8], DM; Jenny Hainsworth[9,10], ClinPsyD; Vicki Johnson[7], BSc; Maria Martinez[11], MSc; Andrew C Nixon[12,13], MBChB; Victoria Pursey[14], MSc; Sally Schreder[7], MSc; Noemi Vadaszy[1], MSc; Lucina Wilde[1], MSc; Fiona Willingham[15,16], MSc; Hannah M L Young[4,7], PhD; Thomas Yates[2,17], PhD; Melanie J Davies[2,4], MD; Alice C Smith[1,2], PhD

[1]Leicester Kidney Lifestyle Team, Department of Health Sciences, University of Leicester, Leicester, United Kingdom
[2]Leicester Biomedical Research Centre, Leicester, United Kingdom
[3]National Institute for Health Research Applied Research Collaboration East Midlands, Leicester Diabetes Centre, Leicester, United Kingdom
[4]Diabetes Research Centre, University of Leicester, Leicester, United Kingdom
[5]Department of Cardiovascular Sciences, University of Leicester, Leicester, United Kingdom
[6]Department of Renal Medicine, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom
[7]Leicester Diabetes Centre, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom
[8]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom
[9]Department of Medical Psychology, Leicestershire Partnership NHS Trust, Leicester, United Kingdom
[10]University Hospitals of Leicester NHS Trust, Leicester, United Kingdom
[11]Renal and Transplant Unit, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom
[12]Department of Renal Medicine, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, United Kingdom
[13]Division of Cardiovascular Sciences, University of Manchester, Manchester, United Kingdom
[14]Department of Renal Medicine, York Teaching Hospital NHS Foundation Trust, York, United Kingdom
[15]Nutrition and Dietetics Team, School of Sport and Health Sciences, University of Central Lancashire, Preston, United Kingdom
[16]Nutrition and Dietetic Department, East Lancashire Hospitals NHS Trust, Blackburn, United Kingdom
[17]Diabetes Research Centre, College of Life Sciences, University of Leicester, Leicester, United Kingdom

Corresponding Author:
Courtney J Lightfoot, PhD
Leicester Kidney Lifestyle Team
Department of Health Sciences
University of Leicester
University Road
Leicester, LE1 7RH
United Kingdom
Phone: 44 1162584346
Email: courtney.lightfoot@leicester.ac.uk

Abstract

Background: Health care self-management is important for people living with nondialysis chronic kidney disease (CKD). However, the few available resources are of variable quality.

Objective: This work describes the systematic codevelopment of "My Kidneys & Me" (MK&M), a theory-driven and evidence-based digital self-management resource for people with nondialysis CKD, guided by an established process used for the successful development of the diabetes education program MyDESMOND (Diabetes Education and Self-Management for Ongoing and Newly Diagnosed, DESMOND).

Methods: A multidisciplinary steering group comprising kidney health care professionals and researchers and specialists in the development of complex interventions and digital health provided expertise in the clinical and psychosocial aspects of CKD,

XSL•FO
RenderX

self-management, digital health, and behavior change. A patient and public involvement group helped identify the needs and priorities of MK&M and co-design the resource. MK&M was developed in 2 sequential phases. Phase 1 involved the codevelopment process of the MK&M resource (content and materials), using Intervention Mapping (IM) as a framework. The first 4 IM steps guided the development process: needs assessment was conducted to describe the context of the intervention; intervention outcomes, performance objectives, and behavioral determinants were identified; theory- and evidence-based change methods and practical strategies to deliver change methods were selected; and program components were developed and refined. Phase 2 involved the adoption and adaptation of the existing MyDESMOND digital platform to suit the MK&M resource.

Results: The needs assessment identified that individuals with CKD have multiple differing needs and that delivering a self-management program digitally would enable accessible, tailored, and interactive information and support. The intended outcomes of MK&M were to improve and maintain effective self-management behaviors, including physical activity and lifestyle, improve knowledge, promote self-care skills, increase self-efficacy, and enhance well-being. This was achieved through the provision of content and materials designed to increase CKD knowledge and patient activation, reduce health risks, manage symptoms, and improve physical function. Theories and behavior change techniques selected include Self-Management Framework, Capability, Opportunity, Motivation Behavior model components of Behaviour Change Wheel and taxonomy of behavior change techniques, Health Action Process Approach Model, Common Sense Model, and Social Cognitive Theory. The program components developed comprised educational and behavior change sessions, health trackers (eg, monitoring blood pressure, symptoms, and exercise), goal-setting features, and forums for social support. The MyDESMOND digital platform represented an ideal existing platform to host MK&M; thus, the MyDESMOND interface and features were adopted and adapted for MK&M.

Conclusions: Applying the IM framework enabled the systematic application of theory, empirical evidence, and practical perspectives in the codevelopment of MK&M content and materials. Adopting and adapting a preexisting platform provided a cost- and time-efficient approach for developing our digital intervention. In the next stage of work, the efficacy of MK&M in increasing patient activation will be tested in a randomized controlled trial.

(J Med Internet Res 2022;24(11):e39657) doi:10.2196/39657

KEYWORDS

kidney disease awareness; kidney disease knowledge; program development; eHealth; digital health; telehealth; mobile health; mHealth; health promotion; self-management behaviors; mobile phone

Introduction

Background

Chronic kidney disease (CKD) affects approximately 5% to 7% of people living in the United Kingdom [1]. People with CKD have an increased risk of morbidity and mortality, which is largely attributable to cardiovascular disease [2]. Progressive kidney damage can lead to end-stage kidney disease (ESKD), where kidney replacement therapy (ie, dialysis or transplantation) may be needed [3]. However, most people with reduced kidney function never progress to ESKD but, nevertheless, live with significant poor health and high symptom burden [4]. Many individuals with nondialysis CKD are diagnosed and managed by primary care clinicians with infrequent or no input from kidney specialists in secondary care or other members of the multidisciplinary team (eg, specialist or nonspecialist nurses or allied health professionals) [5]. These individuals are mainly required to self-manage their condition, which involves balancing the medical, emotional, and psychosocial consequences of living with CKD and any additional comorbidities with their day-to-day responsibilities [6,7].

The Chronic Care Model has identified self-management as an essential component in chronic disease management to empower patients to take a more active role in their health [8]. Self-management is a complex set of processes and tasks involving the following: (1) use of an individual's knowledge (education), skills, and confidence in managing their disease; (2) the ability to identify and access resources and support; and (3) learning to cope and live with the condition, including the impact on an individual's life and the emotional consequences [9-11]. The goal of effective self-management support is to identify strategies that can help individuals manage their condition(s) while leading full, active, and productive lives. For people with CKD, specific self-management behaviors range from understanding and adhering to medication, health and symptom monitoring, and making lifestyle modifications (eg, increasing physical activity and eating an appropriate diet), which reduce cardiovascular, CKD, and general health risk factors [12].

Effective self-management requires appropriate knowledge, skills, and confidence to manage one's own health and care; this is termed *patient activation* [13]. However, significant deficiencies in CKD awareness and knowledge among people with CKD have been identified [14-16], with many reporting limited or no understanding of their condition and little or no awareness about how to manage their health [17]. Individuals living with CKD, who are managed in primary care, are often unaware of their diagnosis and the detrimental effect that poor risk factor control (such as hypertension and diabetes) may have on their health [18].

Evidence suggests that primary health care providers infrequently discuss CKD with their patients and that the quality of these discussions is limited [19]. Furthermore, self-management support and education through face-to-face interactions in clinical settings are minimal [6,20-22]. For clinicians, barriers to successful education include lack of time and clinical confidence combined with competing priorities

XSL•FO
RenderX

[17]. For people living with CKD, the complex nature of CKD information, limited awareness, health literacy and numeracy combined with the lack of readily available and understandable information and reduced readiness to learn form the greatest challenges [17]. Clearly, innovative approaches to support self-management in people with CKD are urgently required.

Although various self-management interventions have been evaluated, many are not theoretically driven or evidence-based, nor are they informed by patient preferences [22]. Digital health interventions (DHIs) are an efficient and effective method of providing interventions to improve knowledge and self-management behaviors and to actively involve individuals in their care, resulting in improved outcomes for people with long-term conditions [23]. Implementation barriers often associated with face-to-face interventions, such as time and transport, can be addressed using DHIs, which provide more accessible, acceptable, tailored, and interactive information and support [24,25]. When developing evidence- and theory-based DHIs, it is essential that appropriate methods are adopted to ensure effective implementation [26].

Despite DHIs becoming more readily available for people with CKD, few are theory-based, and the development strategies and processes are unclear. Conversely, DHIs in other long-term conditions, such as type 2 diabetes, have appropriately recorded and systematically communicated their development processes. An example is MyDESMOND, which adapted a face-to-face diabetes self-management program (Diabetes Education and Self-Management for Ongoing and Newly Diagnosed, DESMOND) [9] and translated it onto a digital platform [27]. The adaptation and development of MyDESMOND were guided by the intervention mapping (IM) framework.

Objective

The aim of this paper is to provide an overview of the process used in the systematic development of “My Kidneys & Me” (MK&M), a digital self-management program for people living with nondialysis CKD.

Methods

Overview of Predevelopment Work

Review of Previous Research Projects

Before the work described in this paper, we reviewed the experiences and learning from our previous research projects, which were conducted between 2016 and 2018 and involved the development and preliminary evaluation of 2 physical activity behavior change interventions for people with CKD.

Physical Activity Changing Together

Underpinned by the social cognitive theory [28], we previously developed PACT (Physical Activity Changing Together) [29,30], a group-based program aimed at increasing the levels of physical activity in people with nondialysis CKD via structured education (face-to-face sessions comprising 6 modules and 3 booklets) and self-monitoring strategies (pedometer and action planner) [29,30]. In a small uncontrolled evaluation, the PACT program was deemed feasible and acceptable to the participants, and the group-based and interactive delivery was well received by them. However, recruitment was poor (recruitment rate of 23%) [30], suggesting that recruiting to a similar intervention in practice may not be viable. Participants suggested refinements to PACT, which predominantly included broadening the focus to provide a more holistic approach to lifestyle management, including dietary information, emotional management, and other types of exercise.

Self-directed Programme to Increase Physical Activity in CKD

Underpinned by the theory of planned behavior [31], SPARK (Self-directed Programme to Increase Physical Activity in Chronic Kidney Disease) [32,33] was a one-to-one, self-directed physical activity intervention designed to increase physical activity levels in nondialysis CKD [32,33]. SPARK was an 8-week walking and strength training program that was delivered using one-to-one motivational interviewing sessions, with supportive written educational material (two booklets), biweekly telephone calls, and self-monitoring strategies (pedometers and exercise diary). Initial testing, in a small uncontrolled trial, showed that SPARK was feasible and acceptable, with high levels of engagement and completion of the program [33]. Findings suggested positive changes in physical activity behavior, functional ability, symptom burden, disease knowledge, confidence to exercise, patient activation, and quality of life (QoL). However, similar to PACT, recruitment to the program was poor (recruitment rate of 27% from secondary care and 5% from primary care). Although the completion of the physical activity diary was high, interviews with participants highlighted that strength training was perceived as less important and that they needed improved guidance regarding technique.

Learning to Take Forward

Although PACT and SPARK increased disease knowledge and exercise behaviors, there was a clear need for a more holistic approach to support health and lifestyle management, with more guidance on how to engage with and perform desired behaviors. Recruitment to the PACT and SPARK programs was low, and the need to attend the hospital or clinic for face-to-face sessions was cited as a major barrier to participation and engagement. The staff time commitment involved in delivering sessions and follow-up telephone calls also significantly reduced the likely practicality and feasibility of both interventions outside a research setting. Therefore, the overall learning indicated that an alternative method may be needed to deliver the required knowledge and skills, which (1) provides a holistic approach, (2) is likely to be cost-effective, and (3) reduces the number of visits to the clinic or hospital.

Development of MK&M

Overview

There were 2 sequential phases in the development of MK&M:

1. Phase 1: development process of the MK&M resource (content and materials) using IM as a framework (March 2019 to April 2020).

XSL•FO
RenderX

2. Phase 2: adoption and adaptation of the existing MyDESMOND platform to suit the MK&M content and materials (March 2020 to April 2021).

The *Methods* section describes the process and steps of the development work conducted in phases 1 and 2, and the *Results* section presents the details of the needs assessment and theories identified and selected phases 1 and 2.

Steering Group

Prior to embarking on Phase 1, a steering group comprising of kidney researchers from the Leicester Kidney Lifestyle team, representatives of the multidisciplinary kidney health care team (including doctors, nurse, pharmacist, dietician, physiotherapist, psychologist), and experts from the Leicester Diabetes Centre with experience in developing complex interventions (including members of the Innovative Management, Prevention and Care for long term Conditions [IMPACT] team) and digital health (including those who were involved in the development of the digital-based self-management programme for type 2 diabetes, MyDESMOND[27]) was convened. The patient and caregiver perspectives were provided by a patient and public involvement (PPI) group comprising 10 (83%) people with CKD (aged between 33 and 78 years; 7/10, 70% female; and 9/10, 90% White British) and 2 (17%) spouses or relatives. The multidisciplinary steering and PPI groups worked together to identify the needs and priorities for a self-management intervention for people with CKD, develop the program content and materials, and design program features. Both groups met regularly to provide input into the development of MK&M throughout phases 1 and 2.

Phase 1: Codevelopment Process of MK&M Using IM

Overview

IM is a framework designed to systematically develop effective theory- and evidence-based behavior change interventions for real-world implementation [34]. IM was created for, and is widely used in, health promotion but can be applied to any situation in which behavior change is desirable [35]. IM has previously been used in the development of rehabilitation and self-management interventions for people with chronic diseases [36], including nondialysis CKD [37,38], kidney transplantation [39], and type 2 diabetes [27]. This approach provides an iterative, six-step decision-making framework for effective intervention planning, implementation, and evaluation. The creation of MK&M encompassed steps 1 to 4 of the IM process: (1) conduct needs assessment to create a logic model of the problem; (2) identify intervention outcomes, performance objectives, behavioral determinants, and construct matrices of change objectives; (3) select theory-informed intervention methods and practical strategies to deliver change; and (4) intervention development, including the drafting and refining of materials. A subsequent clinical trial (SMILE-K [Self-Management Intervention through Lifestyle Education for Kidney health]; ISRCTN18314195) will address the final 2 stages of IM steps: (5) adoption and implementation of the intervention and (6) evaluation and feasibility testing. The IM approach used was similar to and guided by that used in the development of MyDESMOND [27] but was followed separately and driven and informed by CKD experts and PPI representatives who co-designed MK&M, with advice and support from the MyDESMOND team.

Step 1: Needs Assessment

Overview

The aim of step 1 was to conduct an assessment of the health problem, its related behaviors, and the environmental determinants of the behavior. The needs assessment comprised 4 elements: (1) a literature review of the existing self-management interventions in nondialysis CKD; (2) a previous observational study conducted by our group; (3) a PPI CKD priority-setting workshop; and (4) the creation of "Your Kidneys and Your Health" information booklet as an aid to the review and revision of the concept and content of the nascent resource. The needs assessment was synthesized into a logic model, which outlined the causes of the problem identified and key determinants to target for improvement.

Literature Review of Existing Self-management Interventions in CKD

The evidence discussed in the introduction serves as a justification for the focus on self-management in CKD. A further literature review (unpublished) was subsequently carried out to expand on self-management in CKD and identify the existing CKD-specific self-management interventions and ascertain the (1) components of effective self-management interventions for adults with CKD and (2) patient perspectives of CKD self-management interventions.

Previous Work Conducted by Our Group

During the early part of phase 1, we conducted a mixed methods cross-sectional observational study (DIMENSION-KD [Investigating lifestyle determinants of muscle and physical function, and the impact on patient experience and support needs in kidney disease]; ISRCTN84422148) investigating the lifestyle determinants of physical function and QoL and their impact on patient experience and support needs in CKD. The study involved several parts, including a survey to identify and evaluate lifestyle determinants and factors associated with living with CKD, semistructured interviews with people living with CKD about their health and experiences of living with a kidney condition, and semistructured interviews with health care professionals (HCPs) about the role of self-management and lifestyle interventions for people living with CKD. To participate in the study, individuals had to be (1) diagnosed with kidney disease, (2) male or female, (3) aged ≥18 years, and (4) willing and able to give informed consent and comply with the study protocol. Individuals were excluded if they (1) were aged <18 years; (2) had any other significant disease or disorder that, in the opinion of the patient's own clinician, may either put the participants at risk because of participation in the study or may influence the result of the study or the participant's ability to participate in the study; or (3) had an inability to give informed consent or comply with the protocol for any reason. HCPs were defined as any clinician who cares for patients or professionals on the "Health and Care Professions Council" (eg, physician, general practitioner [GP], practice nurse, dietician, physiotherapist, and occupational therapist).

XSL•FO
RenderX

A PPI CKD Priority Workshop

A research priority–setting workshop relating to healthy lifestyle, physical activity, and diet and weight management was conducted to identify topics that were important to people with nondialysis CKD. Participants with stages 3 and 4 nondialysis CKD were invited from the Leicester Kidney Lifestyle Team PPI contact list and included people with CKD who were managed in primary care and their relatives. Participants attended a half-day workshop addressing 3 broad topics: (1) healthy lifestyle, (2) physical activity, and (3) diet and weight management. For each topic, the participants were divided into 2 facilitated discussion groups, and the points arising were then reviewed and evaluated among the whole group to formulate indicative research priority questions. Finally, the participants independently ranked their top three research priorities based on these questions.

"Your Kidneys and Your Health" Information Booklet

Moreover, 3 workshops, separate from the priority-setting workshop, were conducted to develop a prototype educational resource to provide a basis for subsequent review and refinement of the content, design, and delivery format. Similar to the priority-setting workshop, participants with stage 3 and 4 nondialysis CKD were invited from the Leicester Kidney Lifestyle Team PPI contact list and included people with CKD who were managed in primary care and their relatives. The workshops were interactive, with brainstorming discussions about content and style. For practical reasons, the initial resource was coproduced in a booklet form to provide a physical document that could be easily examined and refined in real time. The participants annotated the booklet, while graphic designers from the Leicester Biomedical Research Centre updated the document "on the spot" based on the feedback, producing new versions for further comments. Following this process, the booklet was printed, and copies were sent to all PPI members. We also formed a collaborative relationship with a local GP practice who distributed the booklet and collated the feedback. The booklet was developed as an aid to understand the concept and content of the nascent resource and thus was not intended to be the final intervention but to facilitate the early development of the educational program (MK&M).

Step 2: Identification of Outcomes, Performance Objectives, and Change Objectives

The next step involved specifying what needed to change in order to achieve the desired outcomes of the intervention. The overall desired outcome was to improve the knowledge, skills, and confidence of people with CKD in managing their health (termed *patient activation*) and consequent self-management behaviors.

The determinants (factors that could be expected to influence behavior) of each desired outcome were identified and relevant performance objectives (factors associated with the expected behaviors required to achieve the program goal) were developed. Combining the determinants and performance objectives, we produced a list of change objectives that refer to what needs to change to affect the performance objective and ultimately the program outcome. For example, for individuals to increase their access to social support and resources, they need to be aware of the support available, understand which support and resources are beneficial to their own personal situation, and have the confidence to access the relevant support and resources. The output of this process is a matrix of change objectives detailing what will be targeted in the intervention. To illustrate the proposed relationships between change methods, the determinants they were expected to influence, and the behavioral and environmental outcomes that would address the problem, we constructed a logic model of change.

Step 3: Selection of Theory-Informed Intervention Methods and Practical Strategies

Step 3 included the selection of theoretical methods considered to influence changes in the determinants in the desired direction. This step involved the review of theoretical methods and their translation into practical applications matching the health behavior change of the individual. The criteria for choosing theories to inform interventions, suggested by Hardeman and Sutton [40], were used to help guide the selection of theories. The criteria are as follows: (1) use in interventions aimed at similar target behaviors; (2) applicability to the target group; (3) clear definition of causal, testable pathways between behavioral determinants and behavior; (4) strength of evidence about predictive validity; and (5) clear guidelines for measurement [40]. Appropriate behavior change theories were identified from the literature and expert steering group. These were then translated into strategies for practical application.

Step 4: Development of the Intervention Program

The aim of step 4 was to combine the learning from the previous steps into the program and to develop working documents to guide the intervention production [34]. The steering group collaborated with the PPI group to provide guidance on the scope and content of the intervention and the most appropriate format of delivery. Program materials were codeveloped with the PPI group, and feedback was elicited. A description of the scope, content, and delivery of the program was outlined.

Phase 2: Adoption and Adaptation of the MyDESMOND Platform to Suit MK&M Content and Materials

The MyDESMOND platform interface and features [27] were adopted for MK&M (justification is described in the *Results* section). The developed MK&M materials and content (presented in the *Results* section) were mapped onto the existing MyDESMOND platform [27]. Key components of the intervention were selected based on the feasibility of adopting and adapting the existing MyDESMOND platform to suit MK&M content and materials within the resource constraints.

Ethics Approval

For the priority-setting workshop and the booklet workshops, the PPI members provided written informed consent. DIMENSION-KD was granted national research ethical approval (18/EM/0117). All patients provided informed written consent, and the study was conducted in accordance with the Declaration of Helsinki. The SMILE-K study was fully approved by the Research Ethics Committee-Leicester South

XSL•FO
RenderX

(17/EM/0357). All participants are required to provide informed consent online.

Results

Phase 1: Development Process of MK&M Using IM

Step 1: Needs Assessment

Literature Review

The self-management of a chronic illness or disease involves 3 sets of core tasks, namely, medical, behavioral, and emotional management [41], and encompasses 5 key processes, namely, problem-solving, decision-making, resource utilization, forming partnerships, and action planning [42]. Increased awareness and improved decision-making skills enable patients to be more engaged in the management of their condition [43]. Clinical management guidelines for CKD consider the promotion of self-management behaviors as a standard of care, with recommendations to incorporate information, advice, and education to support self-management behaviors into treatment plans at all stages of CKD [44] to potentially slow the disease progression and prevent complications [45]. Despite recommendations to support CKD self-management, many individuals living with CKD do not perceive the need to self-manage their CKD and lack an understanding of the importance of self-management [17,18]. To help people with CKD self-manage, it is widely recommended that individuals are aware of their diagnosis; involved in shared treatment decisions; provided access to their medical data; and given information on blood pressure control, exercise, diet, medication management, and smoking cessation [46]. However, the unique and complex nature of CKD (ie, asymptomatic and multimorbidity) may require distinct self-management support that may differ from that in other long-term conditions [47].

Previous systematic reviews [48-51] and integrative reviews [7,21,22] of self-management interventions for people living with CKD have shown improvements in self-care activities. However, they are variable in their effectiveness in managing and preventing the progression of CKD. A recent review highlighted a wide range of self-management interventions for adults with CKD (not requiring kidney replacement therapy) with considerable heterogeneity in outcomes [6]. The wide variations in the content, delivery, and components of digital self-management interventions (eg, self-monitoring and education), alongside the outcomes assessed and results obtained, make it difficult to compare findings across studies evaluating self-management programs [52]. Consequently, there is a paucity of knowledge about which components of these programs are the most effective [53].

Despite the importance of theory-driven interventions [54], efforts to improve CKD self-management have been largely atheoretical [22]. Major gaps in the literature include the lack of patient involvement and engagement in intervention design (an estimated <1% of programs are codeveloped with patients) and failure to apply behavior change theory (only 20% are based on a theory or framework) [6]. Person-centeredness, applicability to comorbidities associated with CKD, and physiological and nonphysiological outcomes are also often lacking from self-management interventions [6].

Meaningful information relating to the patient perspective of self-management interventions (eg, desired support and delivery of this support) is lacking from the literature but is key to providing valuable information regarding attitudes towards and challenges associated with self-management and self-management interventions [7]. Furthermore, patients have highlighted that interventions need to be individualized and tailored to their specific situations and preferences (eg, awareness of having CKD, stage of CKD, knowledge of the disease, and access to resources [6]). Engaging patients by involving them in the co-design of such interventions will ensure that their values, culture, and psychosocial needs are addressed in the intervention [7,21,22].

Findings From a Mixed Methods Observational Study Exploring Lifestyle Determinants in CKD (DIMENSION-KD)

Our survey findings showed that of 743 people living with nondialysis CKD (mean eGFR 32.3, SD 17.1 mL/min/1.73 m^2; mean age 67.8, SD 13.9 years; 503/743, 68% male), only a minority of individuals were activated for self-management, with 60% (444/743) reporting low activation (knowledge, skills, and confidence in managing own health) [55]. Patient activation declined concomitantly with disease progression, and low activation was associated with being older, having a comorbidity, and lower hemoglobin levels. Individuals with low activation had poor cardiorespiratory fitness and health-related QoL and more cardiovascular disease risk factors [55].

Findings (unpublished) from semistructured interviews with 22 people living with CKD (mean eGFR 42.2, SD 14.3 mL/min/1.73 m^2; mean age 71.4, range 48-88 years; 11/22, 50% male) highlighted a need for education and support for people living with CKD (not requiring kidney replacement therapy), particularly during the initial period following their diagnosis. The participants described a desire to understand CKD and its impact on overall health (including disease progression, health risk factors, and comorbidities) and the role self-management and lifestyle can play in living well with CKD. Information about medication, symptoms, lifestyle behaviors (eg, diet and physical activity) and mental health and well-being and strategies to help manage these and make relevant changes were thought to be key to enabling people to manage their CKD. Peer and familial support, goal setting, coping strategies, and self-monitoring tools were considered as motivators to facilitate health and lifestyle changes. In addition, HCPs were interviewed about the role of self-management and lifestyle interventions for people living with CKD. The findings highlight the need to increase patient awareness of CKD and self-management in primary care. There was a belief that individuals need to be empowered to take responsibility to manage their health and engage in healthy lifestyle behaviors (eg, physical activity and weight management).

XSL•FO
RenderX

CKD Priority-Setting Workshop Findings

A total of 12 participants (people with CKD: n=9, 75%; and family members: n=3, 25%) attended our priority-setting workshop. The dominant recurring theme was a lack of appropriate information from primary care providers. This not only extended to lifestyle, physical activity, and diet but also to kidney function and CKD itself. Many individuals had little knowledge of the roles of the kidney; the causes and nature of CKD as a condition; or its impact, consequences, and relevance to wider health. Both psychological and social well-being were deemed as important as physical health. Findings highlighted a need to focus on the most appropriate means of disseminating knowledge regarding CKD, lifestyle, diet, and exercise to people living with CKD. The top 3 research priorities were (1) How can we improve GP communication? (2) What can I eat and drink with CKD? and (3) What diet will help me maintain physical activity [56]?

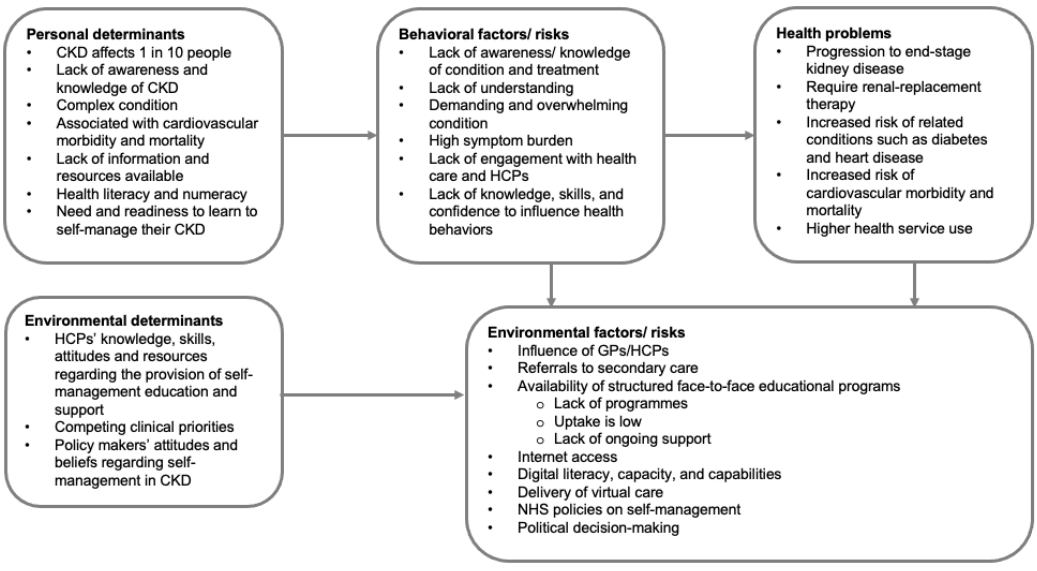
Findings From the Development and Evaluation of "Your Kidneys and Your Health" Information Booklet

An education booklet called "Your Kidneys and Your Health" was codeveloped by our team in collaboration with the PPI group (10 people with CKD and 3 family members). Feedback from other people living with CKD, who were given the booklet by their GP, suggested that the most useful aspects were information relating to CKD stages, kidney function, diet, and symptoms associated with CKD. The participants also reported that the content was easy to read and understand. Although the education booklet provided a basic holistic approach to lifestyle management in CKD and represented a convenient, low-cost alternative to delivery of face-to-face education, it was limited in its ability to provide interactive guidance and support and to engage participants in positive behavior change. It was determined that another method (eg, digital) would be required to deliver CKD education and lifestyle information. However, the booklet provided a useful physical resource to share with the PPI and steering groups to aid discussion about the content and design of the larger self-management program.

Summary of Needs Assessment Findings

The findings from the different elements of the needs assessment, alongside learning from previous interventions and resources, were collated to establish a rationale for a self-management intervention for people with CKD. The needs assessment highlighted that individuals with CKD have multiple needs that can differ based on the complexity of their illness, health knowledge, and confidence to manage the disease. Having explored the issues of CKD self-management, a logic model was created to better understand the problem, which illustrates in detail the issues under investigation and the associated relationships and factors (Figure 1). It was agreed that the goal was to co-design (meaningful stakeholder involvement in the planning and development of the intervention to ensure that it meets their needs and is usable [57,58]) a self-management intervention that could be offered to any individual with nondialysis CKD. On the basis of the evidence discussed in the introduction, needs assessment, and experience of the steering group, it was determined that a potential self-management program would be best delivered digitally to enable accessible, tailored, and interactive information and support and to overcome the barriers experienced in the delivery of PACT, SPARK, and the "Your Kidneys and Your Health" booklet. The PPI steering group named the proposed self-management program MK&M.

Figure 1. Logic model of problem. CKD: chronic kidney disease; HCP: health care professional; NHS: National Health Service; GP: general practitioner.



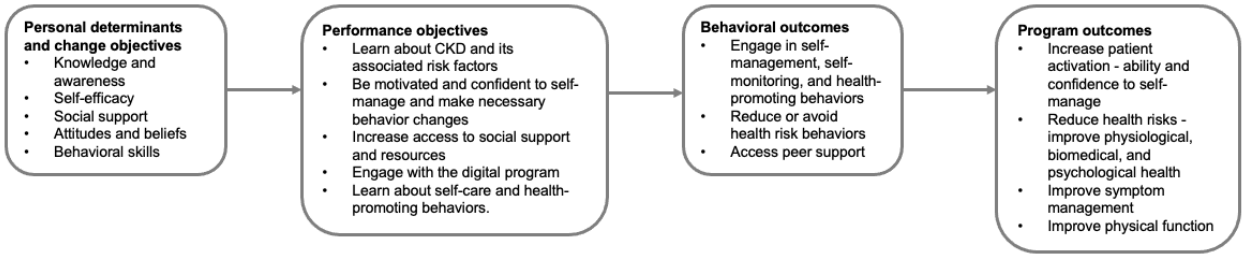
XSL•FO
RenderX

Step 2: Identification of Outcomes, Performance Objectives, and Change Objectives

In step 2, the findings from step 1 were used to specify what would need to be changed for people to successfully self-manage their CKD. The findings from the needs assessment indicated that people with CKD would benefit from education about the condition and lifestyle behaviors as early as possible following diagnosis. In addition to containing information to improve knowledge, the intervention should also aim to improve and maintain self-management behaviors, promote self-care skills, increase self-efficacy, and improve well-being. Thus, the program goals were to improve the knowledge, skills, and confidence of people with CKD in managing their health and consequent self-management behaviors. The key outcomes were to (1) increase patient activation (ie, knowledge, skills, and confidence); (2) reduce health risks; (3) manage symptoms; and (4) increase physical function.

The behavioral determinants were (1) knowledge and awareness, (2) self-efficacy, (3) social support, (4) attitudes and beliefs (intention), and (5) behavioral skills. The performance objectives were (1) learn about CKD and its associated risk factors, (2) be motivated and confident to self-manage and make necessary behavior changes, (3) increase access to social support and resources, (4) engage with the digital program, and (5) learn about self-care and health-promoting behaviors. A matrix of change objectives detailing the behavioral determinants required to influence the performance objectives, and subsequently self-management behavior, in people with CKD was created. The change objectives required to achieve the performance objectives are detailed in Table 1. A logic model of change illustrating the proposed relationships between change methods, the determinants they were expected to influence, and the behavioral and environmental outcomes that would address the problem are presented in Figure 2.

Figure 2. Logic model of change. CKD: chronic kidney disease.



XSL•FO
RenderX

Table 1. Change objectives.

Performance objectives	Personal determinants				External determinants
	Knowledge and awareness	Self-efficacy	Attitudes and beliefs (intention)	Behavioral skills	Social support
Learn about CKD[a] and its associated risk factors	• Awareness and understanding about CKD, its treatment options, and its symptoms and associated health risks and conditions	• Demonstrates ability to process information about CKD and self-management • Expresses confidence in applying the information learnt	• Recognizes that CKD is a long-term condition that requires patients to take an active role in looking after their health and reducing health risks	• Be able to identify own personal health risk factors and methods to modify and/or reduce health risk factors and problems	N/A[b]
Be motivated and confident to self-manage and make necessary behavior changes	• Understand the importance of self-managing and identify which factors and behaviors can be modified	• Be motivated and confident to self-manage and make necessary behavior changes	• Belief that there will be benefits in making necessary behavior changes to their health and lifestyle	• Develop routines that assist in managing their health • Develop habits that enhance self-management	N/A
Increase access to social support and resources	• Increase awareness of the support available and knowledge of which support and resources are beneficial to own personal situation	• Have the confidence to access the relevant support and resources	• N/A	• N/A	Accesses and receives support from both HCPs[c] and others with CKD
Willingness to engage with the program and accept more responsibility for managing their health and health care	• Learn about the importance and benefits of looking after their health and health care	• Express confidence in managing their health and health care • Express confidence in taking action to help prevent or minimize symptoms or health problems	• Increase the motivation to use the program • Expresses willingness to take greater responsibility toward managing their CKD, health, and health care	• Be able to find a solution when new situations or health problems arise • Be able to confidently recognize own limitations and seek necessary help when required	Engage with other program users
Learn about and practice self-care and health-promoting behaviors	• Awareness and understanding of health-promoting behaviors	• Express confidence in self-managing and performing health-promoting behaviors	• Recognizes and believes that practicing self-management behaviors will enhance health and well-being • Expected to reduce health risks and engage in health-promoting behaviors	• Be able to confidently perform and maintain self-care practices and health-promoting behaviors, including lifestyle changes • Demonstrates model behaviors	N/A

[a]CKD: chronic kidney disease.

[b]N/A: not applicable.

[c]HCP: health care professional.

XSL•FO
RenderX

Step 3: Selection of Theory-Informed Intervention Methods and Practical Strategies

In developing MK&M, we were able to draw upon multiple theories that consider the complexity of behavior change, which are commonly used in the development of health behavior change interventions and have been extensively used in chronic disease self-management interventions [59], including those designed to improve self-management behaviors, and thus provide an effective framework for self-management support in people with CKD [60]. The theories and frameworks selected (self-management framework; Capability, Opportunity, Motivation Behavior model components of Behaviour Change Wheel and taxonomy of behavior change techniques; health action process approach model; common sense model; and social cognitive theory) are described in detail, including the justification for inclusion, in Multimedia Appendix 1 [61-80]. A theoretical framework was constructed to map the relevant mechanisms of action (processes that influence behaviors), behavior change techniques, and application of practical strategies to facilitate behavior change in MK&M and is illustrated in Table 2.

XSL•FO
RenderX

Table 2. My Kidneys & Me theoretical framework, target determinants, mechanisms of action (MoA), behavior change techniques (BCTs), and practical applications.

Guiding principle	Underlying theory and constructs	Target determinants[a]	MoA	BCTs	Methods and practical applications
Provide patients with education about CKD[b] and promote self-management	• COM-B[c]: psychological capability • HAPA[d]: risk awareness and outcome expectancies • CSM[e]: interpretation of the problem • SCT[f]: outcome expectations, behavioral capability, and self-regulation	• Environment: address the perceived lack of support in providing clear disease-based information	Knowledge	• Information about the condition, including causes, symptoms, treatment, and consequences • Biofeedback • Feedback on behavior	• Educational material, including videos, providing information about the kidneys; chronic nature of CKD; possible causes and consequences of CKD; conditions associated with CKD; health risks; medication; health care appointments; types of treatments; symptoms; and healthy lifestyle, including diet and physical activity • Interactive quizzes to test knowledge and learning • "How to" sessions proving guidance on how to perform self-care and health-promoting behaviors and strategies to encourage behavior change and the adoption of new behaviors • Trackers to monitor health and activity and to provide feedback on behavior outcomes
Provide patients with education about CKD[b] and promote self-management	• COM- B: psychological capability • HAPA: risk awareness and outcome expectancies • CSM: interpretation of the problem • SCT: outcome expectancy, behavioral capability, and self-regulation	• Behavioral: address outcome expectations (risks and benefits) of engaging in self-management behaviors and other healthy lifestyles	Behavioral regulation	• Information on the consequences of behavior • Instruction on how to perform behavior • Modeling or demonstrating behavior • Focus on past success • Feedback on behavior • Biofeedback • Self-monitoring of behavior	• Educational material, including videos, providing information about health risks associated with CKD and the concept of personal responsibility for health focusing on modifiable CVD[g] risk factors • Interactive quizzes to test knowledge and learning • "How to" sessions encouraging users to engage with the self-monitoring trackers and to complete an action plan for desired behaviors • Trackers (health trackers, symptom tracker, and exercise trackers [step count and strength training]) are available for users to record their physical activity, including monitoring steps, strength training, weight, dietary intake, and symptoms

XSL•FO
RenderX

Guiding principle	Underlying theory and constructs	Target determinants[a]	MoA	BCTs	Methods and practical applications
Provide guidance and opportunity to self-monitor health and self-management behaviors	• COM-B: physical capability • HAPA: risk awareness, outcome expectancies, self-efficacy, planning, and action control • CSM: action plan, coping and appraisal • SCT: outcome expectancy, behavioral capability, self-efficacy, and self-regulation	• Environmental: address the lack of support in providing information regarding self-management and healthy lifestyle behaviors • Behavioral: address the benefits of engaging in health-promoting behaviors and how they can help to support symptom management, for example, fatigue, and help individuals stay independent, re-enforcing innate desires and outcome expectancies for self-management • Personal: Guidelines and advice on what is safe and appropriate may help to reduce fears related to recognizing performance limitations	Skill	• Information on desired behaviors • Instruction on how to perform behavior • Behavioral rehearsal/ practice • Coping strategies • Habit formation	• Educational material provides information about healthy lifestyle, physical activity, and well-being. Information on benefits and consequences of behavior is highlighted • "How to" sessions provide information on how to perform health-promoting behaviors and helpful strategies to encourage adoption and maintenance of health-promoting behaviors. Instructions on how to set SMART[h] goals and create personalized action plans • Trackers can encourage users to track and self-monitor their health, symptoms, and exercise behaviors • Chat forum enables users to share their experiences and strategies with others
Provide advice about how to self-manage through a healthy lifestyle, goal setting (set health and behavior-related goals), self-monitoring health, and self-management behaviors	• COM-B: reflective motivation • HAPA: outcome expectancies, self-efficacy, planning, and action control • CSM: action plan, coping and appraisal • SCT: self-efficacy, behavior capability, observational learning, and self-regulation	• Behavioral: address outcome expectations (risks and benefits) of engaging in self-management and health-promoting behaviors	Beliefs about capabilities	• Information and examples of desired behaviors • Verbal persuasion to encourage the adoption of new behaviors • Verbal persuasion to reduce self-doubts and enhance self-efficacy • Focus on past successes • Habit formation	• Videos highlighting important messages are presented by experts to encourage users to set and review personalized goals, create and continue their action plans, and engage in changes of health and lifestyle behaviors • Educational material addressing knowledge about health-promoting behaviors, factors related to health concerns, and physiological responses, including those to exercise • Chat forum enables verbal reassurance and encouragement among users that may increase their self-efficacy
Provide advice about how to self-manage through a healthy lifestyle, goal setting (set health and behavior-related goals), self-monitoring health, and self-management behaviors	• COM-B: reflective motivation • HAPA: outcome expectancies, self-efficacy, planning, and action control • CSM: action plan, coping and appraisal • SCT: self-efficacy, behavior capability, observational learning, and self-regulation		Motivations and goals	• Information on the consequences of behaviors • Information on goal setting, reviewing goals, action planning, problem-solving, self-monitoring	

XSL•FO
RenderX

Guiding principle	Underlying theory and constructs	Target determinants[a]	MoA	BCTs	Methods and practical applications
		• Environmental: the lack of guidance on how to set health- and behavior-related goals • Behavioral: Target self-management and health motivations by including self-regulation strategies • Personal: Target the innate drives for self-management by asking participants to consider why self-management and health-promoting behaviors may be important to them and how it would specifically benefit them in their lives			• Educational material, including videos, provides information and guidance on how to set personalized goals, review goals, create action plans, and evaluate progress. Information was also provided on barrier identification and how to address/ overcome potential barriers • Users are encouraged to set goals and complete an action plan for target behaviors, considering any potential barriers and strategies to overcome these • Access to goal-setting functions enables users to set health-related and exercise goals
Provide advice about how to self-manage through a healthy lifestyle, goal setting (set health and behavior-related goals), self-monitoring health, and self-management behaviors	• COM-B: reflective motivation • HAPA: outcome expectancies, self-efficacy, planning, and action control • CSM: action plan, coping and appraisal • SCT: self-efficacy, behavior capability, observational learning, and self-regulation	• Behavioral: address the perceptions of living with CKD, expectations of self-managing the condition, and importance of engaging in health-promoting behaviors	Role and identity	• Information on desired behaviors • Verbal persuasion to encourage the adoption of new behaviors • Message framing/reframing	• Videos presented by experts provide key information and important messages about CKD, associated health risks, symptoms, health care appointments, physical activity and exercise, diet, well-being, sleep, and other relevant information
Provide advice about how to manage the emotional consequences of CKD and the opportunity to share experiences	• COM-B: autonomic motivation • SCT: outcome expectancy, collective efficacy, and self-regulation	• Environment: addresses the need for emotional support • Personal: enable participants to share their experiences of living with CKD	Emotions	• Information on appropriate support and resources • Reduce negative emotions • Coping strategies • Social support • Social comparison	• Educational material provides information around well-being, including how to cope and live with CKD. Videos are presented by experts on managing emotions and accessing support • Chat forum encourages users to talk to others, including HCPs, family and friends, and other people with CKD. It also enables users to provide and receive peer support
Provide the opportunity to share positive self-management experiences and advice from peers	• COM-B: social opportunity • SCT: outcome expectancy and collective efficacy	• Environment: addresses the need for peer support and facilitate peer learning • Personal: enable participants to share positive experiences of engaging in self-care and health-promoting behaviors and seek advice from peers	Social influences	• Social support • Social comparison • Encouragement	• Chat forum enables users to discuss and share their thoughts and experiences, providing encouragement for each other. It also enables users to seek and provide approval from others

XSL•FO
RenderX

[a]Target determinants include personal, behavioral, and environmental influences identified from previous work to target within the intervention.

[b]CKD: chronic kidney disease.

[c]COM-B: Capability, Opportunity, Motivation Behavior model.

[d]HAPA: health action process approach model.

[e]CSM: common sense model.

[f]SCT: social cognitive theory.

[g]CVD: cardiovascular disease.

[h]SMART: Smart, Measurable, Achievable, Realistic, Timely.

Step 4: Development of the Intervention Program

Defining the Content

The overall requirements of MK&M were derived from the needs assessment and theory described earlier. Discussions with stakeholders, including people living with CKD, and findings from qualitative interviews and PPI workshops highlighted an interest in digital interactive information, including videos, as a way to understand CKD self-management. We codeveloped content and material that was designed to improve kidney-specific self-management practices, including increasing CKD knowledge, reducing health risks, managing symptoms, and improving physical function. The content was reviewed and refined by a wide range of CKD experts representing the kidney multidisciplinary team, including specialist kidney clinicians, nurses, dieticians, pharmacists, physiotherapists, and psychologists, and the PPI group. Educational "Learn About" sessions were developed to provide information about kidneys; CKD; its treatment, symptoms, and associated health risks; and lifestyle-related factors (eg, diet and physical activity [including strength training]). Behavior change–focused "How To" sessions were designed to provide practical strategies to assist individuals with making small modifications to improve their health and lifestyle behaviors. The "How To" sessions also contained information about how to set goals, self-monitor health and behaviors, and create action plans, as goal setting is a motivator to facilitate health and lifestyle behavior changes. Moreover, enabling individuals to set, review, and adapt their own goals is important [49,81]; therefore, goal-setting features were included to support people in selecting and self-monitoring their own goals. Good physical function is important for patients with CKD [82] but is often overlooked and underemphasized in exercise guidance. Therefore, based on the findings from the SPARK evaluation and feedback, we developed a separate strength training educational session with strength exercise programs, videos demonstrating the exercises, and a bespoke strength tracker.

The development of the content for MK&M was an iterative process, with 12 PPI members providing regular review and input, such as suggesting (1) a symptom tracker to enable self-monitoring of symptoms; (2) the inclusion of new sessions on sleep and well-being, emphasizing the importance of looking after mental health, including coping and living with CKD; (3) the use of "myth and fact" quizzes as a way to test knowledge; and (4) stating how long sessions should take to complete. These suggestions resulted in the following actions: (1) designing a bespoke symptom tracker, (2) creating both educational and "how to" sessions on improving sleep quality and looking after well-being, (3) adding quizzes to both "learn about" and "how to" sessions, and (4) including the anticipated time it would take people to work through and complete each session (based on the mean time of completion for PPI group). Content topics included are detailed in Multimedia Appendix 2.

Defining the Format of Delivery

Although the content and materials of MK&M were developed with a future digital platform in mind (based on previous learning, needs assessment, and PPI feedback), before this point, the exact web platform had not yet been decided. Following discussions about the design and format of the web platform for MK&M, it was agreed among the steering group that the MyDESMOND platform represented an ideal existing platform (reasons detailed in the subsequent section) to host MK&M; therefore, the final stage of development involved the required adaptation of the MyDESMOND interface and features [27] for MK&M content and materials.

Phase 2: Adoption and Adaptation of the Existing MyDESMOND Platform to Suit MK&M Content and Materials

Given the successful development and implementation of the MyDESMOND program [27,83,84], it was advantageous to use the preexisting and established MyDESMOND platform to host the MK&M program. MyDESMOND is an award-winning program on a quality-assured platform, which is known to be well accepted and effective [83]. The MyDESMOND platform was built using progressive web apps and is accessible on any digital device (tablet, desktop computer, and smartphone) and operating system. The system was developed using the Zend framework and MySQL database. The app includes an app programming interface built with Zend, which is used to manage data between the AngularJS progressive web app and the database. The MyDESMOND platform hosts several evidence-based web-based self-management programs, including programs for people at risk of or with type 2 diabetes. The existing MyDESMOND programs are used throughout the United Kingdom and Australia. These MyDESMOND programs have gained accreditation from Quality Institute for Self-Management Education and Training. The Quality Institute for Self-Management Education and Training (QISMET) defines good practice in self-management education [85]. MyDESMOND has been successfully reviewed by Organisation for the Review of Care and Health Apps (ORCHA) [86], an independent leading digital health reviewer that assesses digital health programs against the latest standards and guidelines that cover clinical/ professional assurance, data and privacy, and usability and accessibility. MyDESMOND also conforms to the Digital Technology Assessment Criteria (DTAC) for health and social care. This is a new national baseline criterion used

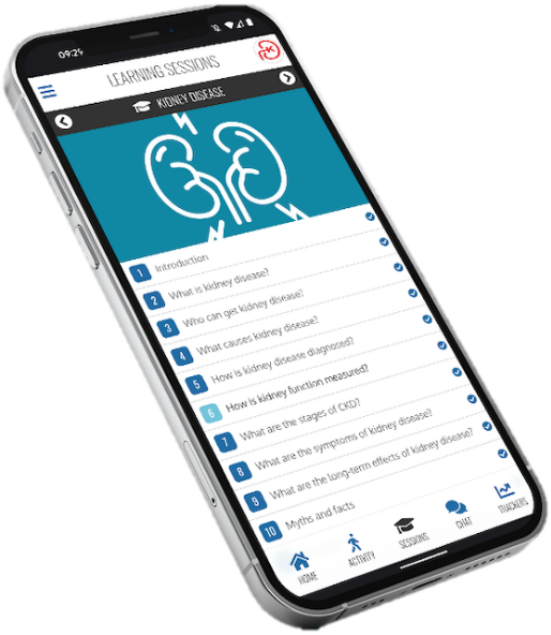
XSL•FO
RenderX

to ensure that the latest standards within digital health systems are being met. It covers clinical safety, data protection, technical security, interoperability, and usability and accessibility standards and is used for the commissioning of digital health technologies [87]. The MyDESMOND platform is flexible to easily allow for additional programs to be added and adapted. Adding the MK&M content and materials to the MyDESMOND platform [27] and adapting the existing platform to suit the program needs provided a cost- and time-efficient approach for developing our digital intervention. In addition, the MyDESMOND platform meets many National Health Service (NHS) requirements and quality standards, which would enable the future implementation of MK&M in clinical practice.

We mapped the developed MK&M content as closely as possible to the preexisting features already available on the MyDESMOND platform [27]. For example, we mapped the "Learn About" and "How To" sessions onto the preexisting "Learning" and "Booster" sessions on the platform. Figures 3 and 4 present examples of the sessions available. The MK&M goal-setting materials were mapped onto the MyDESMOND "Decision Maker" feature. We adopted the health (eg, blood pressure, shape, and healthy eating) and physical activity (eg, step counting) trackers that were already available on the MyDESMOND platform [27], as these were perceived to adequately meet the needs of people with CKD.

Figure 3. Example of learning sessions available on My Kidneys & Me.

Figure 4. Example of the "Kidney Disease" learning sessions available on My Kidneys & Me.



XSL•FO
RenderX

Although we were able to adopt the interface and features of the MyDESMOND platform [27] and map most of the MK&M content and materials onto it, there were some key MK&M components that were missing. Thus, we further adapted the platform to suit our needs. This involved creating and realizing a bespoke CKD symptom (detailing 13 common symptoms identified by people with CKD not receiving kidney replacement therapy [88]) (Figure 5) and a bespoke strength tracker, which were conceptualized during the codevelopment of materials and content in phase 1. Links to strength training materials, including flashcards and demonstration videos, were embedded within MK&M. To aid individuals' understanding of the program and navigation through the platform, we also created a welcome video detailing the aims of the program and a screencast (ie, digital video recording of our computer screen, including audio narration) explaining how to navigate the program and the key features available. The finalized MK&M materials and content were uploaded onto the platform by the MyDESMOND digital team. Finally, we added an MK&M-specific launch page to the MyDESMOND platform, as shown in Figure 6. Usability (functionality, navigation, and interactivity) and experience testing of MK&M was conducted through think-aloud interviews to identify potential areas for refinement (manuscript in preparation).

Figure 5. Chronic kidney disease symptom tracker on My Kidneys & Me.

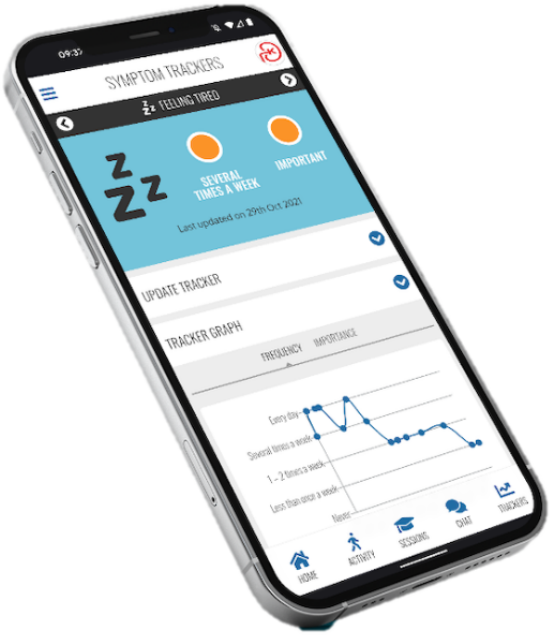


Figure 6. Example of My Kidneys & Me launch page and the digital devices it is accessible on.

XSL•FO
RenderX

Discussion

Principal Findings

Self-management has been identified as a top priority by individuals with nondialysis CKD, their caregivers, clinicians, and policy makers [89]. The concept of patient empowerment is integrated in the "NHS Long Term Plan," which aims to support people with long-term conditions to improve self-management skills and provide access to resources [90]. Here, we have described the codevelopment of a theory- and evidence-based digital self-management program for people with CKD, guided by the IM framework and the MyDESMOND development process [27].

Steering Group

The formation, participation, and engagement of the multidisciplinary planning group and PPI group were valuable as it enabled knowledge, experiences, and visions to be shared. The development process was guided by the inputs from HCPs and the PPI group, with both groups reviewing and refining the program content and materials and the method of delivery alongside the researchers. This ensured that the program was appropriately designed and relevant to people with CKD and that people with CKD and HCPs support MK&M. Having input from the MyDESMOND and IMPACT teams greatly facilitated the development of MK&M, as learning first-hand from their experiences in developing MyDESMOND [27], and adapting their processes, was invaluable. In addition, the ability to map our own program onto the MyDESMOND platform was advantageous.

Applying Theoretical Framework

The IM process helped the rigorous development of MK&M and ensured that we explored our program goals in depth and were able to specify what it was that we needed to change and which method and practical applications would elicit the desired change. This ensured that MK&M, similar to MyDESMOND, was underpinned by a strong theoretical framework and evidence-based foundation. Using IM enabled a meaningful analysis of the underlying mechanisms that are hypothesized to affect the desired intervention outcomes by enabling the explicit linking of interventional components to theory, which should result in improved outcomes for the target population [91]. The use of behavior change theories and techniques to determine the anticipated behavior change ensured a pragmatic application of change methods.

Practical Implications

The IM framework provides a comprehensive approach that clearly articulates the development process [34], which allows for more transparency in the design of the intervention to others outside the development team. The process may be different for different interventions, and each step of the framework allows for the adaptation of the different requirements of the specific population, setting, and nature of the intervention [34]. Using IM enabled us to produce an intervention that is potentially efficacious and acceptable and can be implemented in CKD. The needs assessment conducted in step 1, alongside lessons learnt from our previous work, provided us with a clear understanding of the different key components that needed to be incorporated into MK&M. Identifying the desired outcomes of MK&M ensured that appropriate content and materials were designed to address our program goals.

Aside from using IM to develop interventions, the approach may aid clinicians in assessing the suitability of an intervention, its applicability to their setting and population, or what change may need to occur to the intervention to enhance its suitability [92]. The dissemination of the development of complex interventions offers clinicians more detailed information on which they can base their decision regarding the implementation of these interventions [92].

Comparison With Other Interventions

To our knowledge, this is the first systematically developed theory- and evidence-based digital self-management program for people with nondialysis CKD in the United Kingdom. Another digital intervention for people with CKD, called Kidney Beam, has been developed in the United Kingdom to provide support for exercise, physical activity, and emotional well-being [93]. Similar to MK&M, Kidney Beam is theory informed and guided by the Behaviour Change Wheel. The full development process, application of theory, and selection of BCTs have yet to be published. Unlike MK&M, which provides a holistic approach to self-management, encompassing the core tasks alongside the key processes, Kidney Beam aims to specifically address exercise, physical activity, and emotional well-being. Results of the feasibility study were promising and showed good levels of engagement, with 43% of individuals signing up to the platform and completing at least one exercise class [93]. Similar to MK&M, Kidney Beam is currently being evaluated in a randomized controlled trial to evaluate the feasibility, clinical value, and cost-effectiveness of the Kidney Beam program delivered as part of clinical care [93]. It is to be hoped that MK&M and Kidney Beam will provide complementary options for future kidney care.

Digital self-management programs for people with CKD have also been developed in other countries such as Canada [94] and the United States [95]. There are similarities and differences between these programs. Like us, Donald et al [94] and Markossian et al [95] co-designed their program with stakeholders, involving people living with CKD and their relatives, clinicians, researchers, and graphic designers. Although there are similarities in methodologies, both MK&M and "My Kidneys My Health" [94] developed a web-based self-management resource, whereas Markossian et al [95] developed a mobile app. In our program, we used the IM framework to guide the development, whereas Donald et al [94] used the "Knowledge-to-Action" framework to guide the multiphased activities for determining CKD patient self-management support intervention. It is not clear what framework was used to guide the development of the mobile app by Markossian et al [95]. Like us, Donald et al [94] and Markossian et al [95] are evaluating their programs in a study, which will further provide an insight into the acceptability and engagement of the digital programs and identification of patient-reported outcomes and potential factors for implementation.

XSL•FO
RenderX

Evaluation and Implementation

The efficacy of MK&M in improving patient activation and subsequent self-management behaviors is currently being evaluated in a multisite clinical trial in the United Kingdom, with results expected in 2023 (protocol paper published elsewhere [96]). Patient activation is considered the cornerstone of effective self-management in CKD, as it encompasses all the necessary ingredients for effective self-management (ie, knowledge, skills, and confidence) [45]. Findings from the trial will provide the first step in determining whether MK&M is effective in increasing patient activation and promoting self-management behaviors in people with nondialysis CKD. In addition to evaluating MK&M, we also look to implement MK&M in renal services as a part of clinical care. Currently renal services are undergoing redesigning as a part of the Renal Service Transformation Programme [97]; consequently, we are conducting preliminary work before implementing MK&M into clinical care to ensure MK&M can be successfully embedded within the new model of care.

Strengths and Limitations

The inclusion of a stakeholder steering group comprising multidisciplinary professionals and PPI members with a variety of experience and expertise, who were involved in the codevelopment of MK&M, was vital to the successful development of MK&M. The PPI group was key in identifying the needs and preferences of people with CKD; however, the PPI members were predominantly White British, female, and well educated. Consequently, as our PPI group may not be representative of the CKD population, MK&M may not be suitable for all individuals with CKD. Successful use of MK&M is likely to require a certain level of health literacy and activation. It was developed in partnership with individuals who participate and engage in activities such as PPI and are therefore likely to be more engaged than the typical patient. The generalizability and applicability of MK&M will be established and addressed during the evaluation, refinement, and implementation phases of the program.

Although the IM approach provides a useful planning template to develop a tool that is theoretically grounded and evidence-based, it is very time consuming. The process described in this paper took approximately 2 years to complete, which was longer than initially anticipated; this was partly because of the COVID-19 pandemic, where members of the participatory steering group had competing clinical interests that took priority during this period, and the time taken to adapt a preexisting platform to suit our needs. Despite this, others have also reported similar experiences when using IM [36,98,99].

Despite the ambition, and initially setting out, to coproduce MK&M, it was not feasible owing to the COVID-19 pandemic. The inability to meet face-to-face during this period influenced the level of involvement of the members of the participatory steering group; moreover, clinical members were required to work on the frontline of the United Kingdom's NHS, and people with kidney disease were required to shield themselves because of being classed as clinically vulnerable. Although the steering group actively inputted during this time, it was not to the degree that we initially envisaged when we began this work; however, we did what we were able to during such challenging circumstances. Thus, we consider this work to be co-designed with the elements of coproduction.

Future Suggestions

We acknowledge that MK&M may not meet the needs of those with lower levels of activation and/or health literacy. The findings from our mixed methods evaluation of MK&M will provide a wealth of information, including details regarding program usage, engagement, and adherence. Exploring patients' and HCPs' views and experiences of using MK&M will provide an in-depth understanding of its usability, functionality, and acceptability. The findings will inform the potential improvements and adaptations we could make to the program to ensure that it is more acceptable and accessible to disadvantaged groups. In addition, the program requires digital literacy and thus may exclude individuals who are not computer literate or do not have access to digital technology. More work is needed to improve digital literacy and ensure the provision of internet access. There is also scope for digital programs such as MK&M to address the health inequalities that exist in the CKD population, and further work is needed to explore how these programs could effectively reduce health inequalities in CKD.

Conclusions

This paper provides a detailed example of the application of IM in the development of a theory- and evidence-based digital self-management program for people living with CKD. We have been successful in developing a digital holistic CKD-specific self-management program that provides accessible, tailored, and interactive information and support to help people with CKD improve their awareness and understanding of their condition, health knowledge, and confidence in managing their condition. MK&M aims to provide ongoing self-management education, support, and guidance for people living with CKD to manage their condition and live full, productive lives. Results from the multicenter randomized controlled trial evaluating MK&M will provide evidence of the efficacy, effectiveness, and cost-effectiveness of the program. If MK&M is successful in increasing patient activation and subsequent self-management behaviors, we anticipate that it will also improve people's QoL, physical function, and symptom burden while reducing the number of hospital admissions and saving costs to the NHS.

Acknowledgments

The authors would like to thank Professor John Feehally, Charlie Franklin, Mike Bonar, and the patient and public involvement steering group for their invaluable input and contribution to the development of My Kidneys and Me. This work is independent research funded by the Stoneygate Trust and the National Institute of Health and Care Research (NIHR) Applied Research

XSL•FO
RenderX

Collaboration East Midlands and supported by the NIHR Leicester Biomedical Research Centre. Patient and public involvement activities were supported by a grant from the Kidney Care Appeal awarded to TJW. HMLY was supported by a fellowship from the NIHR (DRF-2016-09-015). The funders had no role in the study design, collection, analysis, or interpretation of the data. The views expressed are those of the authors and not necessarily those of the NIHR, National Health Service (NHS), or Department of Health and Social Care.

Authors' Contributions

CJL, TJW, and ACS conceptualized and led on the idea of My Kidneys and Me (MK&M). VJ, SS, TY, and MJD helped develop the idea of MK&M and provided guidance based on their experiences of developing MyDESMOND. CJL, with input from TJW, LW, VJ, SS, CB, MH, TY, MJD, and ACS, led the development and adaptation of the program content, including the underpinning theories and behavior change. CJL, TJW, MG-B, HMLY, ACN, JOB, JB, NV, MM, JH, VP, FW, and ACS contributed to the development of the educational content of MK&M. CJL drafted the manuscript. TJW, MH, MJD, and ACS revised the initial content and structure of the manuscript. All authors have read, revised, and approved the final manuscript.

Conflicts of Interest

CB, VJ, and SS are employed through the University Hospitals of Leicester National Health Service Trust, which holds the intellectual property rights for and receives not-for-profit income for the Diabetes Education and Self-Management for Ongoing and Newly Diagnosed (DESMOND) suite of programs (which includes MyDESMOND).

Multimedia Appendix 1
Description of theories and frameworks selected and the rationale.
[DOCX File , 30 KB - jmir_v24i11e39657_app1.docx]

Multimedia Appendix 2 [DOCX File , 16 KB - jmir_v24i11e39657_app2.docx]

References

1. Kim LG, Cleary F, Wheeler DC, Caplin B, Nitsch D, Hull SA, UK National Chronic Kidney Disease Audit. How do primary care doctors in England and Wales code and manage people with chronic kidney disease? Results from the National Chronic Kidney Disease Audit. Nephrol Dial Transplant 2018 Aug 01;33(8):1373-1379 [FREE Full text] [doi: 10.1093/ndt/gfx280] [Medline: 29045728]
2. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. Lancet 2017 Mar;389(10075):1238-1252 [FREE Full text] [doi: 10.1016/s0140-6736(16)32064-5]
3. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 2004 Sep 23;351(13):1296-1305. [doi: 10.1056/NEJMoa041031] [Medline: 15385656]
4. Kalantar-Zadeh K, Jafar TH, Nitsch D, Neuen BL, Perkovic V. Chronic kidney disease. Lancet 2021 Aug;398(10302):786-802. [doi: 10.1016/s0140-6736(21)00519-5]
5. Neale EP, Middleton J, Lambert K. Barriers and enablers to detection and management of chronic kidney disease in primary healthcare: a systematic review. BMC Nephrol 2020 Mar 12;21(1):83 [FREE Full text] [doi: 10.1186/s12882-020-01731-x] [Medline: 32160886]
6. Donald M, Kahlon BK, Beanlands H, Straus S, Ronksley P, Herrington G, et al. Self-management interventions for adults with chronic kidney disease: a scoping review. BMJ Open 2018 Mar 22;8(3):e019814 [FREE Full text] [doi: 10.1136/bmjopen-2017-019814] [Medline: 29567848]
7. Havas K, Bonner A, Douglas C. Self-management support for people with chronic kidney disease: patient perspectives. J Ren Care 2016 Mar;42(1):7-14. [doi: 10.1111/jorc.12140] [Medline: 26404772]
8. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4. [Medline: 10345255]
9. Davies MJ, Heller S, Skinner TC, Campbell MJ, Carey ME, Cradock S, Diabetes Education Self Management for Ongoing Newly Diagnosed Collaborative. 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 Mar 01;336(7642):491-495 [FREE Full text] [doi: 10.1136/bmj.39474.922025.BE] [Medline: 18276664]
10. Novak M, Costantini L, Schneider S, Beanlands H. Approaches to self-management in chronic illness. Semin Dial 2013;26(2):188-194. [doi: 10.1111/sdi.12080] [Medline: 23520989]
11. 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]

XSL•FO
RenderX

12. Peng S, He J, Huang J, Lun L, Zeng J, Zeng S, et al. Self-management interventions for chronic kidney disease: a systematic review and meta-analysis. BMC Nephrol 2019 Apr 26;20(1):142 [FREE Full text] [doi: 10.1186/s12882-019-1309-y] [Medline: 31027481]
13. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]
14. Welch JL, Bartlett Ellis RJ, Perkins SM, Johnson CS, Zimmerman LM, Russell CL, et al. Knowledge and awareness among patients with chronic kidney disease stage 3. Nephrol Nurs J 2016;43(6):513-519. [Medline: 30550080]
15. Molnar AO, Akbari A, Brimble KS. Perceived and objective kidney disease knowledge in patients with advanced CKD followed in a multidisciplinary CKD clinic. Can J Kidney Health Dis 2020 Feb 11;7:2054358120903156 [FREE Full text] [doi: 10.1177/2054358120903156] [Medline: 32110417]
16. Finkelstein FO, Story K, Firanek C, Barre P, Takano T, Soroka S, et al. Perceived knowledge among patients cared for by nephrologists about chronic kidney disease and end-stage renal disease therapies. Kidney Int 2008 Nov;74(9):1178-1184 [FREE Full text] [doi: 10.1038/ki.2008.376] [Medline: 18668024]
17. Narva AS, Norton JM, Boulware LE. Educating patients about CKD: the path to self-management and patient-centered care. Clin J Am Soc Nephrology 2015 Nov 04;11(4):694-703. [doi: 10.2215/cjn.07680715]
18. Greer RC, Crews DC, Boulware LE. Challenges perceived by primary care providers to educating patients about chronic kidney disease. J Ren Care 2012 Dec;38(4):174-181 [FREE Full text] [doi: 10.1111/j.1755-6686.2012.00323.x] [Medline: 23176576]
19. Greer RC, Cooper LA, Crews DC, Powe NR, Boulware LE. Quality of patient-physician discussions about CKD in primary care: a cross-sectional study. Am J Kidney Dis 2011 Apr;57(4):583-591 [FREE Full text] [doi: 10.1053/j.ajkd.2010.08.027] [Medline: 21131116]
20. Lee YJ, Shin SJ, Wang RH, Lin KD, Lee YL, Wang YH. Pathways of empowerment perceptions, health literacy, self-efficacy, and self-care behaviors to glycemic control in patients with type 2 diabetes mellitus. Patient Educ Couns 2016 Feb;99(2):287-294 [FREE Full text] [doi: 10.1016/j.pec.2015.08.021] [Medline: 26341940]
21. Welch JL, Johnson M, Zimmerman L, Russell CL, Perkins SM, Decker BS. Self-management interventions in stages 1 to 4 chronic kidney disease: an integrative review. West J Nurs Res 2015 May;37(5):652-678 [FREE Full text] [doi: 10.1177/0193945914551007] [Medline: 25239136]
22. Bonner A, Havas K, Douglas C, Thepha T, Bennett P, Clark R. Self-management programmes in stages 1-4 chronic kidney disease: a literature review. J Ren Care 2014 Sep;40(3):194-204. [doi: 10.1111/jorc.12058] [Medline: 24628848]
23. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
24. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
25. Slater H, Dear BF, Merolli MA, Li LC, Briggs AM. Use of eHealth technologies to enable the implementation of musculoskeletal Models of Care: evidence and practice. Best Pract Res Clin Rheumatol 2016 Jun;30(3):483-502 [FREE Full text] [doi: 10.1016/j.berh.2016.08.006] [Medline: 27886943]
26. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
27. Hadjiconstantinou M, Schreder S, Brough C, Northern A, Stribling B, Khunti K, et al. Using intervention mapping to develop a digital self-management program for people with type 2 diabetes: tutorial on MyDESMOND. J Med Internet Res 2020 May 11;22(5):e17316 [FREE Full text] [doi: 10.2196/17316] [Medline: 32391797]
28. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. New Jersey: Prentice Hall; 1986.
29. Clarke A, MacKinnon H, Yates T, Smith A. Mp412 The “person-based approach” to developing a structured group education programme to increase physical activity in CKD: the pact-project. Nephrology Dialysis Transplantation 2016 May;31(suppl_1):i477. [doi: 10.1093/ndt/gfw191.16]
30. Clarke A. Developing behaviour change interventions to increase levels of physical activity in patients with chronic kidney disease. University of Leicester. 2018. URL: https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.745848 [accessed 2022-01-10]
31. Ajzen I. The theory of planned behavior. Organizational Behav Human Decision Processes 1991 Dec;50(2):179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-t]
32. MacKinnon H, Clarke A, Singh S, Smith A. MP409 Development of a self-directed programme to increase physical activity in chronic kidney disease. Nephrology Dialysis Transplantation 2016 May;31(suppl_1):i476,. [doi: 10.1093/ndt/gfw191.13]
33. Mackinnon H. Development and feasibility testing of a self-directed intervention to increase physical activity in chronic kidney disease: University of Leicester; 2022.

XSL•FO
RenderX

34. Eldredge LK, Markham CM, Ruiter RA, Fernández ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach, 4th Edition. Hoboken, New Jersey, United States: Wiley; Feb 2016.
35. Kok G, Peters LW, Ruiter RA. Planning theory- and evidence-based behavior change interventions: a conceptual review of the intervention mapping protocol. Psicol Reflex Crit 2017 Oct 18;30(1):19 [FREE Full text] [doi: 10.1186/s41155-017-0072-x] [Medline: 32026109]
36. Detaille SI, van der Gulden JW, Engels JA, Heerkens YF, van Dijk FJH. Using intervention mapping (IM) to develop a self-management programme for employees with a chronic disease in the Netherlands. BMC Public Health 2010 Jun 21;10:353 [FREE Full text] [doi: 10.1186/1471-2458-10-353] [Medline: 20565925]
37. Geense WW, van Gaal BG, Knoll JL, Cornelissen EA, Schoonhoven L, Kok G. Online support program for parents of children with a chronic kidney disease using intervention mapping: a development and evaluation protocol. JMIR Res Protoc 2016 Jan 13;5(1):e1 [FREE Full text] [doi: 10.2196/resprot.4837] [Medline: 26764218]
38. Shen H, van der Kleij R, van der Boog PJ, Song X, Wang W, Zhang T, et al. Development and evaluation of an eHealth self-management intervention for patients with chronic kidney disease in China: protocol for a mixed-method hybrid type 2 trial. BMC Nephrol 2020 Nov 19;21(1):495 [FREE Full text] [doi: 10.1186/s12882-020-02160-6] [Medline: 33213398]
39. Been-Dahmen JM, Beck DK, Peeters MA, van der Stege H, Tielen M, van Buren MC, et al. Evaluating the feasibility of a nurse-led self-management support intervention for kidney transplant recipients: a pilot study. BMC Nephrol 2019 Apr 27;20(1):143 [FREE Full text] [doi: 10.1186/s12882-019-1300-7] [Medline: 31029107]
40. Hardeman W, Sutton S, Griffin S, Johnston M, White A, Wareham NJ, et al. A causal modelling approach to the development of theory-based behaviour change programmes for trial evaluation. Health Educ Res 2005 Dec;20(6):676-687. [doi: 10.1093/her/cyh022] [Medline: 15781446]
41. Corbin J, Strauss A. Managing chronic illness at home: three lines of work. Qual Sociol 1985;8(3):224-247. [doi: 10.1007/BF00989485]
42. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
43. Morton K, Dennison L, May C, Murray E, Little P, McManus R, 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]
44. Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Update Work Group. KDIGO 2017 Clinical Practice Guideline Update for the diagnosis, evaluation, prevention, and treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). Kidney Int Suppl 2017 Dec;7:1-59 [FREE Full text] [doi: 10.1016/j.kisu.2017.10.001] [Medline: 30681074]
45. Lightfoot CJ, Nair D, Bennett PN, Smith AC, Griffin AD, Warren M, et al. Patient activation: the cornerstone of effective self-management in chronic kidney disease? Kidney Dialysis 2022 Mar 03;2(1):91-105. [doi: 10.3390/kidneydial2010012]
46. Chronic Kidney Disease: Assessment and Management. London: National Institute for Health and Care Excellence (NICE); Nov 24, 2021.
47. Donald M, Beanlands H, Straus S, Ronksley P, Tam-Tham H, Finlay J, et al. Identifying needs for self-management interventions for adults with CKD and their caregivers: a qualitative study. Am J Kidney Dis 2019 Oct;74(4):474-482. [doi: 10.1053/j.ajkd.2019.02.006] [Medline: 30952486]
48. Lin MY, Liu MF, Hsu LF, Tsai PS. Effects of self-management on chronic kidney disease: a meta-analysis. Int J Nurs Stud 2017 Sep;74:128-137. [doi: 10.1016/j.ijnurstu.2017.06.008] [Medline: 28689160]
49. Lopez-Vargas PA, Tong A, Howell M, Craig JC. Educational interventions for patients with CKD: a systematic review. Am J Kidney Dis 2016 Sep;68(3):353-370. [doi: 10.1053/j.ajkd.2016.01.022] [Medline: 27020884]
50. Lee MC, Wu SV, Hsieh NC, Tsai JM. Self-management programs on eGFR, depression, and quality of life among patients with chronic kidney disease: a meta-analysis. Asian Nurs Res (Korean Soc Nurs Sci) 2016 Dec;10(4):255-262 [FREE Full text] [doi: 10.1016/j.anr.2016.04.002] [Medline: 28057311]
51. Mason J, Khunti K, Stone M, Farooqi A, Carr S. Educational interventions in kidney disease care: a systematic review of randomized trials. Am J Kidney Dis 2008 Jun;51(6):933-951. [doi: 10.1053/j.ajkd.2008.01.024] [Medline: 18440681]
52. Shen H, van der Kleij RM, van der Boog PJ, Chang X, Chavannes NH. Electronic health self-management interventions for patients with chronic kidney disease: systematic review of quantitative and qualitative evidence. J Med Internet Res 2019 Nov 05;21(11):e12384 [FREE Full text] [doi: 10.2196/12384] [Medline: 31687937]
53. Zimbudzi E, Lo C, Misso ML, Ranasinha S, Kerr PG, Teede HJ, et al. Effectiveness of self-management support interventions for people with comorbid diabetes and chronic kidney disease: a systematic review and meta-analysis. Syst Rev 2018 Jun 13;7(1):84 [FREE Full text] [doi: 10.1186/s13643-018-0748-z] [Medline: 29898785]
54. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010 Mar 01;31(1):399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
55. Wilkinson TJ, Memory K, Lightfoot CJ, Palmer J, Smith AC. Determinants of patient activation and its association with cardiovascular disease risk in chronic kidney disease: a cross-sectional study. Health Expect 2021 Jun 09;24(3):843-852 [FREE Full text] [doi: 10.1111/hex.13225] [Medline: 33835670]

XSL•FO
RenderX

56. Wilkinson T, Goodliffe S, Wilde L, Smith A. P176 -"I've been diagnosed since 1970 and nobody has ever given me advice on diet or lifestyle": a research priority setting workshop with CKD patients treated in primary care. University Of Leicester. URL: https://vo2k0qci4747qecahf07gktt-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/P176.pdf [accessed 2022-01-10]
57. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health - a perspective on definition and distinctions. Public Health Res Pract 2022 Jun 15;32(2) [FREE Full text] [doi: 10.17061/phrp3222211] [Medline: 35702744]
58. Clarke D, Jones F, Harris R, Robert G, Collaborative Rehabilitation Environments in Acute Stroke (CREATE) team. What outcomes are associated with developing and implementing co-produced interventions in acute healthcare settings? A rapid evidence synthesis. BMJ Open 2017 Jul 11;7(7):e014650 [FREE Full text] [doi: 10.1136/bmjopen-2016-014650] [Medline: 28701409]
59. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019 Apr 01;40(1):127-146 [FREE Full text] [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]
60. Havas K, Douglas C, Bonner A. Meeting patients where they are: improving outcomes in early chronic kidney disease with tailored self-management support (the CKD-SMS study). BMC Nephrol 2018 Oct 20;19(1):279 [FREE Full text] [doi: 10.1186/s12882-018-1075-2] [Medline: 30342487]
61. Corbin JM, Strauss A. Unending Work and Care: Managing Chronic Illness at Home. San Francisco: Jossey-Bass; 1988.
62. 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]
63. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Applied Psychol 2008;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
64. 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]
65. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London, UK: Silverback Publishing; 2014.
66. Schwarzer R. Self-efficacy in the adoption and maintenance of health behaviors: theoretical approaches and a new model. In: Self-efficacy: Thought Control of Action. Washington DC, US: Hemisphere Publishing Corp; 1992.
67. MacPhail M, Mullan B, Sharpe L, MacCann C, Todd J. Using the health action process approach to predict and improve health outcomes in individuals with type 2 diabetes mellitus. Diabetes Metab Syndr Obes 2014;7:469-479 [FREE Full text] [doi: 10.2147/DMSO.S68428] [Medline: 25342914]
68. Bandura A. Self-efficacy: The Exercise of Control. New York, US: W H Freeman/Times Books/ Henry Holt & Co; 1997.
69. Leventhal H, Meyer D, Nerenz D. The common sense model of illness danger. In: Medical Psychology. Oxford, UK: Pergamon; 1980.
70. Leventhal H, Phillips LA, Burns E. The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. J Behav Med 2016 Dec;39(6):935-946. [doi: 10.1007/s10865-016-9782-2] [Medline: 27515801]
71. Hale ED, Treharne GJ, Kitas GD. The common-sense model of self-regulation of health and illness: how can we use it to understand and respond to our patients' needs? Rheumatology (Oxford) 2007 Jun 27;46(6):904-906. [doi: 10.1093/rheumatology/kem060] [Medline: 17449488]
72. Pagels AA, Söderquist BK, Heiwe S. Differences in illness representations in patients with chronic kidney disease. J Ren Care 2015 Sep 05;41(3):146-155. [doi: 10.1111/jorc.12117] [Medline: 25753248]
73. Wu C, Lin C, Hsieh H, Chang S. Lived experiences and illness representation of Taiwanese patients with late-stage chronic kidney disease. J Health Psychol 2016 Dec 10;21(12):2788-2798. [doi: 10.1177/1359105315587134] [Medline: 26060242]
74. Lin CC, Chen MC, Hsieh HF, Chang SC. Illness representations and coping processes of Taiwanese patients with early-stage chronic kidney disease. J Nurs Res 2013 Jun;21(2):120-128. [doi: 10.1097/jnr.0b013e3182921fb8] [Medline: 23681348]
75. Schwarzer R, Fuchs R. Changing risk behaviors and adopting health behaviors: the role of self-efficacy beliefs. In: Self-efficacy in Changing Societies. Cambridge, UK: Cambridge University Press; 2010.
76. Wu SV, Courtney M, Edwards H, McDowell J, Shortridge-Baggett LM, Chang P. Self-efficacy, outcome expectations and self-care behaviour in people with type 2 diabetes in Taiwan. J Clin Nurs 2007 Nov;16(11C):250-257. [doi: 10.1111/j.1365-2702.2006.01930.x] [Medline: 17931318]
77. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
78. Curtin RB, Walters BA, Schatell D, Pennell P, Wise M, Klicko K. Self-efficacy and self-management behaviors in patients with chronic kidney disease. Adv Chronic Kidney Dis 2008 Apr;15(2):191-205. [doi: 10.1053/j.ackd.2008.01.006] [Medline: 18334246]

XSL•FO
RenderX

79. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
80. Clarke AL, Young HM, Hull KL, Hudson N, Burton JO, Smith AC. Motivations and barriers to exercise in chronic kidney disease: a qualitative study. Nephrol Dial Transplant 2015 Nov;30(11):1885-1892. [doi: 10.1093/ndt/gfv208] [Medline: 26056174]
81. Evangelidis N, Craig J, Bauman A, Manera K, Saglimbene V, Tong A. Lifestyle behaviour change for preventing the progression of chronic kidney disease: a systematic review. BMJ Open 2019 Oct 28;9(10):e031625 [FREE Full text] [doi: 10.1136/bmjopen-2019-031625] [Medline: 31662393]
82. MacKinnon HJ, Wilkinson TJ, Clarke AL, Gould DW, O'Sullivan TF, Xenophontos S, et al. The association of physical function and physical activity with all-cause mortality and adverse clinical outcomes in nondialysis chronic kidney disease: a systematic review. Ther Adv Chronic Dis 2018 Nov;9(11):209-226 [FREE Full text] [doi: 10.1177/2040622318785575] [Medline: 30364521]
83. Quinn LM, Davies MJ, Northern A, Brough C, Schreder S, Stribling B, et al. Use of MyDesmond digital education programme to support self-management in people with type 2 diabetes during the COVID-19 pandemic. Diabet Med 2021 Mar 23;38(3):e14469 [FREE Full text] [doi: 10.1111/dme.14469] [Medline: 33274489]
84. Hadjiconstantinou M, Barker MM, Brough C, Schreder S, Northern A, Stribling B, et al. Improved diabetes-related distress and self-efficacy outcomes in a self-management digital programme for people with type 2 diabetes, myDESMOND. Diabet Med 2021 May 15;38(5):e14551. [doi: 10.1111/dme.14551] [Medline: 33675068]
85. Qismet. URL: https://www.qismet.org.uk [accessed 2022-10-25]
86. ORCHA Health. URL: https://orchahealth.com [accessed 2022-10-25]
87. Digital Technology Assessment Criteria. National Health Service. URL: https://transform.england.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/ [accessed 2022-10-25]
88. Brown SA, Tyrer F, Clarke AL, Lloyd-Davies LH, Niyi-Odumosu FA, Nah RG, et al. Kidney symptom questionnaire: development, content validation and relationship with quality of life. J Ren Care 2018 May 24 (forthcoming). [doi: 10.1111/jorc.12247] [Medline: 29797783]
89. Hemmelgarn BR, Pannu N, Ahmed SB, Elliott MJ, Tam-Tham H, Lillie E, et al. Determining the research priorities for patients with chronic kidney disease not on dialysis. Nephrol Dial Transplant 2017 May 01;32(5):847-854 [FREE Full text] [doi: 10.1093/ndt/gfw065] [Medline: 27190349]
90. NHS Long Term Plan. National Health Service. URL: https://www.longtermplan.nhs.uk/ [accessed 2022-01-10]
91. French SD, Green SE, O'Connor DA, McKenzie JE, Francis JJ, Michie S, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. Implement Sci 2012 Apr 24;7(1):38 [FREE Full text] [doi: 10.1186/1748-5908-7-38] [Medline: 22531013]
92. Jones TM, Dear BF, Hush JM, Titov N, Dean CM. Application of intervention mapping to the development of a complex physical therapist intervention. Phys Ther 2016 Dec;96(12):1994-2004. [doi: 10.2522/ptj.20150387] [Medline: 27256070]
93. Mayes J, Billany RE, Vadaszy N, Young HM, Castle EM, Bishop NC, et al. The rapid development of a novel kidney-specific digital intervention for self-management of physical activity and emotional well-being during the COVID-19 pandemic and beyond: kidney beam. Clin Kidney J 2022 Mar;15(3):571-573 [FREE Full text] [doi: 10.1093/ckj/sfab239] [Medline: 35198162]
94. Donald M, Beanlands H, Straus SE, Smekal M, Gil S, Elliott MJ, et al. A web-based self-management support prototype for adults with chronic kidney disease (my kidneys my health): co-design and usability testing. JMIR Form Res 2021 Feb 09;5(2):e22220 [FREE Full text] [doi: 10.2196/22220] [Medline: 33560245]
95. Markossian TW, Boyda J, Taylor J, Etingen B, Modave F, Price R, et al. A mobile app to support self-management of chronic kidney disease: development study. JMIR Hum Factors 2021 Dec 15;8(4):e29197 [FREE Full text] [doi: 10.2196/29197] [Medline: 34914614]
96. Lightfoot CJ, Wilkinson TJ, Yates T, Davies MJ, Smith AC. 'Self-Management Intervention through Lifestyle Education for Kidney health' (the SMILE-K study): protocol for a single-blind longitudinal randomised control trial with nested pilot study. medRxiv 2022 May 19. [doi: 10.1101/2022.05.17.22275207]
97. Jenkins K. J Kidney Care 2021 Jul 02;6(4):160-161. [doi: 10.12968/jokc.2021.6.4.160]
98. McEachan RR, Lawton RJ, Jackson C, Conner M, Lunt J. Evidence, theory and context: using intervention mapping to develop a worksite physical activity intervention. BMC Public Health 2008 Sep 22;8(1):326 [FREE Full text] [doi: 10.1186/1471-2458-8-326] [Medline: 18808709]
99. Munir F, Kalawsky K, Wallis DJ, Donaldson-Feilder E. Using intervention mapping to develop a work-related guidance tool for those affected by cancer. BMC Public Health 2013 Jan 05;13:6 [FREE Full text] [doi: 10.1186/1471-2458-13-6] [Medline: 23289708]

Abbreviations

BCT: behavior change technique
CKD: chronic kidney disease

XSL•FO
RenderX

DESMOND: Diabetes Education and Self-Management for Ongoing and Newly Diagnosed
DHI: digital health intervention
DIMENSION-KD: Investigating lifestyle determinants of muscle and physical function, and the impact on patient experience and support needs in kidney disease
GP: general practitioner
HCP: health care professional
IM: intervention mapping
MK&M: My Kidneys & Me
NHS: National Health Service
PACT: Physical Activity Changing Together
PPI: patient and public involvement
QoL: quality of life
SMILE-K: Self-Management Intervention through Lifestyle Education for Kidney health
SPARK: Self-directed Programme to Increase Physical Activity in Chronic Kidney Disease

Edited by R Kukafka; submitted 17.05.22; peer-reviewed by H Mehdizadeh, ER Khalilian, M Donald, PS Tsai; comments to author 07.07.22; revised version received 05.08.22; accepted 11.08.22; published 14.11.22.

Please cite as:
Lightfoot CJ, Wilkinson TJ, Hadjiconstantinou M, Graham-Brown M, Barratt J, Brough C, Burton JO, Hainsworth J, Johnson V, Martinez M, Nixon AC, Pursey V, Schreder S, Vadaszy N, Wilde L, Willingham F, Young HML, Yates T, Davies MJ, Smith AC
The Codevelopment of "My Kidneys & Me": A Digital Self-management Program for People With Chronic Kidney Disease
J Med Internet Res 2022;24(11):e39657
URL: https://www.jmir.org/2022/11/e39657
doi:10.2196/39657
PMID:36374538

©Courtney J Lightfoot, Thomas J Wilkinson, Michelle Hadjiconstantinou, Matthew Graham-Brown, Jonathan Barratt, Christopher Brough, James O Burton, Jenny Hainsworth, Vicki Johnson, Maria Martinez, Andrew C Nixon, Victoria Pursey, Sally Schreder, Noemi Vadaszy, Lucina Wilde, Fiona Willingham, Hannah M L Young, Thomas Yates, Melanie J Davies, Alice C Smith. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of a 4-Week Internet-Delivered Mindfulness-Based Cancer Recovery Intervention on the Symptom Burden and Quality of Life of Patients With Breast Cancer: Randomized Controlled Trial

Luyao Wang[1], MD; Xing Chen[2], MD; Yueyang Peng[1], MD; Kun Zhang[1], MD; Jun Ma[1], PhD; Lin Xu[1], MD; Zixuan Liu[3], MD; Li Liu[1], PhD; Yang Luo[1], PhD; Can Gu[1], PhD

[1]Xiangya School of Nursing, Central South University, Changsha, China

[2]Hunan Key Laboratory of Oral Health Research & Xiangya Stomatological Hospital, Central South University, Changsha, China

[3]School of Nursing, Xinjiang Medical University, Urumqi, China

Corresponding Author:
Can Gu, PhD
Xiangya School of Nursing
Central South University
XiaoQian Bldg, 6th FI
172 TongZiPo St
Changsha, 410013
China
Phone: 86 82650269
Email: gucan_cs@csu.edu.cn

Abstract

Background: Mindfulness-based interventions (MBIs) can improve the symptoms and psychological well-being of patients with breast cancer. However, standard MBIs are an 8-week program delivered face-to-face, which may be inconvenient for patients with cancer. Many attempts have been made to adapt MBIs to increase their accessibility for patients with cancer while maintaining their therapeutic components and efficacy.

Objective: This study aimed to investigate the effectiveness of a 4-week internet-delivered mindfulness-based cancer recovery (iMBCR) program in reducing symptom burden and enhancing the health-related quality of life (HRQoL) of patients with breast cancer.

Methods: A total of 103 postoperative patients with breast cancer (stages 0 to IV) were randomly assigned to an iMBCR group (4-week iMBCR; n=51, 49.5%) or a control group (usual care and 4-week program of health education information; n=52, 50.5%). The study outcomes included symptom burden and HRQoL, as measured by the MD Anderson Symptom Inventory and the Functional Assessment of Cancer Therapy-Breast scale. All data were collected at baseline (T0), after the intervention (T1), and at 1-month follow-up (T2). Data analysis followed the intention-to-treat principle. Linear mixed models were used to assess the effects over time of the iMBCR program.

Results: Participants in the iMBCR group had significantly larger decreases in symptom burden than those in the control group at T1 (mean difference –11.67, 95% CI –16.99 to –6.36), and the decreases were maintained at T2 (mean difference –11.83, 95% CI –18.19 to –5.46). The HRQoL score in the iMBCR group had significantly larger improvements than that in the control group at T1 and T2 (mean difference 6.66, 95% CI 3.43-9.90 and mean difference 11.94, 95% CI 7.56-16.32, respectively).

Conclusions: Our preliminary findings suggest that the iMBCR program effectively improved the symptom burden and HRQoL of patients with breast cancer, and the participants in the iMBCR group demonstrated good adherence and completion rates. These results indicate that the iMBCR intervention might be a promising way to reduce symptom burden and improve HRQoL of patients with cancer.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000038980; http://www.chictr.org.cn/showproj.aspx?proj=62659

(J Med Internet Res 2022;24(11):e40059) doi:10.2196/40059

XSL•FO
RenderX

KEYWORDS

mindfulness-based cancer recovery; mindfulness-based intervention; cancer-related symptom; quality of life; breast cancer; internet-delivered intervention; mobile phone

Introduction

Background

Breast cancer has become the most frequent cancer type among women. The International Agency for Research on Cancer has reported that the 5-year prevalence of breast cancer is approximately 7.8 million cases globally [1]. In China, the estimated 5-year prevalence of breast cancer is approximately 1,390,095 cases, with a prevalence rate of nearly 197 per 100,000 [2]. The breast cancer burden has grown over time in China [3].

Cancer diagnosis and treatment are highly stressful experiences that cause patients to experience a range of physical or psychological symptoms; for example, up to 86% of patients with breast cancer report fatigue during treatment [4], 54% to 78% report sleep disturbances [5], 53% experience pain [6], and 70% to 81% report psychological distress [7]. Such symptom burdens negatively affect their health-related quality of life (HRQoL) and increase their need for supportive care service. Recently, a large cohort study evaluating HRQoL in this population demonstrated that the restrictions in HRQoL can persist for >10 years [8]. Evidence suggests that psychosocial interventions such as mindfulness-based interventions (MBIs) are effective in improving stress-related symptoms and HRQoL [9,10]. Mindfulness-based stress reduction (MBSR) is a structured 8-week group program initially designed to reduce chronic pain and stress-related symptoms by developing mindfulness, meaning a nonjudgmental and accepting moment-by-moment awareness [11]. It has been adapted for multiple patient populations, and mindfulness-based cancer recovery (MBCR) is an adaptation of MBSR specifically for patients with cancer [12]. This program is not used for cancer treatment but for stressors related to the disease. MBCR retains the core principles and practices of MBSR, and it adds specific intervention material for coping with cancer, further focusing on symptoms such as sleep problems, pain, and fear of cancer recurrence. Growing evidence supports the MBCR program's benefit for symptoms in patients with breast cancer; for example, previous MBCR studies have repeatedly shown improvements in fatigue [13,14], and several systematic reviews have summarized the benefits of MBCR and other MBIs across outcomes of sleep disturbance, depression, stress, and HRQoL in patients with breast cancer [15,16].

Although the efficacy of MBIs has been proven, patients with cancer still report some barriers to participation. Eyles et al [17] found that the 8-week commitment to the course is the main reason for nonparticipation. Even for those who participated in the intervention, the completion rate of 8-week courses was low; for example, Carlson et al [14] reported that only 64.9% (87/134) of the patients with breast cancer completed the 8-week course. Another study has identified that the sample attrition rate is high (52%) in the MBCR group [18]. Compared with the 8-week intervention, short-term MBIs may be more acceptable to patients with cancer and could save time and human resources. Many attempts have been made to abbreviate 8-week MBIs [19]; for example, Demarzo et al [20] explored the efficacy of standard 8-week MBIs and abbreviated 4-week MBIs for improving well-being in a nonclinical population. They found a similar efficacy between the 4-week and 8-week MBIs. Similarly, another study by Wirth et al [21] identified that a 4-week MBCR program significantly improved sleep quality among patients with cancer. Both of these studies showed high retention in participants (4-week group attrition rate: 4.2% and 5%, respectively). Thus, the effectiveness of a 4-week MBCR program for patients with breast cancer is worth exploring.

Studies have also identified other practical barriers that may diminish access of patients with cancer to face-to-face programs, including but not limited to cancer-related illnesses, limited mobility, fatigue, transportation inconvenience, and scarcity of trained therapists [22-24]. Moreover, COVID-19 regulations have restricted people's activities. An internet-delivered intervention might be a promising method to overcome these barriers. Zernicke et al [25] assessed the feasibility of delivering a web-based MBCR program to patients with cancer. Their findings support providing the web-based MBCR program to patients with cancer, that is, feasibility targets for recruitment and retention are achieved, and participants are satisfied with a web-based MBCR program. However, evidence of the effect of an internet-delivered MBCR (iMBCR) program in women with breast cancer is limited. Further research is required to explore the efficacy of an iMBCR program in patients with breast cancer.

Objectives

In this study, we constructed a 4-week iMBCR program, and the heuristic framework developed by Barrera and Castro [26] guided the process of cross-cultural intervention adaptation [26]. This study aimed to investigate the effectiveness of the 4-week iMBCR program in regard to changes in patients' symptom burden and HRQoL. We developed the following hypotheses: (1) women allocated to the iMBCR group would report greater reductions in symptom burden and improvements in HRQoL after the intervention and at 1-month follow-up than those allocated to the control group, and (2) a 4-week iMBCR program would perform well in terms of adherence and intervention completion rates.

Methods

Participants

Our study used the following inclusion criteria: (1) women aged between 18 and 70 years, (2) having a prior diagnosis with stages 0 to IV breast cancer and aware of their cancer diagnosis, (3) having completed 1 to 24 months after surgery, (4) with normal cognitive capacity and functional status (Mini-Mental State Examination score ≥27 points and Karnofsky performance status scale score >60 points), and (5) able to operate a

XSL•FO
RenderX

smartphone and WeChat (the most popular smartphone app used for communication in China). The exclusion criteria were as follows: (1) participating in other psychological interventions or consultations, (2) having a history of mental illness or a combination of other severe somatic diseases, and (3) refusing to participate.

Sample size was calculated by power analyses using G*Power software (version 3.1; Heinrich Heine University). According to a previous study that explored the effectiveness of the MBCR program in women with cancer, the effect size for HRQoL scores was 0.66 [27]. Thus, to predict the difference between the 2 groups at a 5% level of significance and a power of 0.8, 38 participants were required in each group. Allowing for a 20% attrition rate, an additional 10 participants were needed in each group. Thus, the total sample size required for this study was 96, with 48 participants in each group.

Study Design and Randomization

This study was a 2-arm, parallel-group randomized controlled trial. Randomization was performed after the participants agreed to participate and signed the informed consent. Patients were randomly divided into the iMBCR group and the control group in a 1:1 ratio according to a list of computer-generated random numbers. To guarantee allocation concealment, an independent researcher who was not involved in the recruitment performed the random assignments by delivering an opaque, sealed envelope to each participant. Because of the nature of the intervention, participants could not be blinded. The research assistants who collected the data were blinded to each participant's group allocation throughout the study.

Ethics Approval

This study was approved by the institutional review board of the Xiangya School of Nursing (E2020153), Central South University, and was registered at the Chinese Clinical Trial Registry (ChiCTR2000038980).

Recruitment and Data Collection

Participants were recruited at the breast cancer wards of a tertiary hospital in Changsha, Hunan province, China, between November 1, 2020, and August 15, 2021. Recruitment was conducted primarily through referrals from ward nurses and research posters displayed at the gynecological clinic and wards. We contacted interested participants to screen them for eligibility. During this time, 1 researcher used the Mini-Mental State Examination [28] and Karnofsky performance status scale [29] to assess each participant's cognitive and functional status. Next, we contacted the eligible participants and provided further information about our study to them. Participants were given the choice to participate or decline and were informed that they had the right to withdraw at any time without reprisal. Baseline data were collected using written questionnaires at the wards of the hospitals. Postintervention and 1-month follow-up results were collected through web-based questionnaires. Participants who attended <2 sessions were considered dropouts.

Intervention

iMBCR Group

Guided by the heuristic framework developed by Barrera and Castro [26], the original English version of the MBCR program was translated into Chinese with a cross-cultural adaptation process. We used the following steps:

1. Information gathering: the goal of this step is to identify the form and content of needed adaptations, as well as the characteristics and preferences of potential participants. We conducted a mixed study. The quantitative study investigated the supportive care needs, mindfulness levels, and HRQoL of Chinese patients with cancer. The results showed that patients with cancer had a high level of supportive care needs. Health system and information needs and psychological needs were the top 2 needs of Chinese patients with cancer. Multiple linear regression analyses revealed that psychological needs and mindfulness levels could significantly predict HRQoL in patients with cancer. This suggested that the HRQoL of patients with cancer might be improved with mindfulness-based psychological interventions. The qualitative study was conducted to understand fully the patients' cancer-related troubles or distress or discomfort, what they thought were the main reasons for these issues, and their attitudes toward participating in an 8-week psychosocial intervention. The interviews showed that patients with cancer experienced numerous symptoms but lacked strategies for coping with them. In addition, most (6/10, 60%) of the patients felt that an 8-week intervention course was too long for them.
2. Preliminary adaptation design: we translated the text of the 8-week MBCR program into Chinese and organized an expert-panel meeting to discuss the content and delivery format of an MBCR program in the Chinese cultural context. Combining the results of the previous phase with those of the expert-panel meeting, we created the preliminary 8-week web-based MBCR program.
3. Preliminary adaptation tests: we conducted a pilot study to test the feasibility and acceptability of an 8-week iMBCR program in patients with breast cancer. The results showed that the web-based MBCR intervention was acceptable to the participants, but only 40% (4/10) of the participants completed the full 8-week session. More than half (7/10, 70%) of the participants felt that the intervention was too long. Interviews with the participants who completed the intervention (4/10, 40%) showed that mindful practice techniques (eg, mindful breathing, body scan meditation, and mindful walking) were practical for daily life activities, but yoga exercises were difficult for some (3/4, 75%) of the patients to complete.
4. Adaptation refinement: on the basis of the problems in the pilot study and participant feedback, we adjusted the duration and content of the iMBCR program. Five experts were invited to evaluate the content importance and rationality of a 4-week MBCR program. Through 2 rounds of expert consultations, we constructed the 4-week iMBCR program. Table 1 presents the detailed contents of the 4-week iMBCR program.

XSL•FO
RenderX

Table 1. Content of the 4-week mindfulness-based cancer recovery program.

Sessions and modules	Content
Session 1: first experience of mindfulness—connecting mind and body	
Participative section	Self-introduction; review of the agenda and content of sessions and main rules; group discussion on changes and distress caused by cancer
Didactic presentation	Understand the core concepts and related knowledge of mindfulness; develop a mindful attitude
Mindfulness techniques	Raisin meditation; body scan practice
Home practice	Guided audio body scan practice; mindful eating; record pleasant and unpleasant events
Session 2: power of awareness—emotion and thought	
Participative section	Recording pleasant and unpleasant events; mindfulness practice experience and body reaction
Didactic presentation	Explain nonjudgmental attitude; self-acceptance of emotions and thoughts
Mindfulness techniques	Mindful breathing; mindful stretching; sitting meditation
Home practice	Guided audio sitting meditation; mindful breathing; observe individual responses to stressful events
Session 3: stress management and self-compassion	
Participative section	Group discussion: individual responses to stressful events; confusion or discovery in practice
Didactic presentation	Mindfulness coping with cancer-related symptoms and self-compassion
Mindfulness techniques	Mindful walking; loving-kindness meditation; sitting meditation
Home practice	Guided audio body scan practice; mindful walking; loving-kindness meditation
Session 4: new life—incorporating mindfulness into daily life	
Participative section	How to bring mindfulness into daily living
Didactic presentation	Attitude toward mindfulness practices; experience sharing; recommendation of resources for mindfulness practice
Mindfulness techniques	"Who am I?" meditation exercise; mindful breathing; mindful stretching
Home practice	Guided audio sitting meditation; loving-kindness meditation

The iMBCR group received a 4-week MBCR program (1.5 hours per week) and at least 30 minutes of daily mindfulness home practice. All participants were invited to scan a quick-response code to join a WeChat group. We used this group mainly to send to the participants links to web-based courses and intervention materials, as well as for instant web-based communication. Participants were invited to attend web-based courses on Saturday mornings through a videoconferencing app (Tencent) extensively used in China. They were required to use their own smartphone to access the web-based course, which was also accessible by a desktop computer. Our assistant provided systematic training on the use of videoconferencing to ensure that the courses were accessible to participants. Session attendance was recorded, and all courses were recorded by video. The course video was provided to those who were absent for any reason. A Chinese-version MBCR book and some assisted–mindfulness practice audios were provided to the participants for home practice. Participants were asked to use the WeChat applet to make a note after completing their daily home practice. The intervention was delivered by a therapist who had completed mindfulness training and had 4 years of experience in teaching MBCR, accompanied by an assistant with 2 years of experience in mindfulness practice.

Control Group

Participants in the control group received usual care (routine standard oncology care) and 4 cancer-themed health education sessions: recognizing stress and managing negative emotions, coping with the adverse effects of therapy, dietary guidance, and exercise guidance, which do not involve any mindfulness component. These 4 sessions were conducted through Tencent. The number and frequency of web-based courses were the same as those in the iMBCR group.

Treatment Fidelity

Several strategies were used to ensure treatment fidelity. A treatment manual specifying the content of each course was developed, and the interventions were followed strictly. All of the intervention courses were video recorded, and an investigator reviewed the video recording after each course to ensure proper implementation of the treatment manual. Furthermore, our team met at the end of each intervention week to discuss the intervention implementation quality.

Measures

Demographic and Clinical Characteristics

A self-designed demographic questionnaire was used to collect the demographic data of the participants, including age, education, marital status, employment status, religion, and meditation and yoga practice experience. Clinical information included time since diagnosis, stage of breast cancer, and type of treatment.

XSL•FO
RenderX

Symptom Burden

The Chinese version of the MD Anderson Symptom Inventory (MDASI-C) was used to evaluate the severity of symptoms and symptom interference with daily life [30]. We used the MDASI-C to evaluate the symptom burden of women diagnosed with breast cancer. The MDASI-C includes 13 core symptom-severity items (pain, fatigue, sleep disturbance, drowsiness, lack of appetite, nausea, vomiting, shortness of breath, numbness, difficulty remembering, dry mouth, distress, and sadness) and 6 symptom-interference items (general activity, walking, work, mood, relations with other people, and enjoyment of life). Each item was measured using a numeric rating scale ranging from 0 (not present) to 10 (as bad as I can imagine) to evaluate the participants' status over the past 24 hours. The internal consistency for the MDASI-C ranged from Cronbach α=.84 to Cronbach α=.90 [31].

HRQoL of Women Diagnosed With Breast Cancer

The Chinese version of the Functional Assessment of Cancer Therapy-Breast (FACT-B) scale was used to assess the HRQoL of women diagnosed with breast cancer. This is a 36-item questionnaire that includes the 27 items of general HRQoL associated with cancer and the 9 items of HRQoL related to breast cancer. FACT-B comprises the following subscale domains: physical well-being, social well-being, emotional well-being, functional well-being, and breast cancer–specific subscale. Each item was measured using a numeric rating scale ranging from 0 (not at all) to 4 (very much) to evaluate the participants' status over the past week. A higher score indicates better HRQoL. The test-retest reliability of the Chinese version of the FACT-B scale ranged from .82 to .89, and the Cronbach α coefficient ranged from .61 to .84 [32].

Intervention Adherence and Completion

Adherence to the intervention was calculated by dividing the number of performed training sessions by the number of recommended training sessions. The 4-week iMBCR program corresponded with 4 web-based sessions and 28 days of home practice. Completion was estimated by dividing the number of consenting participants by the number of participants who completed the study.

Statistical Methods

Data were analyzed using SPSS software (version 18.0; IBM Corp). The intention-to-treat analysis was applied. Missing data (<20%) were handled with the participants' average response on the remaining scale items, and the missing data in the FACT-B questionnaire were treated by the proportion method per the manual instructions [33]. Descriptive statistics were applied to calculate the mean and SD for continuous data and frequency and percentage for categorical data. Baseline differences between the groups were explored using a chi-square test or Fisher exact test for categorical variables and a 2-tailed independent sample *t* test for continuous variables. Linear mixed models were used to compare the groups over time on all outcome variables. The data were hierarchically arranged in a 2-level structure with time at level 1 nested within individuals at level 2. Fixed effects were specified for intercept, time, group, and the group×time interaction, whereas the random effect was the participant. Effect sizes for the mean changes between the groups were calculated using Cohen *d*, with 0.2, 0.5, and 0.8 considered a small effect size, medium effect size, and large effect size, respectively [34]. We assumed a 2-tailed *P* value of <.05 to be statistically significant.

Results

Participant Characteristics

A total of 103 participants were recruited; 51 (49.5%) participants in the iMBCR group and 52 (50.5%) participants in the control group. The age of the participants ranged from 28 to 67 (average 46.8, SD 7.9) years. The median length of time since diagnosis was 4 months. The majority (86/103, 83.5%) of the patients were diagnosed with stage II or stage III breast cancer. No significant differences were detected in the baseline characteristics between the iMBCR and control groups (none of the *P* values met the threshold for statistical significance). Details are listed in Table 2.

XSL•FO
RenderX

Table 2. Sociodemographic and clinical characteristics of the participants (N=103).

Characteristics	iMBCR[a] group (n=51)	Control group (n=52)	*t* test (*df*) or chi-square value (*df*)	*P* value
Age (years), mean (SD; range)	45.37 (7.59; 28-67)	48.17 (8.05; 29-64)	*t* (*df*)=–1.816 (101)	.07
Education, n (%)			χ^2 (*df*)=4.798 (3)	.19
College or university	18 (35)	9 (17)		
High school or vocational	18 (35)	21 (40)		
Secondary	9 (18)	15 (29)		
≤Primary	6 (12)	7 (14)		
Marital status, n (%)			χ^2 (*df*)=0.000 (1)	>.99
Married	49 (96)	49 (94)		
Other marital status	2 (4)	3 (6)		
Employment status, n (%)			χ^2 (*df*)=0.816 (2)	.67
Employed	30 (59)	30 (58)		
Unemployed	16 (31)	14 (27)		
Retired	5 (10)	8 (15)		
Religion, n (%)			χ^2 (*df*)=1.414 (1)	.23
No	48 (94)	52 (100)		
Yes	3 (6)	0 (0)		
Experience of meditation practice, n (%)			χ^2 (*df*)=0.344 (1)	.56
No	42 (82)	45 (87)		
Yes	9 (18)	7 (13)		
Experience of yoga practice, n (%)			χ^2 (*df*)=0.079 (1)	.78
No	36 (71)	38 (73)		
Yes	15 (29)	14 (27)		
Time since diagnosis, months, n (%)			χ^2 (*df*)=3.945 (2)	.14
≤3	25 (49)	24 (46)		
4 to 12	15 (29)	23 (44)		
≥13	11 (22)	5 (10)		
Stage of breast cancer, n (%)			χ^2 (*df*)=6.728 (4)	.13
0 (carcinoma in situ)	2 (4)	1 (2)		
I	6 (12)	2 (4)		
II	28 (55)	31 (59)		
III	10 (19)	17 (33)		
IV	5 (10)	1 (2)		
Treatment, n (%)				
Breast surgery	51 (100)	52 (100)	N/A[b]	N/A
Chemotherapy	48 (94)	50 (96)	χ^2 (*df*)=0.000 (1)	.98
Radiotherapy	16 (31)	12 (23)	χ^2 (*df*)=0.895 (1)	.34
Immunotherapy	14 (27)	20 (38)	χ^2 (*df*)=1.412 (1)	.24
Baseline psychometric scores, mean (SD)				
MDASI[c]	63.12 (20.37)	61.44 (19.58)	*t* (*df*)=0.427 (101)	.67

XSL•FO
RenderX

Characteristics	iMBCR[a] group (n=51)	Control group (n=52)	*t* test (*df*) or chi-square value (*df*)	*P* value
FACT-B[d]	82.64 (15.74)	82.86 (16.27)	*t* (*df*)=−0.070 (101)	.95

[a]iMBCR: internet-delivered mindfulness-based cancer recovery.

[b]N/A: not applicable.

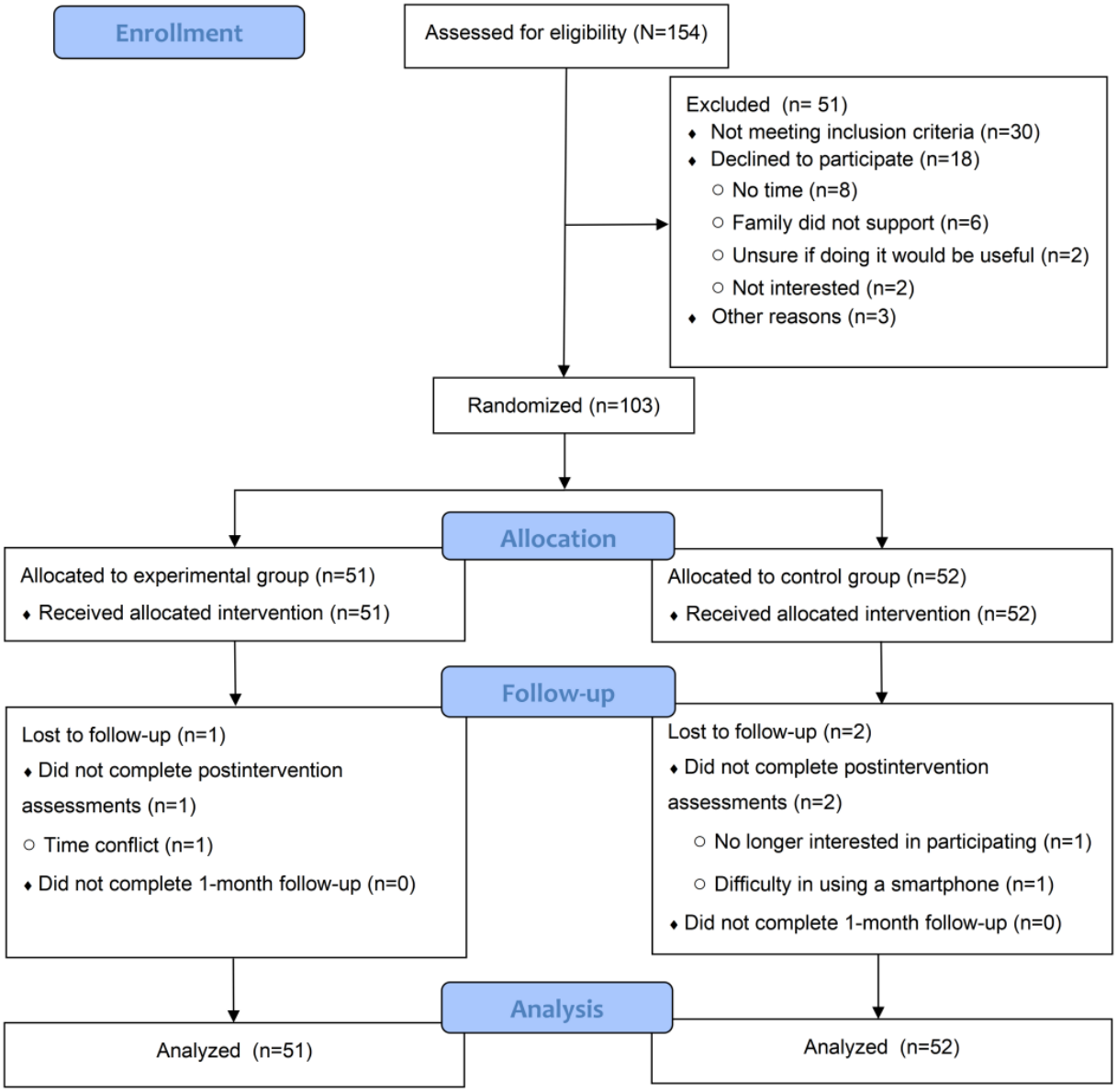
[c]MDASI: MD Anderson Symptom Inventory.

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

Intervention Adherence and Completion

Figure 1 shows the flowchart of participants throughout the study. A total of 100 participants completed all assessments, of whom 1 (1%) participant was lost from the iMBCR group (dropout rate=1/51, 2%), and 2 (2%) were lost from the control group (dropout rate=2/52, 4%). No significant differences were detected between the lost sample and the sample that completed all assessments (none of the *P* values met the threshold for statistical significance). The mean number of attended iMBCR courses was 3.6 (SD 0.7; adherence rate=3.6/4, 90%). The mean number of days of mindfulness home practice was 19.6 (SD 4.6; adherence rate=19.6/28, 70%).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



XSL•FO
RenderX

Effect on Outcomes

Symptom Burden

Statistically significant group×time effects were observed for MDASI total score ($F_{2,100}$=9.86; P<.001) and the symptom-severity subscale ($F_{2,100}$=11.73; P<.001) between the iMBCR group and control group, indicating that the MDASI total score and symptom-severity scores in both groups had different trends over the 3 time points (Table 3). Compared with the participants in the control group, those in the iMBCR group had significantly larger decreases in the MDASI total score (mean difference –11.67, 95% CI –16.99 to –6.36, Cohen d=–0.65) and symptom-severity scores (mean difference –8.87, 95% CI –12.54 to –5.21, Cohen d=–0.69) at T1, and the difference remained significant at T2 (mean difference –11.83, 95% CI –18.19 to –5.46, Cohen d=–0.98 and mean difference –9.51, 95% CI –14.19 to –4.82, Cohen d=–0.96, respectively). Table 4 shows the details of mean symptom-severity scores. Sleep disturbance, fatigue, and pain were the 3 most severe physical symptoms in both groups at the baseline. Regarding the between-group comparisons, fatigue was significantly decreased in the iMBCR group at T1 and T2 (none of the P values met the threshold for statistical significance), and sleep disturbance only significantly decreased at T1 (P=.02). The most severe psychological symptoms were sadness and distress, and only distress had significantly decreased at T1 (P=.01) in the iMBCR group compared with the control group.

The group×time interaction for the symptom-interference subscale was not significant (P=.26). Reductions were also observed in symptom-interference scores at T1 and T2, but the difference did not reach statistical significance.

XSL•FO
RenderX

Table 3. Comparison of symptom burden and health-related quality of life in the experimental and control groups.

Scale	iMBCR[a] group, mean (SD)	Control group, mean (SD)	Cohen *d*	Linear mixed model statistical tests							
				Difference in mean change from baseline between groups[b]		Group effect		Time effect		Group×time	
				Score (95% Cl)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value
MDASI[c] total score						5.55 (1,95)	.02	65.31 (2,100)	<.001	9.86 (2,100)	<.001
T0[d]	63.12 (20.37)	61.44 (19.58)	N/A[e]	N/A	N/A						
T1[f]	45.11 (13.30)	54.78 (16.46)	–0.65	–11.67 (–16.99 to –6.36)	<.001						
T2[g]	39.14 (9.95)	49.13 (10.43)	–0.98	–11.83 (–18.19 to –5.46)	<.001						
MDASI symptom-severity score						4.91 (1,102)	.03	69.61 (2,100)	<.001	11.73 (2,100)	<.001
T0	45.43 (13.79)	43.35 (14.74)	N/A	N/A	N/A						
T1	31.87 (8.85)	38.53 (10.43)	–0.69	–8.87 (–12.54 to –5.21)	<.001						
T2	26.92 (7.49)	34.27 (7.79)	–0.96	–9.51 (–14.19 to –4.82)	<.001						
MDASI symptom-interference score						4.56 (1,161)	.03	20.81 (2,121)	<.001	1.37 (2,121)	.26
T0	17.69 (9.78)	18.10 (7.24)	N/A	N/A	N/A						
T1	13.23 (6.41)	16.24 (7.30)	–0.44	–2.67 (–6.14 to 0.79)	.13						
T2	12.22 (4.20)	14.86 (3.68)	–0.67	–2.30 (–5.14 to 0.54)	.11						
FACT-B[h] total score						6.55 (1,96)	.01	36.49 (2,100)	<.001	14.82 (2,100)	<.001
T0	82.64 (15.74)	82.86 (16.27)	N/A	N/A	N/A						
T1	90.07 (13.74)	84.04 (11.52)	0.48	6.66 (3.43 to 9.90)	<.001						
T2	98.00 (7.51)	86.52 (10.88)	1.23	11.94 (7.56 to 16.32)	<.001						
Physical well-being						1.88 (1,101)	.17	20.48 (2,99)	<.001	5.18 (2,99)	.007
T0	19.41 (4.23)	19.63 (5.41)	N/A	N/A	N/A						
T1	21.38 (3.60)	20.20 (4.66)	0.28	1.50 (0.38 to 2.62)	.009						
T2	22.88 (2.02)	20.86 (4.76)	0.55	2.31 (0.87 to 3.76)	.002						
Emotional well-being						0.41 (1,101)	.53	25.42 (2,96)	<.001	2.20 (2,96)	.12
T0	15.20 (4.81)	15.35 (4.34)	N/A	N/A	N/A						

XSL•FO
RenderX

Scale	iMBCR[a] group, mean (SD)	Control group, mean (SD)	Cohen *d*	Linear mixed model statistical tests							
				Difference in mean change from baseline between groups[b]		Group effect		Time effect		Group×time	
				Score (95% Cl)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value
T1	16.46 (3.58)	16.32 (3.97)	0.04	0.35 (–0.66 to 1.36)	.50						
T2	18.46 (1.95)	17.26 (3.59)	0.42	1.38 (–0.05 to 2.81)	.06						
Functional well-being						3.22 (1,102)	.08	5.25 (2,90)	.007	4.89 (2,90)	.01
T0	11.73 (4.74)	11.66 (4.77)	N/A	N/A	N/A						
T1	12.36 (4.78)	10.10 (4.36)	0.49	1.31 (0.14 to 2.48)	.03						
T2	14.06 (2.68)	11.48 (4.32)	0.72	2.52 (0.90 to 4.15)	.003						
Social well-being						2.33 (1,101)	.13	2.02 (2,96)	.14	1.51 (2,96)	.23
T0	17.19 (5.31)	16.86 (5.21)	N/A	N/A	N/A						
T1	17.87 (4.84)	16.58 (3.70)	0.30	1.03 (–0.23 to 2.29)	.11						
T2	18.69 (3.03)	16.98 (3.58)	0.52	1.41 (–0.38 to 3.21)	.12						
Breast cancer–specific subscale for additional concerns						33.41 (1,99)	<.001	46.72 (2,196)	<.001	29.33 (2,196)	<.001
T0	19.12 (2.42)	19.37 (2.39)	N/A	N/A	N/A						
T1	22.00 (2.37)	19.94 (2.33)	0.88	2.36 (1.26 to 3.47)	<.001						
T2	23.92 (2.16)	19.94 (2.48)	1.70	4.28 (3.18 to 5.39)	<.001						

[a]iMBCR: internet-delivered mindfulness-based cancer recovery.

[b]Difference in mean change from baseline to end point between the groups.

[c]MDASI: MD Anderson Symptom Inventory.

[d]T0: baseline.

[e]N/A: not applicable.

[f]T1: after the intervention.

[g]T2: 1-month follow-up.

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

XSL•FO
RenderX

Table 4. Details of changes in symptom severity.

Symptom	Severity rank	Symptom severity						Difference in mean change between groups[a]			
		iMBCR[b] group, mean (SD)			Control group, mean (SD)			T0-T1		T0-T2	
		T0[c]	T1[d]	T2[e]	T0	T1	T2	Score (95% Cl)	*P* value	Score (95% Cl)	*P* value
Sleep disturbance	1	4.8 (2.1)	3.6 (1.4)	3.0 (1.0)	4.7 (2.5)	4.3 (1.9)	3.6 (1.2)	–0.81 (–1.46 to –0.16)	.02	–0.71 (–1.52 to 0.09)	.08
Sadness	2	4.6 (2.2)	3.1 (1.3)	2.6 (0.9)	4.2 (2.2)	3.5 (1.8)	2.8 (1.1)	–0.87 (–1.76 to 0.019)	.06	–0.68 (–1.46 to 0.09)	.08
Distress	3	4.3 (2.1)	3.0 (1.2)	2.8 (1.0)	4.3 (2.1)	3.8 (1.7)	3.3 (1.2)	–0.72 (–1.29 to –0.16)	.01	–0.52 (–1.26 to 0.22)	.17
Fatigue	4	4.2 (1.9)	2.9 (1.3)	2.6 (1.1)	3.8 (1.9)	3.3 (1.5)	2.8 (1.1)	–0.92 (–1.62 to –0.21)	.01	–0.69 (–1.36 to –0.02)	.04
Pain	5	3.6 (2.4)	2.6 (1.3)	2.2 (1.1)	3.9 (2.3)	3.2 (1.5)	3.1 (1.0)	–0.28 (–0.92 to 0.37)	.40	–0.58 (–1.43 to 0.27)	.18
Drowsiness	6	3.9 (2.3)	2.5 (1.4)	1.9 (0.9)	3.4 (1.6)	3.0 (1.4)	2.8 (1.1)	–0.95 (–1.66 to –0.25)	.009	–1.10 (–1.76 to –0.44)	<.001
Lack of appetite	7	3.6 (2.3)	2.3 (1.3)	1.6 (0.7)	3.3 (2.1)	3.0 (1.4)	2.4 (1.0)	–1.07 (–1.71 to –0.43)	<.001	–1.18 (–2.01 to –0.35)	.006
Dry mouth	8	3.4 (1.8)	3.0 (1.6)	2.5 (1.1)	3.3 (1.9)	3.4 (1.4)	3.1 (1.1)	–0.68 (–1.25 to –0.12)	.02	–1.07 (–1.76 to –0.37)	.003
Difficulty remembering	9	3.2 (2.4)	2.4 (1.6)	1.9 (1.1)	3.4 (2.2)	3.2 (1.8)	2.8 (1.4)	–0.65 (–1.18 to –0.09)	.02	–0.81 (–1.61 to –0.01)	.048
Numbness	10	2.9 (1.8)	2.1 (1.2)	1.9 (0.8)	3.2 (1.9)	2.7 (1.6)	2.2 (1.1)	–0.34 (–1.11 to 0.44)	.14	–0.02 (–0.71 to 0.67)	.04
Nausea	11	2.7 (2.1)	1.8 (1.3)	1.6 (0.9)	2.6 (1.9)	2.4 (1.6)	2.3 (0.9)	–0.76 (–1.28 to –0.24)	.005	–0.85(–1.62 to –0.08)	.03
Vomiting	12	2.3 (2.1)	1.4 (0.8)	1.1 (0.5)	2.1 (1.5)	1.8 (1.4)	1.7 (0.8)	–0.58 (–1.36 to 0.19)	.39	–0.74 (–1.43 to –0.05)	.96
Shortness of breath	13	1.8 (1.3)	1.2 (0.9)	1.2 (0.6)	1.3 (1.4)	0.9 (1.1)	1.2 (0.7)	–0.24 (–0.64 to 0.15)	.23	–0.57 (–1.11 to –0.03)	.04

[a]Difference in mean change from baseline to end point between the groups.

[b]iMBCR: internet-delivered mindfulness-based cancer recovery.

[c]T0: baseline.

[d]T1: after the intervention.

[e]T2: 1-month follow-up.

HRQoL Assessment

The results indicated a significant group×time interaction for the FACT-B total score ($F_{2,100}$=14.82; *P*<.001), physical well-being ($F_{2,99}$=5.18; *P*=.007), functional well-being ($F_{2,90}$=4.89; *P*=.01), and the breast cancer–specific subscale ($F_{2,196}$=29.33; *P*<.001). Participants in the iMBCR group had larger improvements in the FACT-B total score at T1 and T2 than the control group (mean difference 6.66, 95% CI 3.43-9.90, Cohen *d*=0.48 and mean difference 11.94, 95% CI 7.56-16.32, Cohen *d*=1.23, respectively). Statistical improvements were also observed in physical well-being, functional well-being, and the breast cancer–specific subscale at T1 and T2. Emotional well-being exhibited only a time-based effect ($F_{2,96}$=25.42; *P*<.001), and the difference in mean change between groups was not significant for T1 and T2. For social well-being, we found no significant group, time, or group×time interaction effects, indicating that the effect of the iMBCR program on patients' social well-being was not significant.

Discussion

Principal Findings

To our knowledge, this study is the first randomized controlled trial to investigate the effect of an iMBCR program in Chinese women diagnosed with breast cancer and adds to the few available studies on short-term MBIs. We found that the outcomes of symptom burden and HRQoL were improved immediately after the intervention, and the effect was maintained at the 1-month follow-up. This finding demonstrated that the 4-week iMBCR program is effective for women with breast cancer. Moreover, our intervention demonstrated good adherence and intervention completion rates.

XSL•FO
RenderX

Our study found that the MDASI total score in the iMBCR group had significantly decreased after the intervention and at 1-month follow-up compared with the control group. The findings supported our first hypothesis that the 4-week iMBCR program can relieve the symptom burden of women diagnosed with breast cancer. The results of reduced symptom burden after the intervention were consistent with a previous face-to-face 6-week group MBSR study in patients with breast cancer [35], which may indicate that the web-based mindfulness intervention may have an effect on symptoms that is similar to the effect of face-to-face interventions. The symptom-severity scores in the iMBCR group significantly decreased after the intervention. However, the effect was stronger at 1-month follow-up. The reasons may be that the severity of symptoms decreased over time and that the 4-week iMBCR intervention activated the patients' mindfulness behavior in daily life. Techniques taught in iMBCR courses, such as mindful breathing, body scan meditation, mindful stretching, and sitting meditation, had been internalized by the patients and could be useful when they experience symptoms or stressful events in daily life.

Consistent with existing symptom research on patients with breast cancer, we found that the top 3 severe physical symptoms at baseline were fatigue, sleep disturbance, and pain [36]. Current systematic reviews have concluded that standard 8-week MBIs could improve fatigue and quality of sleep in short-term (end of intervention) to medium-term (up to 6 months after baseline) effects for women diagnosed with breast cancer [15,37]. Regarding fatigue, our study found that the 4-week iMBCR intervention achieved results that were similar to those of previous studies. However, we did not find significant reductions in sleep disturbance during follow-up, which may be due to the fact that the 4-week iMBCR program was not intended to treat sleep problems and that sleep disturbance was assessed only by a single item instead of a standard sleep scale. Similar to previous study results [38,39], we did not find any significant effects of the iMBCR intervention on pain in patients with breast cancer.

As expected, the FACT-B total scores and most domain scores (physical well-being, function well-being, and breast cancer–specific additional concerns) significantly increased among the iMBCR group compared with the control group after the intervention and at 1-month follow-up, indicating that the 4-week iMBCR program could improve patients' HRQoL. The postintervention effect sizes for HRQoL found in our study (Cohen d=0.48) were within the same range as that in a previous study on a face-to-face standard 8-week group MBSR intervention for women diagnosed with stages 0 to III breast cancer (Cohen d=0.60) [40]. This finding suggests that the 4-week iMBCR intervention could achieve an effect on HRQoL that is similar to that achieved by standard 8-week MBIs. These results may be caused by multiple reasons. First, the iMBCR program in this study developed the patients' capacity for the intentional self-regulation of attention as well as the attitude and practice of acceptance, which has been proven to be effective in reducing negative reactivity and improving stress-related health outcomes [41]. Meanwhile, the improvement in the patients' symptom burden positively affected the HRQoL [42]. Besides, the group intervention format provided a path of communication and emotional support for patients with breast cancer. Social support reportedly predicted better adjustment to cancer and better quality of life [43].

Unlike previous studies, we found no intervention effect on emotional well-being [39], which was also reflected in the results of symptom severity (no intervention effect on sadness, and distress significantly reduced only during the intervention). The results of another 4-week MBCR program study also showed no positive intervention effects on psychological outcomes (depression and perceived stress) [21]. It seemed that a short-term mindfulness intervention was more effective for physical well-being than for emotional well-being, and the improvement in psychological outcomes required a longer mindfulness intervention period. Nevertheless, given the relative scarcity of studies on short-term mindfulness interventions, we were limited in our ability to draw any conclusion in this regard.

In this study, the attrition rate among the participants in the iMBCR group (dropout rate=1/51, 2%) was relatively low, and it was lower than that reported in the standard 8-week MBIs (typically 20%-30%) [44]. Furthermore, the adherence rate was high in both iMBCR courses and mindfulness home practice. These findings confirmed our second hypothesis. Our results demonstrated that the 4-week mindfulness intervention is acceptable for patients with breast cancer.

Limitations

This study includes several limitations. First, the completion of mindfulness home practice relied on the self-reporting of participants and was calculated by days and not specific practice minutes. This setup may affect the accuracy of intervention adherence data because self-reporting is subjective. In a future study, we will consider using more advanced technology to record user log-ins or the time spent on the web to measure intervention completion and adherence rates. Second, the heterogeneity of the sample with regard to cancer stage and treatment type may affect symptom burden and HRQoL. However, the differences in the cancer stages and treatment types that existed between the 2 groups at baseline were not statistically significant, which could have reduced the bias to some extent. Third, the patients in this study were followed up for only 1 month. We were unable to determine the medium- or long-term effects of the 4-week iMBCR program on symptom burden and HRQoL. As such, future research should incorporate longer follow-up periods to examine the durability of the effects of an iMBCR program. Fourth, our participants were aged <70 years, and it is unclear whether the program is feasible for, or would benefit, patients aged >70 years. Finally, because the study was conducted on the web, patients who did not own technological devices may have lost the opportunity to participate in the study.

Research Implications

Our study used a relatively abbreviated intervention and a novel technology that made the MBCR program more acceptable for patients with breast cancer. The results were encouraging for implementing this 4-week web-based mindfulness intervention to reduce symptom burden and improve HRQoL of patients with cancer. However, because of the short follow-up period in

XSL•FO
RenderX

this study, a long-term follow-up study is required to confirm the results more precisely. Our findings also provided evidence for eHealth services for patients with breast cancer and the effect of short-term MBIs. Future research is warranted to continue the investigation on the mechanisms of change in short-term web-based MBCR interventions. Researchers should examine the efficacy of the components of mindfulness and emphasize not only the techniques but also to what extent they are effective. They should also determine how much exposure to the intervention is needed to efficiently develop a feasible and effective program that increases its accessibility for a larger number of clinical populations.

Conclusions

Our study explored the effect of a 4-week iMBCR program in patients with breast cancer. The 4-week iMBCR program showed positive effects for symptom burden and HRQoL immediately after the intervention and at 1-month follow-up, and our intervention also demonstrated good adherence and completion rates. This low-cost web-based intervention can be more acceptable for patients and be easily translated into clinical practice to reach numerous patients. Further studies are warranted to explore the long-term effects of, and mechanisms of change in, short-term web-based MBCR interventions.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (81703084, 82272924).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH (V 1.6.1)
[PDF File (Adobe PDF File), 1517 KB - jmir_v24i11e40059_app1.pdf]

References

1. Latest global cancer data: cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020. World Health Organization. URL: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheets.pdf [accessed 2020-11-05]
2. Population fact sheets. The International Agency for Research on Cancer. URL: https://gco.iarc.fr/today/data/factsheets/populations/160-china-fact-sheets.pdf [accessed 2021-04-05]
3. Li T, Mello-Thoms C, Brennan PC. Descriptive epidemiology of breast cancer in China: incidence, mortality, survival and prevalence. Breast Cancer Res Treat 2016 Oct;159(3):395-406. [doi: 10.1007/s10549-016-3947-0] [Medline: 27562585]
4. Berger AM, Kumar G, LeVan TD, Meza JL. Symptom clusters and quality of life over 1 year in breast cancer patients receiving adjuvant chemotherapy. Asia Pac J Oncol Nurs 2020;7(2):134-140 [FREE Full text] [doi: 10.4103/apjon.apjon_57_19] [Medline: 32478130]
5. Browall M, Brandberg Y, Nasic S, Rydberg P, Bergh J, Rydén A, et al. A prospective exploration of symptom burden clusters in women with breast cancer during chemotherapy treatment. Support Care Cancer 2017 May 15;25(5):1423-1429 [FREE Full text] [doi: 10.1007/s00520-016-3527-1] [Medline: 27981366]
6. Chow S, Wan BA, Pidduck W, Zhang L, DeAngelis C, Chan S, et al. Symptom clusters in patients with breast cancer receiving radiation therapy. Eur J Oncol Nurs 2019 Oct;42:14-20. [doi: 10.1016/j.ejon.2019.07.004] [Medline: 31446259]
7. Dodd MJ, Cho MH, Cooper BA, Miaskowski C. The effect of symptom clusters on functional status and quality of life in women with breast cancer. Eur J Oncol Nurs 2010 Apr;14(2):101-110 [FREE Full text] [doi: 10.1016/j.ejon.2009.09.005] [Medline: 19897417]
8. Maurer T, Thöne K, Obi N, Jung AY, Behrens S, Becher H, et al. Health-related quality of life in a cohort of breast cancer survivors over more than 10 years post-diagnosis and in comparison to a control cohort. Cancers (Basel) 2021 Apr 13;13(8):1854 [FREE Full text] [doi: 10.3390/cancers13081854] [Medline: 33924513]
9. Jing S, Zhang A, Chen Y, Shen C, Currin-McCulloch J, Zhu C. Mindfulness-based interventions for breast cancer patients in China across outcome domains: a systematic review and meta-analysis of the Chinese literature. Support Care Cancer 2021 Oct;29(10):5611-5621. [doi: 10.1007/s00520-021-06166-0] [Medline: 33770259]
10. Pedro J, Monteiro-Reis S, Carvalho-Maia C, Henrique R, Jerónimo C, Silva ER. Evidence of psychological and biological effects of structured Mindfulness-Based Interventions for cancer patients and survivors: a meta-review. Psychooncology 2021 Nov;30(11):1836-1848 [FREE Full text] [doi: 10.1002/pon.5771] [Medline: 34288218]
11. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol 2003;10(2):144-156. [doi: 10.1093/clipsy.bpg016]
12. Carlson L, Speca M. Mindfulness-Based Cancer Recovery A Step-by-Step Mbsr Approach to Help You Cope with Treatment and Reclaim Your Life. Online: ReadHowYouWant.com; 2011.
13. Carlson LE, Garland SN. Impact of mindfulness-based stress reduction (MBSR) on sleep, mood, stress and fatigue symptoms in cancer outpatients. Int J Behav Med 2005;12(4):278-285. [doi: 10.1207/s15327558ijbm1204_9] [Medline: 16262547]

XSL•FO
RenderX

14. Carlson LE, Tamagawa R, Stephen J, Drysdale E, Zhong L, Speca M. Randomized-controlled trial of mindfulness-based cancer recovery versus supportive expressive group therapy among distressed breast cancer survivors (MINDSET): long-term follow-up results. Psychooncology 2016 Jul;25(7):750-759. [doi: 10.1002/pon.4150] [Medline: 27193737]
15. Schell LK, Monsef I, Wöckel A, Skoetz N. Mindfulness-based stress reduction for women diagnosed with breast cancer. Cochrane Database Syst Rev 2019 Mar 27;3:CD011518 [FREE Full text] [doi: 10.1002/14651858.CD011518.pub2] [Medline: 30916356]
16. Zhang J, Xu R, Wang B, Wang J. Effects of mindfulness-based therapy for patients with breast cancer: a systematic review and meta-analysis. Complement Ther Med 2016 Jun;26:1-10. [doi: 10.1016/j.ctim.2016.02.012] [Medline: 27261975]
17. Eyles C, Leydon GM, Hoffman CJ, Copson ER, Prescott P, Chorozoglou M, et al. Mindfulness for the self-management of fatigue, anxiety, and depression in women with metastatic breast cancer: a mixed methods feasibility study. Integr Cancer Ther 2015 Jan;14(1):42-56 [FREE Full text] [doi: 10.1177/1534735414546567] [Medline: 25161198]
18. Schellekens MP, Tamagawa R, Labelle LE, Speca M, Stephen J, Drysdale E, et al. Mindfulness-Based Cancer Recovery (MBCR) versus Supportive Expressive Group Therapy (SET) for distressed breast cancer survivors: evaluating mindfulness and social support as mediators. J Behav Med 2017 Jun;40(3):414-422 [FREE Full text] [doi: 10.1007/s10865-016-9799-6] [Medline: 27722908]
19. Schumer MC, Lindsay EK, Creswell JD. Brief mindfulness training for negative affectivity: a systematic review and meta-analysis. J Consult Clin Psychol 2018 Jul;86(7):569-583 [FREE Full text] [doi: 10.1037/ccp0000324] [Medline: 29939051]
20. Demarzo M, Montero-Marin J, Puebla-Guedea M, Navarro-Gil M, Herrera-Mercadal P, Moreno-González S, et al. Efficacy of 8- and 4-session mindfulness-based interventions in a non-clinical population: a controlled study. Front Psychol 2017;8:1343 [FREE Full text] [doi: 10.3389/fpsyg.2017.01343] [Medline: 28848465]
21. Wirth MD, Franco R, Wagner Robb S, Daniels K, Susko K, Hudson MF, et al. Randomized controlled trial of a 4-week mindfulness intervention among cancer survivors compared to a breathing control. Cancer Invest 2019;37(4-5):227-232. [doi: 10.1080/07357907.2019.1610968] [Medline: 31198066]
22. Robertson MC, Cox-Martin E, Liao Y, Flores SA, Shegog R, Markham CM, et al. Acceptance- and mindfulness-based techniques for physical activity promotion in breast cancer survivors: a qualitative study. Support Care Cancer 2022 Jan;30(1):465-473 [FREE Full text] [doi: 10.1007/s00520-021-06428-x] [Medline: 34313858]
23. Tate KJ, Newbury-Birch D, McGeechan GJ. A systematic review of qualitative evidence of cancer patients' attitudes to mindfulness. Eur J Cancer Care (Engl) 2018 Mar;27(2):e12783. [doi: 10.1111/ecc.12783] [Medline: 28983995]
24. Wang J, Shen Q, Ding N, Zhang T, Yang Z, Liu C, et al. A qualitative exploration of the unmet psychosocial rehabilitation needs of cancer survivors in China. Psychooncology 2016 Aug;25(8):905-912. [doi: 10.1002/pon.4023] [Medline: 26525458]
25. Zernicke KA, Campbell TS, Speca M, McCabe-Ruff K, Flowers S, Carlson LE. A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial. Psychosom Med 2014 May;76(4):257-267. [doi: 10.1097/PSY.0000000000000053] [Medline: 24804884]
26. Barrera M, Castro FG. A heuristic framework for the cultural adaptation of interventions. Clin Psychol Sci Pract 2006;13(4):311-316. [doi: 10.1111/j.1468-2850.2006.00043.x]
27. Lerman R, Jarski R, Rea H, Gellish R, Vicini F. Improving symptoms and quality of life of female cancer survivors: a randomized controlled study. Ann Surg Oncol 2012 Feb;19(2):373-378. [doi: 10.1245/s10434-011-2051-2] [Medline: 21913014]
28. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
29. Clancey JK. Karnofsky performance scale. J Neurosci Nurs 1995 Aug;27(4):220. [Medline: 7499924]
30. Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, et al. Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. Cancer 2000 Oct 01;89(7):1634-1646. [doi: 10.1002/1097-0142(20001001)89:7<1634::aid-cncr29>3.0.co;2-v] [Medline: 11013380]
31. Wang XS, Wang Y, Guo H, Mendoza TR, Hao X, Cleeland CS. Chinese version of the M. D. Anderson Symptom Inventory: validation and application of symptom measurement in cancer patients. Cancer 2004 Oct 15;101(8):1890-1901 [FREE Full text] [doi: 10.1002/cncr.20448] [Medline: 15386315]
32. Wan C, Zhang D, Yang Z, Tu X, Tang W, Feng C, et al. Validation of the simplified Chinese version of the FACT-B for measuring quality of life for patients with breast cancer. Breast Cancer Res Treat 2007 Dec 22;106(3):413-418. [doi: 10.1007/s10549-007-9511-1] [Medline: 17377841]
33. Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the functional assessment of cancer therapy-breast quality-of-life instrument. J Clin Oncol 1997 Mar;15(3):974-986. [doi: 10.1200/JCO.1997.15.3.974] [Medline: 9060536]
34. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Cambridge, Massachusetts, United States: Academic Press; 1969.
35. Lengacher CA, Reich RR, Post-White J, Moscoso M, Shelton MM, Barta M, et al. Mindfulness based stress reduction in post-treatment breast cancer patients: an examination of symptoms and symptom clusters. J Behav Med 2012 Feb;35(1):86-94. [doi: 10.1007/s10865-011-9346-4] [Medline: 21506018]

XSL•FO
RenderX

36. Haidinger R, Bauerfeind I. Long-term side effects of adjuvant therapy in primary breast cancer patients: results of a web-based survey. Breast Care (Basel) 2019 Apr;14(2):111-116 [FREE Full text] [doi: 10.1159/000497233] [Medline: 31798383]
37. Haller H, Winkler MM, Klose P, Dobos G, Kümmel S, Cramer H. Mindfulness-based interventions for women with breast cancer: an updated systematic review and meta-analysis. Acta Oncol 2017 Dec;56(12):1665-1676. [doi: 10.1080/0284186X.2017.1342862] [Medline: 28686520]
38. Lengacher CA, Reich RR, Paterson CL, Ramesar S, Park JY, Alinat C, et al. Examination of broad symptom improvement resulting from mindfulness-based stress reduction in breast cancer survivors: a randomized controlled trial. J Clin Oncol 2016 Aug 20;34(24):2827-2834 [FREE Full text] [doi: 10.1200/JCO.2015.65.7874] [Medline: 27247219]
39. Zhang Q, Zhao H, Zheng Y. Effectiveness of mindfulness-based stress reduction (MBSR) on symptom variables and health-related quality of life in breast cancer patients-a systematic review and meta-analysis. Support Care Cancer 2019 Mar;27(3):771-781. [doi: 10.1007/s00520-018-4570-x] [Medline: 30488223]
40. Hoffman CJ, Ersser SJ, Hopkinson JB, Nicholls PG, Harrington JE, Thomas PW. Effectiveness of mindfulness-based stress reduction in mood, breast- and endocrine-related quality of life, and well-being in stage 0 to iii breast cancer: a randomized, controlled trial. J Clin Oncol 2012 Apr 20;30(12):1335-1342. [doi: 10.1200/jco.2010.34.0331]
41. Lindsay EK, Creswell JD. Mechanisms of mindfulness training: Monitor and Acceptance Theory (MAT). Clin Psychol Rev 2017 Feb;51:48-59 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.011] [Medline: 27835764]
42. Nho J, Kim S, Park M, Kweon S. Symptom clusters and quality of life in breast cancer survivors after cancer treatment in a tertiary hospital in Korea. Eur J Cancer Care (Engl) 2018 Nov;27(6):e12919. [doi: 10.1111/ecc.12919] [Medline: 30253019]
43. Massicotte V, Ivers H, Savard J. COVID-19 pandemic stressors and psychological symptoms in breast cancer patients. Curr Oncol 2021 Jan 08;28(1):294-300 [FREE Full text] [doi: 10.3390/curroncol28010034] [Medline: 33430131]
44. Toivonen K, Hermann M, White J, Speca M, Carlson LE. A mixed-method, multi-perspective investigation of barriers to participation in mindfulness-based cancer recovery. Mindfulness 2020 Aug 04;11(10):2325-2337. [doi: 10.1007/s12671-020-01461-z]

Abbreviations

FACT-B: Functional Assessment of Cancer Therapy-Breast
HRQoL: health-related quality of life
iMBCR: internet-delivered mindfulness-based cancer recovery
MBCR: mindfulness-based cancer recovery
MBI: mindfulness-based intervention
MBSR: mindfulness-based stress reduction
MDASI: MD Anderson Symptom Inventory
MDASI-C: MD Anderson Symptom Inventory, Chinese version

Edited by R Kukafka; submitted 08.06.22; peer-reviewed by SJC Soerensen , T Haddad, IC Hou; comments to author 18.08.22; revised version received 10.10.22; accepted 24.10.22; published 22.11.22.

Please cite as:
Wang L, Chen X, Peng Y, Zhang K, Ma J, Xu L, Liu Z, Liu L, Luo Y, Gu C
Effect of a 4-Week Internet-Delivered Mindfulness-Based Cancer Recovery Intervention on the Symptom Burden and Quality of Life of Patients With Breast Cancer: Randomized Controlled Trial
J Med Internet Res 2022;24(11):e40059
URL: https://www.jmir.org/2022/11/e40059
doi:10.2196/40059
PMID:36413385

©Luyao Wang, Xing Chen, Yueyang Peng, Kun Zhang, Jun Ma, Lin Xu, Zixuan Liu, Li Liu, Yang Luo, Can Gu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of a Smartphone-Based Out-of-Hospital Screening App for Neonatal Hyperbilirubinemia on Neonatal Readmission Rates and Maternal Anxiety: Randomized Controlled Trial

Qin Yan[1*], PhD; Yanhong Gong[1*], PhD; Qing Luo[2,3*], MS; Xiaoxv Yin[1], PhD; Ling Yang[4], MM; Honglin Wang[5], MM; Juan Feng[5], MM; Kaihui Xing[4], MM; Yan Huang[2], MM; Chuican Huang[2], MM; Lichun Fan[2], PhD

[1]Department of Social Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Child Heath Care, Hainan Women and Children's Medical Center, Haikou, China

[3]Department of Child Health Care, National Center for Women and Children's Health, Chinese Center for Disease Control and Prevention, Beijing, China

[4]Department of Neonatology, Hainan Women and Children's Medical Center, Haikou, China

[5]Obstetrical Department, Hainan Women and Children's Medical Center, Haikou, China

[*]these authors contributed equally

Corresponding Author:
Lichun Fan, PhD
Department of Child Heath Care, Hainan Women and Children's Medical Center
No.75 South of Longkun Road
Haikou, 570206
China
Phone: 86 18689852199
Fax: 86 18689852199
Email: flc007@hainmc.edu.cn

Abstract

Background: Neonatal hyperbilirubinemia is one of the leading causes of neonatal readmission—especially severe hyperbilirubinemia and its complications—and it influences disease burden as well as neonatal and maternal health. Smartphones have been shown to have satisfactory accuracy in screening neonatal bilirubin levels, but the impact of this technology on neonatal health care service and maternal health outcomes is still unknown.

Objective: The aim of this study was to evaluate the impact of a smartphone-based out-of-hospital neonatal jaundice screening program on neonatal readmission rates for jaundice and related maternal anxiety.

Methods: This was a 2-arm, unblinded, randomized controlled trial with 30 days of intervention and follow-up periods. From August 2019 to August 2020, healthy mother-infant dyads were recruited on-site from 3 public hospitals in Hainan, China. Intervention group mothers used the smartphone app to routinely monitor neonatal jaundice at home under the web-based guidance of pediatricians. Control group participants received routine care. The primary study outcome was the neonatal readmission rate due to jaundice within 30 days of the first hospital discharge. The secondary outcome was the maternal anxiety score associated with neonatal jaundice. The data were collected through a self-assessed questionnaire. All participants were included in the analysis (intention-to-treat).

Results: In this study, 1424 mother-infant dyads were recruited, comprising 1424 mothers and 1424 newborns. The median age of the mothers was 29 (IQR 26-32) years, and there were 714 (50.1%) male neonates. These mother-infant dyads were randomly assigned to the intervention group and the control group, with 712 dyads in each group; only 1187 of these dyads completed the follow-up. We found that the adjusted 30-day neonatal readmission rate due to jaundice reduced by 10.5% (71/605, 11.7% vs 141/582, 24.2%; 95% CI 5%-15.9%; odds ratio 0.4, 95% CI 0.3-0.5; P<.001) and the relevant maternal anxiety mean score decreased by 3.6 (95% CI –4.4 to –2.8; β=–3.6, 95% CI –4.5 to –2.8; P<.001) in the intervention group compared to those in the routine care group.

Conclusions: Our study shows that the smartphone-based out-of-hospital screening method for neonatal hyperbilirubinemia decreased the neonatal readmission rate within 30 days from the first discharge and improved maternal mental health to some degree, thus demonstrating the usefulness of this screening app for follow-up in pediatric care.

XSL•FO
RenderX

Trial Registration: China Clinical Trial Registration Center, ChiCTR2100049567; http://www.chictr.org.cn/showproj.aspx?proj=64245

(J Med Internet Res 2022;24(11):e37843) doi:10.2196/37843

KEYWORDS

eHealth; mHealth; mobile apps; maternal anxiety; neonatal jaundice; neonatal readmission; neonatal screening; mobile phone

Introduction

Neonatal hyperbilirubinemia (or neonatal jaundice) is a clinical condition in which the skin, sclera, and mucous membranes become yellowish as a result of elevated serum bilirubin levels caused by abnormalities in bilirubin metabolism [1]. It is a common condition in the neonatal period, particularly in the first 2 weeks of birth, and influences approximately 60%-80% of newborns globally [2]. About 8%-9% of infants are diagnosed with severe hyperbilirubinemia in the first week of life, which affects about 1.1 million newborns annually [3,4]. Some cases may develop into acute bilirubin encephalopathy or nuclear jaundice, causing irreversible neurological damage, and even threatening the life of the child [5,6]. Neonatal hyperbilirubinemia is one of the leading causes of neonatal readmission [7]. In most cases, neonatal jaundice is physiological and always subsides spontaneously [8], but patients with severe hyperbilirubinemia and its complications should be hospitalized [9]. In addition, neonatal jaundice not only increases neonatal pain but also raises the stress levels of the parents, especially the mother [7]. Coupled with their lack of knowledge about this disease and the nursing methods, this disease may have a great impact on infant rehospitalization [10,11].

Early identification of high-risk neonates, appropriate follow-ups, and timely intervention contribute to preventing the deterioration of neonatal hyperbilirubinemia [12], thereby reducing neonatal readmission to hospitals and the consumption of health care resources [13]. Although hyperbilirubinemia screening recommendations should be based on risk factors according to the American Academy of Pediatrics [14], they do not obviate the need for universal evaluation of jaundice in the neonatal period [15]. The increasingly short postpartum length of hospital stay of newborns make follow-up monitoring after discharge to home extremely important [16]. The gold standard for the diagnosis of neonatal jaundice is total serum bilirubin (TSB) levels, but it requires specialized testing equipment in the hospital, as invasive procedures increase infant pain [17]. Transcutaneous bilirubinometry (TcB) measurement has been proven to be feasible and noninvasive, but the high cost of TcB meters limit its wide application, particularly in out-of-hospital settings [18,19]. Therefore, it is usually difficult to perform TSB and TcB measurements for most families. Concurrently, the serious underfunding of the health workforce in most low-income countries [20] and the global shortage of health resources, especially pediatricians, have made postpartum follow-up carried out by health care professionals challenging. Therefore, it is necessary to investigate a convenient way to screen neonatal jaundice that does not entirely rely on medical professionals but can be performed by parents at home at any time.

Mobile communication devices have been widely used in the health care sector [21]. Such emerging information technologies may help overcome the barriers existing in the routine monitoring of neonatal jaundice. Several clinical trials have found that smartphone-based jaundice testing is accurate and cost-effective in identifying neonates with high levels of TSB [7,22-25]. However, mobile technology has not yet been formally used in jaundice screening practices. So far, there is no high-quality evidence to confirm the actual effectiveness of out-of-hospital jaundice screening by using smartphone apps, such as the effects on neonatal health care services and maternal health outcomes. On the basis of routine care, we have established a smartphone-based, family-physician collaborative mechanism for an out-of-hospital jaundice screening program. We performed a randomized controlled trial (RCT) to evaluate whether this smartphone-based intervention could reduce the 30-day neonatal readmission rates and maternal anxiety compared with the routine care.

Methods

Study Design

This study was an unblinded, parallel-assignment RCT conducted at 3 public hospitals in Hainan, China from August 2019 to September 2020. Participants were randomly assigned to either the smartphone-based intervention group or the routine care control group and were followed up for 30 days (Multimedia Appendix 1).

Ethics Approval

The ethics committee of the Hainan Women and Children's Medical Center in Hainan approved the trial protocol (institutional review board approval HNWCMC201605), and all participants provided written informed consent. The trial registration number is ChiCTR2100049567 in the China Clinical Trial Registration Center database.

Study Participants

Neonates born in Hainan Women and Children's Medical Center, Wanning People's Hospital, and Chengmai People's Hospital and their mothers were recruited as participants. All mother-infant dyads met the hospital discharge criteria as judged by the bedside doctor, and they were about to be discharged home. The eligible mother was at least 16 years old with clear understanding and communication skills, owned a smartphone, and was able to use it. Exclusion criteria included multiple births (eg, twins); neonates with congenital disease, perinatal asphyxia, neonatal infection, hemolytic disease (positive Coombs test), or other serious organ damage; or family not living in Hainan

XSL•FO
RenderX

within the study period. Qualified mother-infant dyads were identified and screened face-to-face by trained obstetric nurses prior to discharge from the hospital. All enrolled persons were aware of the trial objective and process, and written informed consent was obtained from mothers at study entry.

Randomization and Masking

After baseline information was collected, the trained obstetric nurses carried out simple randomization (draw of lots) and allocated each mother-infant dyad to the intervention and control groups. Because of the nature of the intervention, it was not possible to blind participants or the pediatric nurses who were responsible for participant recruitment and follow-up.

Comparison Condition

Participants randomized to the control group received routine care. Under the National Basic Public Health Service Program in China [26], the primary health care institutions conduct newborn home visits (within 1 week of hospital discharge) and full-term health management for newborns (28-30 days after birth). Neonates with high risk factors such as low birth weight and prematurity are visited more frequently according to the actual situation.

Intervention

Participants randomized to the intervention group received routine care and the smartphone-based, family-physician collaborative neonatal jaundice screening program, that is, under the remote guidance of a pediatrician, mothers routinely monitored neonatal jaundice at home by using a smartphone app. The jaundice mobile monitoring app used in this study is a publicly available, free downloadable software in China that allows remote, noninvasive, and self-service jaundice monitoring and early warnings. It is used in combination with a jaundice colorimetric card, which is required to be placed on the chest of the child. After judging the image quality, light environment, and skin area, the app will automatically scan and take photos to obtain a clear image of the newborn skin. Once the cloud server receives the image data, the jaundice value will be automatically calculated and displayed (automated image-based bilirubin test), along with an indication of the risk level of neonatal jaundice (low risk, 2.6-10.1 mg/L; medium risk, 10.2-17.1 mg/L; high risk: ≥17.2 mg/L). A clinical trial conducted to evaluate the screening accuracy of this app found strong concordance between automated image-based bilirubin test and TSB levels (r=0.97) [22].

Intervention group mothers received free appropriative jaundice colorimetric cards and installed the jaundice mobile monitoring app on their smartphones at hospital discharge. They were instructed by trained obstetric nurses about the data uploading method. They were required to detect neonatal bilirubin values by using the app, when needed, after returning home. We assigned each mother a pediatrician from the hospital where the delivery took place and bound the mother's and the doctor's versions of the jaundice mobile monitoring app together; both parties could use the app platform for communication and consultation. Pediatricians would be able to access the dynamic changes and risk prompts of neonatal bilirubin values synchronously, so as to judge the infants' situation and guide the mother to take the correct preventive measures. If neonatal bilirubin level was considered to be at a high-risk level or the condition was quite severe, the pediatrician would alert the mother to bring the child to the hospital for further examination and specialist treatment. In addition, pediatricians would supervise and remind mothers to measure newborn infants' jaundice level every day through the app.

Outcome Measures

On the 31st day after the mother-infant dyad was discharged from the hospital, the maternal and neonatal outcomes were collected by trained obstetric nurses through telephone follow-up, and the last questionnaire survey was completed in September 2020. The primary outcome was the neonatal readmission rate due to jaundice, which was defined as the ratio of the number of neonates readmitted to hospitals for jaundice within 30 days of the first discharge to the total number of neonates in each group. Mothers were asked to answer, "Has your baby gone to the hospital again because of jaundice in the past 30 days?" and the corresponding options were yes or no. The secondary outcome was the maternal anxiety score associated with neonatal jaundice, which was measured by a self-designed scale (Multimedia Appendix 2). The scale consisted of 10 items with a 4-level scoring method and from "never happened" to "always happened;" the scores were coded with 1 to 4 points, respectively. The higher the total score (range 10-40 points), the more serious was the anxiety. It has good internal consistency, with Cronbach α=.91. Questions were answered by mothers whose neonates developed jaundice after discharge to home. In addition, the intervention group mothers were also asked to rate the jaundice mobile monitoring app in 5 aspects (ie, convenience, trustworthiness, recommendation, generalizability, and satisfaction) to assess the acceptability of the app. At recruitment, maternal and neonatal basic demographic characteristics were collected through a self-assessed questionnaire. Maternal anxiety score associated with neonatal jaundice was also included, and the assessment tool was the same as that used for the secondary outcome. Mothers whose neonates developed jaundice at baseline could respond to that scale.

Sample Size

A sample size calculation was performed based on the primary outcome. Based on previous studies [27,28] and our hypothesis that the 30-day neonatal readmission rate due to jaundice would relatively reduce by 50% in this trial as well as considering about 10% attrition rate, we estimated that about 600 neonates were needed for each group to ensure that the difference could be detected at 80% power (α=.05, 2-sided), that is, there was a total requirement of 1200 mother-infant dyads at least.

Statistical Analysis

All analyses followed the intention-to-treat principle, including all the randomly assigned mother-infant dyads. We examined the distributions of the baseline characteristics by using descriptive statistics and Pearson chi-square test; 2-sided Student t test and Wilcoxon rank sum test were applied to compare the difference between the 2 groups. We conducted analyses using a binary logistic regression model for primary outcomes and a

XSL•FO
RenderX

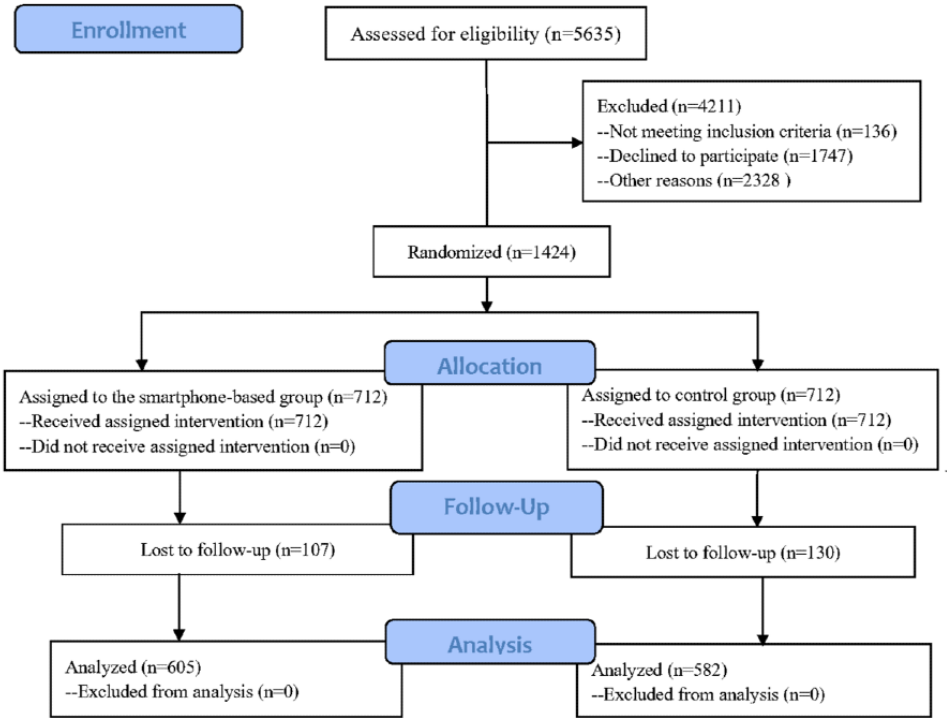
multiple linear regression model for secondary outcomes and reported the odds ratio (OR), regression coefficient (β), and 95% CIs. We adjusted all the models for observed baseline maternal and neonatal characteristics (ie, maternal age, nationality, education level, employment status, medical insurance, residence, convenience of access to health care, economic status and neonatal gender, gestational age, birth weight, feeding patterns, parity, status of jaundice before discharge, and delivery method). We added some post hoc analyses. First, based on least squares regression and identity link function, the generalized linear models were used to enable estimation of the adjusted risk difference for binary outcomes and adjusted mean difference for continuous outcomes. Besides, we fitted a linear mixed-effects repeated measures model to maternal anxiety scores at baseline and follow-up, thus capturing whether the improvement in the smartphone-based intervention group was greater than that in the control group. Sensitivity analyses were performed to assess the robustness of the primary and secondary outcome analysis. We performed multivariate imputation by using chained equations to impute all missing data, including outcomes that were not followed up. All statistical analyses were conducted using SAS version 9.4 (SAS Institute) and R version 4.0.3 (R Development Core Team), and all hypothesis tests were 2-sided with a significance level of .05.

Results

Overview

From August 2019 to August 2020, we screened 5635 mother-infant dyads in 3 public hospitals and 1424 dyads were eligible (Figure 1). Among them, 712 were randomly classified into the smartphone-based intervention group and 712 into the routine-care control group. On the 31st day of the intervention, 605 intervention group and 582 control group mother-infant dyads were followed up, as they were available for the analysis; others were lost because of refusal to follow-up, dropouts, etc, with a total retention rate of 83.4% (1187/1424).

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram for this study.



Baseline Data

The median age of the mothers was 29 (IQR 26-32) years; 32.3% (460/1424) of the mothers lived in the rural areas, and the mean maternal anxiety score associated with neonatal jaundice was 19.6 (SD 6.3) (Table 1). Of the 1424 neonates, 714 (50.1%) were males, 899 (63.1%) were born through natural delivery, and 1062 (74.6%) were breastfed (Table 2). The baseline characteristics were well-balanced between the control and the intervention groups. Moreover, similar demographic characteristics were observed between those who completed this trial and those who did not complete this trial (Multimedia Appendix 3).

XSL•FO
RenderX

Table 1. Baseline characteristics of the mothers.[a]

Variables	Overall (N=1424)	Intervention group (n=712)	Control group (n=712)
Age (years), median (IQR)	29 (26-32)	29 (26-32)	29 (26-32)
Nationality[b], n (%)			
Han nationality	1338 (94.1)	670 (94.4)	668 (93.8)
Minority	84 (5.9)	40 (5.6)	44 (6.2)
Education level, n (%)			
Junior high school or below	376 (26.4)	190 (26.7)	186 (26.1)
High school or technical secondary school	332 (23.3)	167 (23.5)	165 (23.2)
Junior college	338 (23.7)	170 (23.9)	168 (23.6)
Undergraduate or above	378 (26.5)	185 (26)	193 (27.1)
Employment status[c], n (%)			
Formally employed	634 (44.6)	315 (44.3)	319 (44.8)
Temporary employed	77 (5.4)	33 (4.6)	44 (6.2)
Unemployed	712 (50)	363 (51.1)	349 (49.1)
Medical insurance[d], n (%)			
Yes	1254 (88.2)	621 (87.5)	633 (88.9)
No	168 (11.8)	89 (12.5)	79 (11.1)
Residence, n (%)			
Urban	964 (67.7)	479 (67.3)	485 (68.1)
Rural	460 (32.3)	233 (32.7)	227 (31.9)
Convenience of access to health care[e], n (%)			
Yes	1246 (87.6)	616 (86.8)	630 (88.5)
No	176 (12.4)	94 (13.2)	82 (11.5)
Economic status[f], n (%)			
High income	51 (3.6)	25 (3.5)	68 (9.6)
Middle income	1230 (86.5)	612 (86.1)	618 (86.8)
Low income	142 (10)	74 (10.4)	26 (3.7)
Maternal anxiety score[g], mean (SD)	19.61 (6.3)	19.79 (6.3)	19.42 (6.3)

[a]*P* values were obtained using 2-sided Student *t* test or Wilcoxon rank sum test for continuous and χ^2 test for categorical variables; all *P*>.05.

[b]Nationality was unknown for 2 intervention group mothers.

[c]Employment status was unknown for 1 intervention group mother.

[d]Medical insurance was unknown for 2 intervention group mothers.

[e]Convenience of access to health care represents self-assessed convenience of access to health care facilities from residence by mothers; this was unknown for 2 intervention group mothers.

[f]Economic status was unknown for 1 intervention group mother.

[g]Including mothers whose children developed jaundice at baseline (604 in the intervention group, 608 in the control group, as shown in Table 2).

XSL•FO
RenderX

Table 2. Baseline characteristics of the neonates.[a]

Variables	Overall (N=1424)	Intervention group (n=712)	Control group (n=712)
Gender, n (%)			
Male	714 (50.1)	362 (50.8)	352 (49.4)
Female	710 (49.9)	350 (49.2)	360 (50.6)
Gestational age[b] (weeks), median (IQR)	39.1 (38.57-39.9)	39.1 (38.43-39.86)	39.1 (38.6-40.0)
Birth weight[c] (grams), median (IQR)	3200 (2900-3450)	3150 (2950-3450)	3200 (2900-3450)
Feeding patterns[d], n (%)			
Breastfeeding	1062 (74.6)	523 (73.5)	539 (75.8)
Nonbreastfeeding	28 (2)	13 (1.8)	15 (2.1)
Mixed feeding	333 (23.4)	176 (24.7)	157 (22.1)
Parity, n (%)			
First child	630 (44.2)	317 (44.5)	313 (44)
Second child	677 (47.5)	338 (47.5)	339 (47.6)
Third child or later	117 (8.2)	57 (8)	60 (8.4)
Status of jaundice before discharge[e], n (%)			
Not present	211 (14.8)	108 (15.2)	103 (14.5)
Present	1212 (85.2)	604 (84.8)	608 (85.5)
Delivery method, n (%)			
Natural delivery	899 (63.1)	476 (66.9)	482 (67.7)
Cesarean section	466 (32.7)	236 (33.2)	230 (32.3)

[a]*P* values were obtained using 2-sided Student *t* test or Wilcoxon rank sum test for continuous and χ^2 test for categorical variables; all *P*>.05.

[b]Gestational age was unknown for 1 intervention group neonate and 1 control group neonate.

[c]Birth weight was unknown for 2 intervention group neonates and 2 control group neonates.

[d]Feeding pattern was unknown for 1 control group neonate.

[e]Status of jaundice before discharge represents if the neonate developed jaundice at baseline; this was unknown for 1 control group neonate.

Primary Outcome

Within 30 days of the first hospital discharge, 68.1% (412/605) and 70.1% (408/582) of the neonates in the intervention and control groups, respectively, who were successfully followed up showed jaundice symptoms, with no statistical difference between the 2 groups. In comparison with the control group, the smartphone-based intervention group was significantly associated with a decrease (141/582, 24.2% vs 71/605, 11.7%, respectively; risk difference=12.5%, 95% CI 8.2%-16.8%) in the neonatal readmission rate due to jaundice (OR 0.4, 95% CI 0.3-0.6). After adjusting the model, our intervention remained observably effective in reducing the risk of readmission (OR 0.4, 95% CI 0.3-0.5), although the degree of reduction was smaller than before (risk difference=10.5%, 95% CI 5%-15.9%) (Table 3).

XSL•FO
RenderX

Table 3. Differences in the primary and secondary outcomes between the 2 groups.

Outcomes	Intervention group (n=605)	Control group (n=582)	Unadjusted		Adjusted	
			Odds ratio (95% CI) or β (95% CI)	Difference[a] (95% CI)	Odds ratio (95% CI) or β (95% CI)	Difference[a] (95% CI)
Primary outcome, n (%)	71 (11.7)	141 (24.2)	0.4 (0.3 to 0.6)[b]	12.5% (8.2% to 16.8%)	0.4 (0.3 to 0.5)[b]	10.5% (5% to 15.9%)
Secondary outcome[c], mean (SD)	16.8 (4.2)	20.5 (7.4)	–3.7 (–4.6 to –2.9)[b]	–3.7 (–4.6 to –2.9)	–3.6 (–4.5 to –2.8)[b]	–3.6 (–4.4 to –2.8)

[a]Difference represents risk difference for primary outcome and mean difference for secondary outcome. Risk difference represents the absolute value of the difference in the neonatal readmission rates between the 2 groups. Mean difference represents the difference in the mean maternal anxiety scores between the 2 groups.

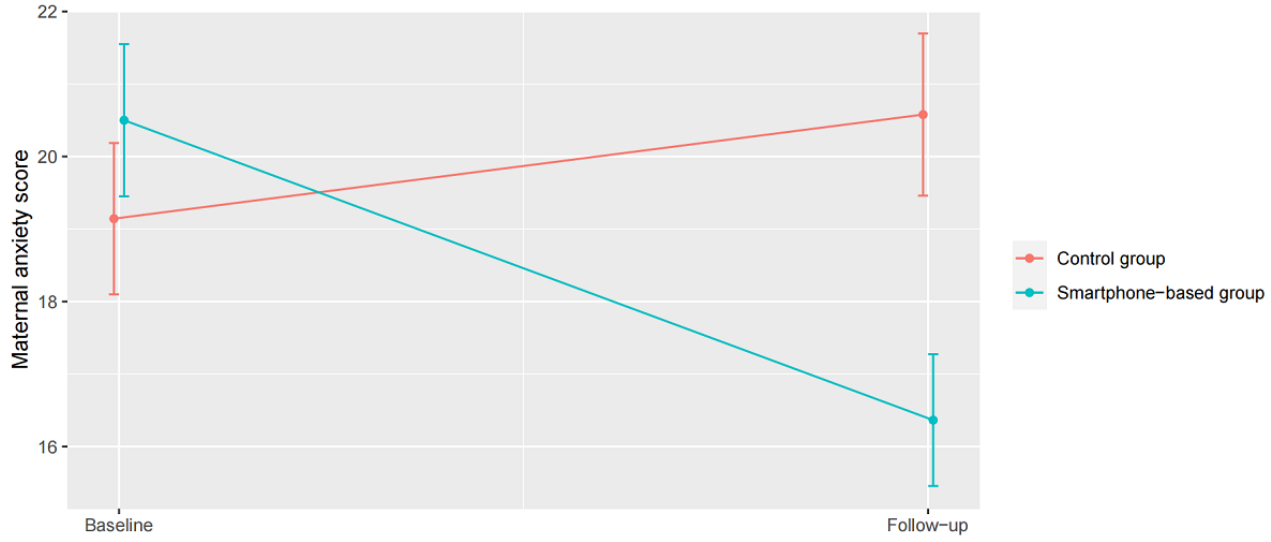
[b]*P*<.001 *P* values were from the odds ratio of the binary logistic regression model and β values were of the multiple linear regression model, with the control group as reference. The primary outcome was neonatal readmission. The secondary outcome was maternal anxiety score due to neonatal jaundice.

[c]Including mothers whose children developed jaundice symptoms after hospital discharge (intervention group, n=412; control group, n=408).

Secondary Outcome

Maternal anxiety scores associated with neonatal jaundice in the smartphone-based intervention group were apparently lower than those in the control group (mean difference=–3.7, 95% CI –4.6 to –2.9). The adjusted model showed a similar intervention effect (mean difference=–3.6, 95% CI –4.5 to –2.8) (Table 3). We performed a post hoc analysis to examine whether improvement in maternal anxiety could be explained by underlying temporal trends. There were 347 neonates in the intervention group and 348 in the control group who developed jaundice both at baseline and during the intervention period follow-up. For these participants, maternal anxiety is shown in Figure 2. From baseline to the 31st day of follow-up, the mean maternal anxiety score in the intervention group reduced by 4.1 (95% CI –4.9 to 3.4); however, that in the control group slightly increased by 1.4 (95% CI 0.5-2.4). After follow-up, the maternal anxiety score in the intervention group was significantly lower than that in the control group (adjusted difference=–4.2, 95% CI –5.1 to –3.3). The results in the sensitivity analysis were similar to the above main results (Multimedia Appendix 4).

Figure 2. Changes in the adjusted mean maternal anxiety score from baseline to the 31st day of follow-up (n=695). Points and error bars show the least squares mean and 95% CIs, respectively, that were derived from the adjusted linear mixed-effects repeated measures model.



Evaluation of the Jaundice Mobile Monitoring App

More than half of intervention group mothers rated the jaundice mobile monitoring app as relatively or very convenient (331/580, 57.1%) and relatively or very credible (295/597, 51%) in measuring their child's jaundice. Approximately 87.1% (504/579) of the mothers were willing to recommend this app to people in need, and 87.7% (508/579) held a positive attitude toward the necessity of promoting this app on a wide scale. In terms of satisfaction, only 11.1% (64/579) of the mothers showed dissatisfaction with the app (Multimedia Appendix 5).

Discussion

Principal Findings

In this smartphone-based RCT, we found that the intervention had a significant impact on reducing the neonatal readmission

XSL•FO
RenderX

for jaundice within 30 days of the initial discharge, wherein the readmission rate in the intervention group was 10.5% (95% CI 5%-15.9%) lower than that in the control group. In contrast with those in the control group, maternal anxiety symptoms induced by neonatal jaundice were also less severe in mothers who had installed and used the jaundice mobile monitoring app. Following-up monitoring of neonatal bilirubin levels is conducive to improving the outcome of neonatal hyperbilirubinemia, including reducing readmission rates [13,29,30]. One RCT implemented in China suggested that the neonatal readmission rate for hyperbilirubinemia was lower when neonatal TcB was regularly monitored by neonatologists compared with the control group wherein regular monitoring was absent (1.3% vs 3.3%, respectively) [31]. Another Australian study used TcB screening and visual assessment to detect the bilirubin levels of apparently healthy newborns before discharge and found that the former was capable of controlling neonatal readmission risk for hyperbilirubinemia (risk ratio 0.25, 95% CI 0.14-0.46) [27]. These studies were mostly based on TcB measurements, wherein parents and infants were asked to head to the follow-up clinic of the hospital or primary health care center on a specified date. The smartphone-based intervention in this trial emphasized more on the screening for neonatal jaundice at home, which may decrease the physical fatigue in mothers caused by repeated visits to the hospital. Our finding was consistent with the above studies basically [27,31], with positive effects in lessening neonatal readmission.

According to some surveys, the shorter the average length of postpartum hospital stay, the higher was the risk of 30-day neonatal readmission [32-34]. One possible reason for this is that both the time spent on neonatal health monitoring and the time parents received guidance from health care professionals are shortened due to the reduced length of postpartum hospital stay, with the consequence that parents may not have sufficient knowledge and confidence to cope with child's various health problems and thus increasing the corresponding health investment, such as hospital readmission [33]. Neonatal bilirubin levels usually peak at 96 hours of life when most infants have returned home, and jaundice symptoms may appear again after subsiding [35]. Nearly 70% of the newborns in our study developed jaundice after hospital discharge. Therefore, subsequent screening of neonatal jaundice is extremely necessary for mother and child. The smartphone-based intervention allowed mothers to observe the child's bilirubin levels at home at any time and acquire professional web-based guidance from pediatricians, which played an essential role in boosting parental confidence and thus avoiding some unnecessary hospital admissions.

Our intervention has proved to be effective in maintaining maternal mental health. Other studies have shown that family-centered parent-involved daily neonatal health care is positively associated with clinical benefits for infants (decreased readmission rate) and improved parental outcomes (decreased anxiety, depression, and stress) [36,37]. In addition, reducing neonatal readmission rate prevents separation of the mother and infant, thereby relieving the maternal emotional stress associated with it [13]. The mother's psychological status, in turn, affects the risk of infant rehospitalization. A previous study has shown that mothers with high anxiety traits were nearly 3 times more likely to take their babies to the hospital than their counterparts [11]. Thus, psychosocial support for mothers could reduce the readmission rates of healthy newborns [38]. The intervention in this study eased maternal negative emotions in the face of neonatal jaundice, which might control newborn readmission indirectly and allow them to experience less pain. It should be noted that mothers who did not receive the intervention had significantly higher anxiety scores after hospital discharge compared to baseline, which possibly, in part, could be because they were away from their health care providers and not able to keep track of the child's health accurately. This indicates that we should enhance the maternal awareness of neonatal jaundice and provide them with adequate psychological comfort in the future management of neonatal hyperbilirubinemia.

The application of the smartphone app in the monitoring of neonatal jaundice fits in with the widespread use of mobile technology currently; the global internet usage rate has reached 65.6% as of March 31, 2021 [39]. The previous cost-benefit analysis found that for every ¥1 invested in the smartphone-based out-of-hospital screening program for neonatal jaundice, ¥18.76 was saved in the treatment cost of postdischarge neonatal jaundice (the average conversion rate at the time of the study in 2021 was US $1=¥6.4515) [40]. The low cost and high accessibility of the jaundice mobile monitoring app make this technology feasible for patients with medical conditions and lack of access to treatment in resource-poor regions. Moreover, this family-centered jaundice monitoring app may contribute to alleviating the shortage of pediatric specialists, thereby promoting the implementation of a basic public health service program.

Although this jaundice mobile monitoring app can be regarded as a good auxiliary screening tool for neonatal jaundice in some cases, there is greater scope for improving its acceptability, because the proportion of mothers who explicitly expressed satisfaction with the app was less than 50%. The potential reason for this low satisfaction rate may be that the app was developed by a non–health care provider, leading to some concerns about the accuracy and credibility of its information. In this regard, we suggest that on the premise of ensuring the quality of jaundice screening technology, telemedicine guidance from health care professionals should be facilitated and related units should optimize their publicity and promotion strategies.

Strengths and Limitations

This study, to our knowledge, is the first RCT to analyze the effects of a smartphone app on neonatal health resource utilization and maternal mental health. Another strength of this study is the inclusion of large samples and enough statistical power to test the effects of the intervention.

Although our research results have important clinical implications, several limitations should be recognized. First, as we did not restrict the hospitals to which participants were readmitted, it was not possible to obtain and check all neonates' visit records in the 3 study hospitals; therefore, self-reported data were uniformly used in all outcomes. Although some recall bias was inevitable, the short follow-up period (only 30 days) and the fact that parents usually attach importance to neonatal

XSL•FO
RenderX

health and remember relatively well whether their child visited a doctor again minimized the recall bias. Second, due to the unavailability of the return visit records, we could not assess whether our intervention increased the readmission rates related to worsening neonatal hyperbilirubinemia, despite a significant drop in the overall rate. However, our intervention approaches are unlikely to delay admission to hospital for serious jaundice under the guidance of pediatricians. Finally, the trial was limited by the study site in comparison to that in other reports [27,28,31]. The 30-day neonatal readmission rate due to jaundice was higher in our study compared to that reported in previous studies [13,27,28,41]. This is partly due to inconsistent postdischarge follow-up periods across studies and due to the high prevalence of neonatal hyperbilirubinemia in Hainan, China—a region with a population with severe erythrocyte glucose-6-phosphate dehydrogenase deficiency [42,43]. This factor may affect the applicability of our findings to other populations. Therefore, it is necessary to validate the effectiveness of this intervention in a larger sample or in a cross-cultural context.

Conclusion

Our study showed that a smartphone-based jaundice screening app was effective for decreasing the risk of neonatal readmission due to jaundice within 30 days of discharge from hospital and the related maternal anxiety. Smartphones are becoming increasingly used around the world widely, as they are highly accessible and can act as low-cost, self-service testing tools for measuring neonatal bilirubin levels. More studies are needed to analyze the effectiveness of this intervention and to ensure large-scale implementation in other settings, subject to adequate resources.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant 71663016), Hainan Province Clinical Medical Center, and the Excellent Talent Team of Hainan Province (QRCBT202121). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication. We appreciate all the medical staff in the Hainan Women and Children's Medical Center, Wanning People's Hospital, and Chengmai People's Hospital for their active cooperation in this study.

Authors' Contributions

QY, YG, and QL contributed equally. LF and YG are joint corresponding authors. LF, XY, YG, QL, and QY conceptualized and designed the study. QL, HW, JF, KX, YH, CH, and LF collected the data, which were supervised by LF and YG. QY, QL, and YG analyzed and interpreted the data. QY, QL drafted the paper, QY, YG mainly completed the paper revision. All authors involved in revising the manuscript for important intellectual content and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 802 KB - jmir_v24i11e37843_app1.pdf]

Multimedia Appendix 2
Items of the self-designed maternal anxiety due to neonatal jaundice scale.
[PDF File (Adobe PDF File), 55 KB - jmir_v24i11e37843_app2.pdf]

Multimedia Appendix 3
Differences in the demographic characteristics between participants who completed and did not complete the trial.
[PDF File (Adobe PDF File), 63 KB - jmir_v24i11e37843_app3.pdf]

Multimedia Appendix 4
Sensitivity analyses.
[PDF File (Adobe PDF File), 69 KB - jmir_v24i11e37843_app4.pdf]

Multimedia Appendix 5
Evaluation of the jaundice mobile monitoring app.
[PDF File (Adobe PDF File), 68 KB - jmir_v24i11e37843_app5.pdf]

References

XSL•FO
RenderX

1. Olusanya BO, Kaplan M, Hansen TWR. Neonatal hyperbilirubinaemia: a global perspective. The Lancet Child & Adolescent Health 2018 Aug;2(8):610-620. [doi: 10.1016/s2352-4642(18)30139-1]
2. Mitra S, Rennie J. Neonatal jaundice: aetiology, diagnosis and treatment. Br J Hosp Med (Lond) 2017 Dec 02;78(12):699-704. [doi: 10.12968/hmed.2017.78.12.699] [Medline: 29240507]
3. Bhutani VK, Zipursky A, Blencowe H, Khanna R, Sgro M, Ebbesen F, et al. Neonatal hyperbilirubinemia and Rhesus disease of the newborn: incidence and impairment estimates for 2010 at regional and global levels. Pediatr Res 2013 Dec;74 Suppl 1:86-100 [FREE Full text] [doi: 10.1038/pr.2013.208] [Medline: 24366465]
4. Smitherman H, Stark AR, Bhutani VK. Early recognition of neonatal hyperbilirubinemia and its emergent management. Semin Fetal Neonatal Med 2006 Jun;11(3):214-224. [doi: 10.1016/j.siny.2006.02.002] [Medline: 16603425]
5. Fakhri M, Farhadi R, Mousavinasab N, Hosseinimehr SJ, Yousefi SS, Davoodi A, et al. Preventive effect of purgative manna on neonatal jaundice: A double blind randomized controlled clinical trial. J Ethnopharmacol 2019 May 23;236:240-249. [doi: 10.1016/j.jep.2019.03.009] [Medline: 30853647]
6. Wang J, Yin J, Xue M, Lyu J, Wan Y. Roles of UGT1A1 Gly71Arg and TATA promoter polymorphisms in neonatal hyperbilirubinemia: A meta-analysis. Gene 2020 Apr 30;736:144409. [doi: 10.1016/j.gene.2020.144409] [Medline: 32007587]
7. Munkholm SB, Krøgholt T, Ebbesen F, Szecsi PB, Kristensen SR. The smartphone camera as a potential method for transcutaneous bilirubin measurement. PLoS One 2018;13(6):e0197938 [FREE Full text] [doi: 10.1371/journal.pone.0197938] [Medline: 29856793]
8. Moreno MA. Common questions about neonatal jaundice. JAMA Pediatr 2015 Mar;169(3):296. [doi: 10.1001/jamapediatrics.2014.2116] [Medline: 25730694]
9. Hameed NN, Na' Ma AM, Vilms R, Bhutani VK. Severe neonatal hyperbilirubinemia and adverse short-term consequences in Baghdad, Iraq. Neonatology 2011;100(1):57-63. [doi: 10.1159/000321990] [Medline: 21212697]
10. Batu ED, Yeni S, Teksam O. The factors affecting neonatal presentations to the pediatric emergency department. J Emerg Med 2015 May;48(5):542-547. [doi: 10.1016/j.jemermed.2014.12.031] [Medline: 25648054]
11. Vonderheid SC, Rankin K, Norr K, Vasa R, Hill S, White-Traut R. Health Care Use Outcomes of an Integrated Hospital-to-Home Mother-Preterm Infant Intervention. J Obstet Gynecol Neonatal Nurs 2016;45(5):625-638. [doi: 10.1016/j.jogn.2016.05.007] [Medline: 27505467]
12. Thanomsingh P. Clinical predictive score of predischarge screening for severe hyperbilirubinemia in late preterm and term infants. Pediatr Neonatol 2020 Aug;61(4):378-384 [FREE Full text] [doi: 10.1016/j.pedneo.2020.02.003] [Medline: 32156481]
13. Alkalay AL, Bresee CJ, Simmons CF. Decreased neonatal jaundice readmission rate after implementing hyperbilirubinemia guidelines and universal screening for bilirubin. Clin Pediatr (Phila) 2010 Sep;49(9):830-833. [doi: 10.1177/0009922810363728] [Medline: 20693521]
14. Muchowski KE. Evaluation and treatment of neonatal hyperbilirubinemia. Am Fam Physician 2014 Jun 01;89(11):873-878 [FREE Full text] [Medline: 25077393]
15. Wainer S, Parmar S, Allegro D, Rabi Y, Lyon M. Impact of a transcutaneous bilirubinometry program on resource utilization and severe hyperbilirubinemia. Pediatrics 2012 Jan;129(1):77-86. [doi: 10.1542/peds.2011-0599] [Medline: 22184646]
16. Campbell OMR, Cegolon L, Macleod D, Benova L. Length of Stay After Childbirth in 92 Countries and Associated Factors in 30 Low- and Middle-Income Countries: Compilation of Reported Data and a Cross-sectional Analysis from Nationally Representative Surveys. PLoS Med 2016 Mar;13(3):e1001972 [FREE Full text] [doi: 10.1371/journal.pmed.1001972] [Medline: 26954561]
17. Halder A, Pal SK, Banerjee M, Singh S, Adhikari A, Sarkar PK, et al. A Novel Whole Spectrum-Based Non-Invasive Screening Device for Neonatal Hyperbilirubinemia. IEEE J. Biomed. Health Inform 2019 Nov;23(6):2347-2353 [FREE Full text] [doi: 10.1109/jbhi.2019.2892946]
18. Neonatal jaundice. National Collaborating Centre for Women's and Children's Health (UK). London: Royal College of Obstetricians and Gynaecologists; 2010. [Medline: 22132434]
19. Maisels M, Bhutani V, Bogen D, Newman T, Stark A, Watchko J. Hyperbilirubinemia in the newborn infant > or =35 weeks' gestation: an update with clarifications. Pediatrics 2009;124(4):1193-1198. [doi: 10.1542/peds.2009-0329] [Medline: 19786452]
20. Hernandez-Peña P, Poullier JP, Van Mosseveld CJM, Van de Maele N, Cherilova V, Indikadahena C, et al. Health worker remuneration in WHO Member States. Bull World Health Organ 2013 Nov 01;91(11):808-815 [FREE Full text] [doi: 10.2471/BLT.13.120840] [Medline: 24347704]
21. Manning ML, Davis J, Sparnon E, Ballard RM. iPads, droids, and bugs: Infection prevention for mobile handheld devices at the point of care. Am J Infect Control 2013 Nov;41(11):1073-1076. [doi: 10.1016/j.ajic.2013.03.304] [Medline: 23816356]
22. Han M, Zhu S, Tong S, Qiu S, Gong L, Gao X. Analysis of monitoring neonatal jaundice detected by smartphone image-based bilirubin. Chinese Journal of Child Health Care 2021;29(05):550-554 [FREE Full text]
23. Outlaw F, Nixon M, Odeyemi O, MacDonald LW, Meek J, Leung TS. Smartphone screening for neonatal jaundice via ambient-subtracted sclera chromaticity. PLoS One 2020;15(3):e0216970 [FREE Full text] [doi: 10.1371/journal.pone.0216970] [Medline: 32119664]

XSL•FO
RenderX

24. Tabatabaee RS, Golmohammadi H, Ahmadi SH. Easy Diagnosis of Jaundice: A Smartphone-Based Nanosensor Bioplatform Using Photoluminescent Bacterial Nanopaper for Point-of-Care Diagnosis of Hyperbilirubinemia. ACS Sens 2019 Apr 26;4(4):1063-1071. [doi: 10.1021/acssensors.9b00275] [Medline: 30896150]
25. Taylor JA, Stout JW, de Greef L, Goel M, Patel S, Chung EK, et al. Use of a Smartphone App to Assess Neonatal Jaundice. Pediatrics 2017 Sep;140(3):e20170312 [FREE Full text] [doi: 10.1542/peds.2017-0312] [Medline: 28842403]
26. National basic public health service specification. National Health and Family Planning Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=d20c37e23e1f4c7db7b8e25f34473e1b [accessed 2022-03-05]
27. Okwundu C, Bhutani VK, Smith J, Esterhuizen TM, Wiysonge C. Predischarge transcutaneous bilirubin screening reduces readmission rate for hyperbilirubinaemia in diverse South African newborns: A randomised controlled trial. S Afr Med J 2020 Feb 26;110(3):249-254 [FREE Full text] [doi: 10.7196/samj.2020.v110i3.14186] [Medline: 32657704]
28. VanderWal B, Kyle CL. Decreasing Newborn Readmissions for Hyperbilirubinemia. Journal of Obstetric, Gynecologic & Neonatal Nursing 2015 Jun;44:S25-S25. [doi: 10.1111/1552-6909.12698]
29. Bhutani VK, Johnson LH, Schwoebel A, Gennaro S. A systems approach for neonatal hyperbilirubinemia in term and near-term newborns. J Obstet Gynecol Neonatal Nurs 2006;35(4):444-455. [doi: 10.1111/j.1552-6909.2006.00044.x] [Medline: 16881988]
30. Petersen J, Okorodudu A, Mohammad A, Fernando A, Shattuck K. Association of transcutaneous bilirubin testing in hospital with decreased readmission rate for hyperbilirubinemia. Clin Chem 2005 Mar 2005;51(3):540-544. [doi: 10.1373/clinchem.2004.037804]
31. Wang Q, Xing Y, Piao M, Liu H, Liu Y, Liu H. Close monitoring of transcutaneous bilirubin ensures better outcome in neonates with hyperbilirubinemia. Chin J Perinat Med 2016;19(11):808-812 [FREE Full text]
32. Harron K, Gilbert R, Cromwell D, Oddie S, van der Meulen J. Newborn Length of Stay and Risk of Readmission. Paediatr Perinat Epidemiol 2017 May;31(3):221-232 [FREE Full text] [doi: 10.1111/ppe.12359] [Medline: 28418622]
33. Sievertsen HH, Wüst M. Discharge on the day of birth, parental response and health and schooling outcomes. J Health Econ 2017 Sep;55:121-138. [doi: 10.1016/j.jhealeco.2017.06.012] [Medline: 28743535]
34. Jones E, Taylor B, MacArthur C, Bradshaw S, Hope L, Cummins C. Early Postnatal Discharge for Infants: A Meta-analysis. Pediatrics 2020 Aug 21;146(3):e20193365 [FREE Full text] [doi: 10.1542/peds.2019-3365] [Medline: 32826340]
35. Fouzas S, Mantagou L, Skylogianni E, Mantagos S, Varvarigou A. Transcutaneous bilirubin levels for the first 120 postnatal hours in healthy neonates. Pediatrics 2010;125(1):e52-e57 [FREE Full text] [doi: 10.1542/peds.2009-0403] [Medline: 20008429]
36. Ding X, Zhu L, Zhang R, Wang L, Wang T, Latour JM. Effects of family-centred care interventions on preterm infants and parents in neonatal intensive care units: A systematic review and meta-analysis of randomised controlled trials. Aust Crit Care 2019 Jan;32(1):63-75. [doi: 10.1016/j.aucc.2018.10.007] [Medline: 30554939]
37. O'Brien K, Bracht M, Robson K, Ye XY, Mirea L, Cruz M, et al. Evaluation of the Family Integrated Care model of neonatal intensive care: a cluster randomized controlled trial in Canada and Australia. BMC Pediatr 2015 Dec 15;15:210 [FREE Full text] [doi: 10.1186/s12887-015-0527-0] [Medline: 26671340]
38. Barilla D, Marshak H, Anderson S, Hopp J. Postpartum follow-up: can psychosocial support reduce newborn readmissions? MCN Am J Matern Child Nurs 2010;35(1):33-39 [FREE Full text] [doi: 10.1097/01.nmc.0000366808.75079.cc] [Medline: 20032758]
39. World internet usage and population statistics 2021 year-Q1 estimates. Internet World Stats. 2021. URL: https://internetworldstats.com/stats.htm [accessed 2021-08-24]
40. Luo Q, Dong C, Wu W, Fan L. Cost-benefit analysis of screening for neonatal hyperbilirubinemia using mobile phone application in different modes. Chinese Journal of Social Medicine. 2021 Dec 26. URL: https://tinyurl.com/5edcsk6z [accessed 2022-11-03]
41. De Luca D, Carnielli VP, Paolillo P. Neonatal hyperbilirubinemia and early discharge from the maternity ward. Eur J Pediatr 2009 Sep;168(9):1025-1030. [doi: 10.1007/s00431-009-0969-1] [Medline: 19277705]
42. Lin F, Lou Z, Xing S, Zhang L, Yang L. The gene spectrum of glucose-6-phosphate dehydrogenase (G6PD) deficiency in Guangdong province, China. Gene 2018 Dec 15;678:312-317. [doi: 10.1016/j.gene.2018.07.068] [Medline: 30077011]
43. Yu MW, Hsiao KJ, Wuu KD, Chen CJ. Association between glucose-6-phosphate dehydrogenase deficiency and neonatal jaundice: interaction with multiple risk factors. Int J Epidemiol 1992 Oct;21(5):947-952. [doi: 10.1093/ije/21.5.947] [Medline: 1468858]

Abbreviations

OR: odds ratio
RCT: randomized controlled trial
TcB: transcutaneous bilirubinometry
TSB: total serum bilirubin

XSL•FO
RenderX

Edited by R Kukafka; submitted 12.03.22; peer-reviewed by UH Mohamad, H Mehdizadeh; comments to author 14.07.22; revised version received 16.08.22; accepted 06.09.22; published 23.11.22.

Please cite as:
Yan Q, Gong Y, Luo Q, Yin X, Yang L, Wang H, Feng J, Xing K, Huang Y, Huang C, Fan L
Effects of a Smartphone-Based Out-of-Hospital Screening App for Neonatal Hyperbilirubinemia on Neonatal Readmission Rates and Maternal Anxiety: Randomized Controlled Trial
J Med Internet Res 2022;24(11):e37843
URL: https://www.jmir.org/2022/11/e37843
doi:10.2196/37843
PMID:36416882

©Qin Yan, Yanhong Gong, Qing Luo, Xiaoxv Yin, Ling Yang, Honglin Wang, Juan Feng, Kaihui Xing, Yan Huang, Chuican Huang, Lichun Fan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Use of Smartphone Keystroke Dynamics to Passively Monitor Upper Limb and Cognitive Function in Multiple Sclerosis: Longitudinal Analysis

Ka-Hoo Lam[1], MD, MSc; James Twose[2], MSc; Birgit Lissenberg-Witte[3], PhD; Giovanni Licitra[2], PhD; Kim Meijer[2], PhD; Bernard Uitdehaag[1], MD, PhD; Vincent De Groot[4], MD, PhD; Joep Killestein[1], MD, PhD

[1]Department of Neurology, Amsterdam University Medical Centers (VU University Medical Center location), Amsterdam, Netherlands

[2]Neurocast BV, Amsterdam, Netherlands

[3]Department of Epidemiology and Data Science, Amsterdam University Medical Centers (VU University Medical Center location), Amsterdam, Netherlands

[4]Department of Rehabilitation Medicine, Amsterdam University Medical Centers (VU University Medical Center location), Amsterdam, Netherlands

Corresponding Author:
Ka-Hoo Lam, MD, MSc
Department of Neurology
Amsterdam University Medical Centers (VU University Medical Center location)
De Boelelaan 1117
Amsterdam, 1081HV
Netherlands
Phone: 31 204440717
Fax: 31 204440715
Email: k.lam1@amsterdamumc.nl

Abstract

Background: Typing on smartphones, which has become a near daily activity, requires both upper limb and cognitive function. Analysis of keyboard interactions during regular typing, that is, keystroke dynamics, could therefore potentially be utilized for passive and continuous monitoring of function in patients with multiple sclerosis.

Objective: To determine whether passively acquired smartphone keystroke dynamics correspond to multiple sclerosis outcomes, we investigated the association between keystroke dynamics and clinical outcomes (upper limb and cognitive function). This association was investigated longitudinally in order to study within-patient changes independently of between-patient differences.

Methods: During a 1-year follow-up, arm function and information processing speed were assessed every 3 months in 102 patients with multiple sclerosis with the Nine-Hole Peg Test and Symbol Digit Modalities Test, respectively. Keystroke-dynamics data were continuously obtained from regular typing on the participants' own smartphones. Press-and-release latency of the alphanumeric keys constituted the fine motor score cluster, while latency of the punctuation and backspace keys constituted the cognition score cluster. The association over time between keystroke clusters and the corresponding clinical outcomes was assessed with linear mixed models with subjects as random intercepts. By centering around the mean and calculating deviation scores within subjects, between-subject and within-subject effects were distinguished.

Results: Mean (SD) scores for the fine motor score cluster and cognition score cluster were 0.43 (0.16) and 0.94 (0.41) seconds, respectively. The fine motor score cluster was significantly associated with the Nine-Hole Peg Test: between-subject β was 15.9 (95% CI 12.2-19.6) and within-subject β was 6.9 (95% CI 2.0-11.9). The cognition score cluster was significantly associated with the Symbol Digit Modalities Test between subjects (between-subject β –11.2, 95% CI –17.3 to –5.2) but not within subjects (within-subject β –0.4, 95% CI –5.6 to 4.9).

Conclusions: Smartphone keystroke dynamics were longitudinally associated with multiple sclerosis outcomes. Worse arm function corresponded with longer latency in typing both across and within patients. Worse processing speed corresponded with higher latency in using punctuation and backspace keys across subjects. Hence, keystroke dynamics are a potential digital biomarker for remote monitoring and predicting clinical outcomes in patients with multiple sclerosis.

Trial Registration: Netherlands Trial Register NTR7268; https://trialsearch.who.int/Trial2.aspx?TrialID=NTR7268

XSL•FO
RenderX

(J Med Internet Res 2022;24(11):e37614) doi:10.2196/37614

KEYWORDS

multiple sclerosis; smartphone; mobile app; digital technology; keystroke dynamics; typing; upper extremity; cognition; outpatient monitoring

Introduction

In multiple sclerosis (MS), a vast number of disease-modifying therapies targeting disease activity are available, and therapies preventing (and potentially counteracting) disease progression are emerging [1-3]. Additional treatment modalities include nonpharmacological therapies, such as rehabilitative and cognitive therapies [4,5]. This wide array of expanding treatment options will increasingly lead to patient-centered disease management. The personalized treatment of MS would strongly benefit from early and improved recognition of disability progression or symptom onset. However, disease progression (ie, deterioration of neurological function independent of relapses) and newly occurring symptoms are often subtle in MS [6]. Additionally, the currently most widely used clinical measure in MS, the Expanded Disability Status Scale (EDSS) [7], assesses neurological function over a period spanning a year or almost a year and may need reassessment over time to confirm deterioration [6]. The Multiple Sclerosis Functional Composite (MSFC) consists of brief objective measurements in 3 important domains in MS: ambulatory, upper limb, and cognitive function. It was designed to complement the EDSS and improve sensitivity in capturing disease status [8]. Compared to the extensive implementation of the MSFC in clinical trials, it has been poorly incorporated into clinical practice, as clinical evaluations are too sporadic for the measure to be sensitive or provide meaningful temporal information for monitoring patients on the individual level [9].

The advent of digital devices allows for more continuous and more fine-grained measurements of biometrics that could be related to functioning in patients with MS. With the digitalization of society, smartphones have become widespread and part of everyday living. Consequently, keystroke dynamics (KD) from typing on smartphones has been investigated for quantifying disability in MS. KD encompasses quantitative metrics of keyboard interactions during regular typing. In our previous work, KD was found to be correlated with upper limb and cognitive function, and, to a lesser extent, overall disability, as measured with the EDSS [10]. Across a wide range of KD features and aggregation methods, KD was also found to reach adequate responsiveness to meaningful change in radiological disease activity, ambulatory function, and upper limb function over a period of 3 months [11]. Additionally, analysis of KD data using a nonlinear time-series approach identified potential indicators of clinical change [12]. Based on these previous findings, 2 keystroke clusters were derived, one specific to upper limb function and the other to cognition, since these 2 domains are most directly related to typing. In order to translate this new biomarker into clinical practice for monitoring upper limb function and cognition in MS, the association with clinical measures over time and within individual patients needs investigation.

Our objective was to investigate the longitudinal associations between KD features, passively derived from regular typing on a smartphone, and upper limb function and cognition in patients with MS. Additionally, we sought to differentiate these longitudinal associations for both between-subject differences and within-subject changes in order to enable disease monitoring on the individual-patient level.

Methods

Study Design and Participants

This was a prospective cohort study at the MS Center of the Amsterdam University Medical Centers (VU University Medical Center location). The study design and interim analyses have been reported previously [10,11]. In brief, after a baseline assessment (M_0), we followed the patients for 1 year, with clinical visits every 3 months (M_3, M_6, M_9, and M_{12}). During the study, participants used the Neurokeys keyboard app on their own smartphones [13]. Participants were patients with MS and were consecutively included in the study between August 2018 and December 2019 until a cohort size of 100 participants was reached. Patients were eligible if they were aged between 18 and 65 years, had a definite diagnosis of MS, had an EDSS score below 7.5, had access to a smartphone with the Android (5.0 or higher) or iOS (10 or higher) operating systems, had no visual or upper extremity deficits affecting regular smartphone use, and had no mood or sleep disorders impacting daily living (based on medical history-taking by a screening physician).

Ethical Considerations

The study received ethical approval from the Medisch Ethische Toetsingscommissie Vrije Universiteit medisch centrum (reference 2017.576) and conformed to legislation regarding data privacy and medical devices (Dutch Health and Youth Care Inspectorate; reference VGR2006948). All patients gave written informed consent. The study was registered as trial number NTR7268 at the Netherlands Trial Register.

Clinical Outcomes

Clinical outcomes for important aspects of MS were assessed, including clinically reported relapses, conventional magnetic resonance imaging (MRI) for disease activity, the EDSS, the MSFC, patient-reported outcomes, quantitative MRI, and optical coherence tomography for evaluation of domain-specific and overall disease severity and disease progression over time. As KD is most directly related to upper limb function and cognition, the current analysis focuses on the clinical assessments made every 3 months with the Nine-Hole Peg Test (NHPT) and Symbol Digit Modalities Test (SDMT). The NHPT is a measure of upper limb function that records the time needed to place, with a single hand, 9 pegs into 9 holes and then remove them [14]. The task is performed twice for each hand, and the 4 trials are averaged into a single score, with a higher score reflecting

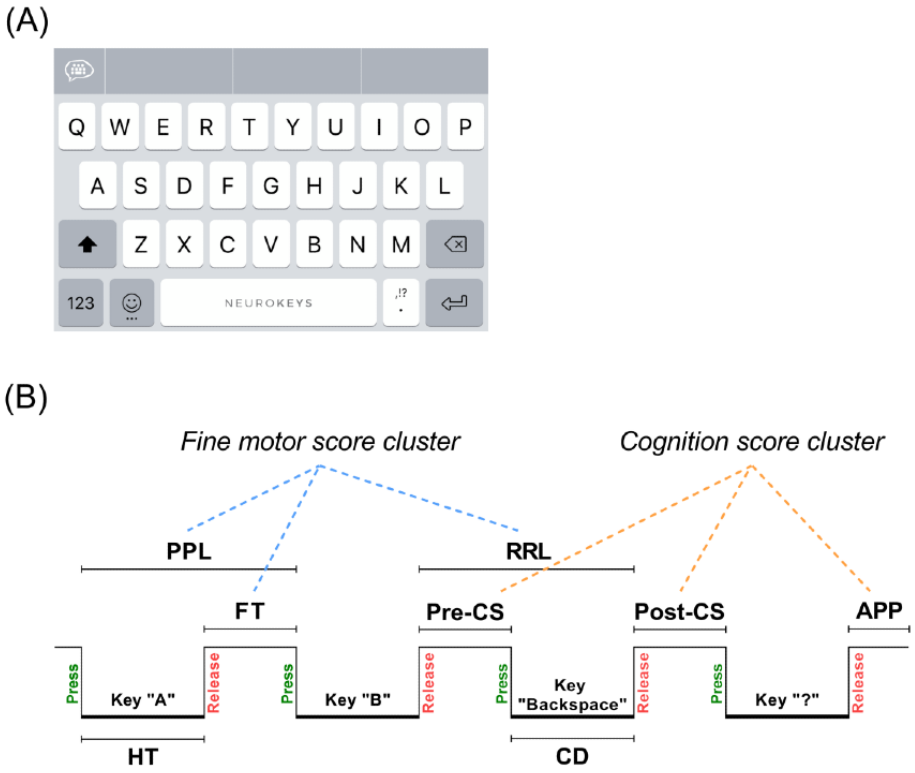
XSL•FO
RenderX

worse performance. The SDMT is a measure of information processing speed, the cognitive domain that is most commonly affected in MS and indicates overall cognitive functioning [15]. Using a key with 9 symbol-digit pairings, the number of correct digits corresponding to symbols during a 90-second trial is recorded as the total score [16]. A higher score reflects better performance.

KD and Keystroke Features

During the 1-year follow-up period, patients used the Neurokeys app (Neurocast BV) on their own smartphones [13]. The Neurokeys app replaces the native keyboard with a similar-looking keyboard (Figure 1A) that passively and continuously collects data on press-and-release typing events during everyday typing. From these keyboard interactions, keystroke features are derived based on key type (Figure 1B). For alphanumeric keys, the features include the latency between presses (press-press latency) and releases (release-release latency), the keypress time (hold time), and the time between keys (flight time). For the backspace key, derived features include latency prior to the use of the key (precorrection slowing), during use (correction duration) and after use (postcorrection slowing). Lastly, the time after a punctuation key was used was also derived (after-punctuation pause). A keystroke event count threshold of 50 events was used to remove days with insufficient data.

Figure 1. Overview of the Neurokeys keyboard (A) and a schematic representation of the keystroke dynamics features and clusters (B). APP: after-punctuation pause; CD: correction duration; FT: flight time; HT: hold time; post-CS: postcorrection slowing; PPL: press-press latency; pre-CS: precorrection slowing; RRL: release-release latency.



Construction of Keystroke Clusters

To compare the continuously collected keystroke data with clinical outcomes, the keystroke features were aggregated and clustered. First, the keystroke features were aggregated per day by the mean and median values, as both statistical measures summarize the data well and remain on the same unit scale (ie, seconds) to retain interpretability. Since mean and median values of the keystroke features were highly correlated, rather than discarding one, both summary values were averaged to reduce potential multicollinearity. Second, the fine motor score cluster (FMSC) and cognition score cluster (CSC) were derived based on the hypothesis that timing-related features (press-press latency, release-release latency, hold time, and flight time) are more related to fine motor skills, while error-related (precorrection slowing, correction duration, and postcorrection slowing) and paralinguistic (after-punctuation pause) features are more specific to events reflecting cognitive function. This concept-based clustering was then analyzed with principal component analysis and correlation analysis (Multimedia Appendices 1 - 3). Only features that contributed equally in the component analysis and were highly correlated (r>0.50) were included in the final cluster [17]. Finally, near the time of each clinical visit, 28-day (the 14 days before and after the clinical visit) and 14-day (the 7 days before and after the clinical visit) aggregation periods for the keystroke clusters was chosen for FMSC and CSC, respectively, since fine motor function can be considered more stable over time than cognitive function. The 28-day and 14-day periods for the keystroke clusters were aggregated by the mean value. Using these criteria, FMSC included press-press latency, release-release latency, and flight time, whereas CSC included precorrection slowing, postcorrection slowing, and after-punctuation pause.

XSL•FO
RenderX

Statistical Analysis

Analysis was performed with SPSS (version 26; IBM Corp) and R (version 4.0.3; R Foundation for Statistical Computing). Categorical data were summarized as the frequency and percentage. Numerical data were summarized as the mean and SD (or median and IQR or range if normally distributed). A linear mixed model analysis was used to determine the longitudinal association between KD clusters and clinical outcomes, so as to take into account clustering of repeated measurements within subjects [18]. Separate intercepts were estimated for each subject, over which a normal distribution was drawn. Then, the variance was estimated from that normal distribution and added to the model as a random intercept (ε), to adjust for repeated measurements within subjects, as follows: $Y = \beta_0 + \beta_1 X + \varepsilon$. For upper limb function, the dependent variable was the NHPT score, the independent variable was FMSC, and the covariates were age and sex. For information processing speed, the dependent variable was the SDMT score, the independent variable was CSC, and the covariates were age, sex, and level of education. Since there was a significant relationship between time and SDMT performance, most likely due to practice effects, an additional random intercept for time (in days) was added to the cognition model. This allowed varying intercepts based on time in order to account for practice effects and imbalances in time intervals between clinical visits across subjects [19].

Importantly, given that the effect estimates of a linear mixed model analysis in a cohort with repeated measures are overall effects (ie, effect estimates entangle both differences across subjects and changes within subjects over time), a "hybrid" linear mixed model analysis was performed to disentangle the between-subject and within-subject effects of the longitudinal association [20]. This was done by centering around the mean and calculating deviation scores at each clinical visit for each subject. The mixed model analysis was then performed with both the centered values and the deviation score of this centered value for each individual, as follows:  , where $\beta_{between}$ is the between-subject effect and β_{within} is the within-subject effect [18].

The output of all linear mixed models included effect estimates, 95% CIs, *P* values, and percentage explained variance. Covariates were considered relevant if the effect estimate between the dependent and independent variables changed by 10% or more after including the covariates into the model [21].

Results

A total of 102 patients with MS were included, of whom 91 completed the follow-up at M_{12}; 6 patients dropped out at M_3, 1 at M_6, 1 at M_9, and 3 at M_{12}. The demographic and clinical characteristics at baseline are summarized in Table 1. The patients had a mean age of 46.4 years, most were female (75/102, 73.5%), and most had the relapsing-remitting MS subtype (61/102, 59.8%). The median disease duration since diagnosis was 5.7 years and the median EDSS score was 3.5. The mean follow-up duration was 376.9 (SD 109.4) days. At M_{12}, the retention rate of patients with active keyboard use was 83.3% (85/102). Figure 2 shows the monthly retention rate and the average number of keystroke events per day. The clinical outcomes per visit and keystroke cluster data corresponding with each clinical visit are summarized in Table 2 and Multimedia Appendices 4 and 5. Part of the study follow-up coincided with the COVID-19 pandemic, which resulted in missing clinical visits, most prominently at M_6 and M_9.

Table 1. Baseline patient demographic and clinical characteristics.

Characteristics	Patients with multiple sclerosis (N=102)
Age (years), mean (SD)	46.4 (10.4)
Sex, n (%)	
Female	75 (73.5)
Male	27 (26.5)
Education level[a], n (%)	
Low	3 (2.9)
Middle	34 (33.3)
High	65 (63.7)
Multiple sclerosis type, n (%)	
Primary progressive	11 (10.8)
Secondary progressive	30 (29.4)
Relapsing remitting	61 (59.8)
Disease duration since diagnosis (years), median (IQR)	5.7 (3.0-13.1)
Expanded Disability Status Scale score, median (range)	3.5 (1.5-7.0)

[a]Education levels were defined according to Rijnen et al [22].

XSL•FO
RenderX

Figure 2. Bar graph depicting the retention rate (left y-axis, "user percentage") of patients per month with superimposed box plots of the number of daily keystroke events (right y-axis, "event count"). The values above the bars show the retention rates as percentages.

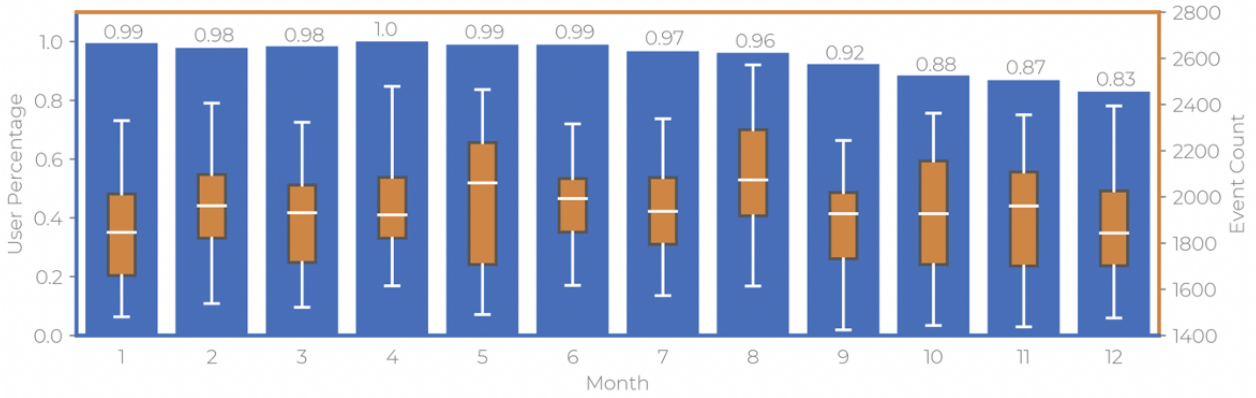


Table 2. Clinical outcomes and keystroke dynamics clusters for each clinical visit.

Outcomes	M_0	M_3	M_6	M_9	M_{12}
Nine-Hole Peg Test[a]					
Subjects, n	102	93	76	58	89
Time (seconds), median (IQR)	21.2 (19.4-25.0)	21.0 (18.7-24.0)	20.5 (18.8-22.5)	20.2 (18.5-22.0)	20.3 (18.7-23.0)
Symbol Digit Modalities Test					
Subjects, n	102	93	76	58	90
Mean score (SD)	54.4 (10.3)	56.8 (10.4)	57.9 (12.0)	61.3 (12.8)	60.3 (12.9)
Fine motor score cluster					
Subjects, n	96	88	72	55	71
Time (seconds), mean (SD)	0.45 (0.16)	0.44 (0.16)	0.44 (0.17)	0.39 (0.15)	0.42 (0.17)
Days[b] (n), mean (SD)	14.3 (2.4)	26.4 (4.8)	26.2 (4.8)	25.5 (5.4)	15.5 (5.5)
Cognition score cluster					
Subjects, n	101	89	70	55	72
Time (seconds), mean (SD)	1.01 (0.42)	0.95 (0.40)	0.92 (0.40)	0.86 (0.38)	0.90 (0.44)
Days[b] (n), mean (SD)	7.8 (1.0)	13.7 (1.0)	13.3 (2.2)	13.1 (2.1)	8.4 (2.7)

[a]For the Nine-Hole Peg Test, an average outlier threshold of 40 seconds was implemented, excluding 14 of 387 samples (3.6%).

[b]Only days with ≥50 keystroke events.

Upper Limb Function

For the association between the NHPT and FMSC, 98 patients with MS were included in the mixed model analysis with an average of 3.9 observations per patient. Overall, the mean (SD) for FMSC was 0.43 (0.16) seconds and the median (IQR) for the NHPT was 20.6 (18.8-23.3) seconds. The results of the mixed model analysis are shown in Table 3 and depicted visually in Figure 3. In the overall model, FMSC was significantly associated with the NHPT and explained 42% of the variance in the NHPT results. Age and sex were not found to be relevant confounders in this association. In the hybrid model, a one-SD (0.16-second) increase in FMSC was significantly associated with an increase in NHPT of 2.5 seconds between patients and 1.1 seconds within patients.

XSL•FO
RenderX

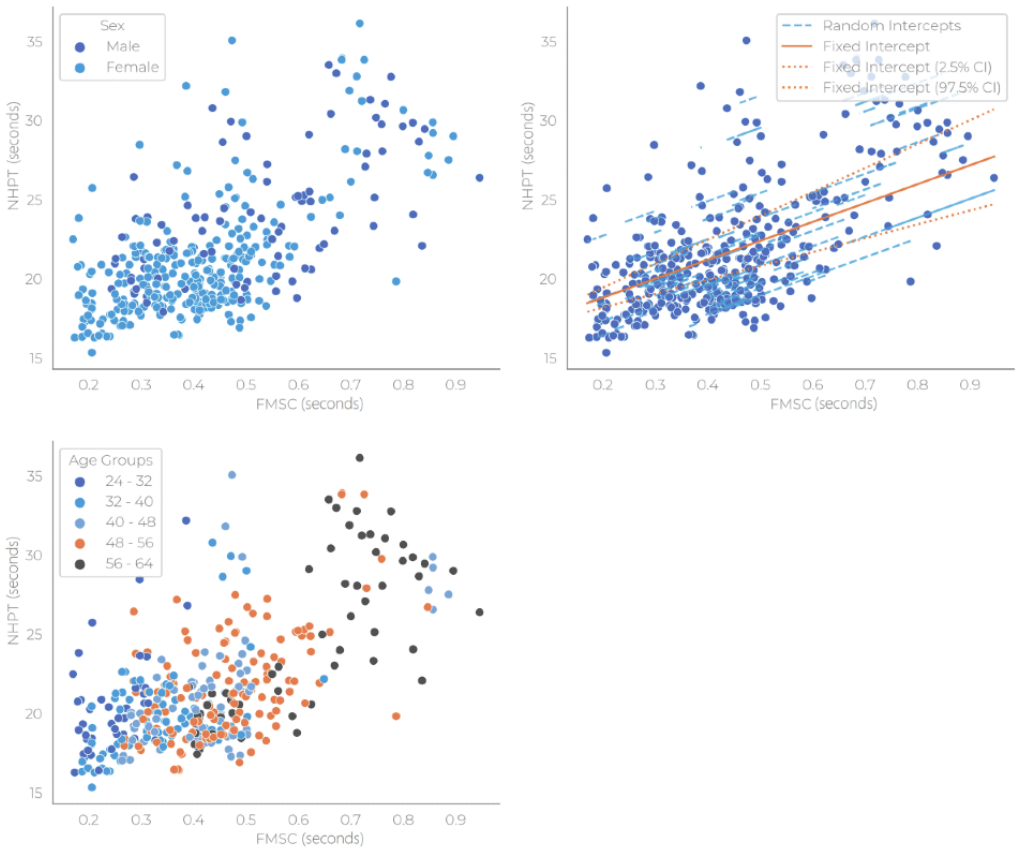
Table 3. Results of linear mixed model analyses of Nine-Hole Peg Test results over time with a random intercept on subject level.

	β (95% CI)	*P* value	Random effect variance, %	Explained variance, %
Intercept only			13.7	N/A[a]
Fine motor score cluster	12.62 (9.61-15.63)	<.001	8	42
Fine motor score cluster and covariates[b]	12.56 (8.96-16.16)	<.001	7.7	43.9
Hybrid model			7.7	43.7
Between subjects	15.91 (12.18-19.63)	<.001	N/A	N/A
Within subjects	6.94 (2.00-11.87)	<.001	N/A	N/A

[a]N/A: not applicable.

[b]Covariates included age and sex.

Figure 3. Scatter plots and linear mixed model fit for the Nine-Hole Peg Test and fine motor score cluster by covariates (sex and age), with random intercepts on subject level, and the number of days that constituted the keystroke cluster data points. FMSC: fine motor score cluster; NHPT: Nine-Hole Peg Test.



Information Processing Speed

All 102 patients with MS were included in the analysis of the association between information processing speed and the cognition keystroke cluster. The patients had an average of 3.8 repeated observations. The overall mean (SD) was 0.94 (0.41) seconds for CSC and 58.9 (12.1) points for SDMT. The output of the mixed model analyses is summarized in Table 4 and shown visually in Figure 4. In the overall model, CSC was significantly associated with SDMT and, together with age, sex, and level of education, explained 30.4% of the variance in SDMT. In the hybrid model, an increase of 1 SD (0.41 seconds) in CSC was significantly associated with a decrease of –4.6 in SDMT between patients. The within-subject association between CSC and SDMT, however, was not statistically significant.

XSL•FO
RenderX

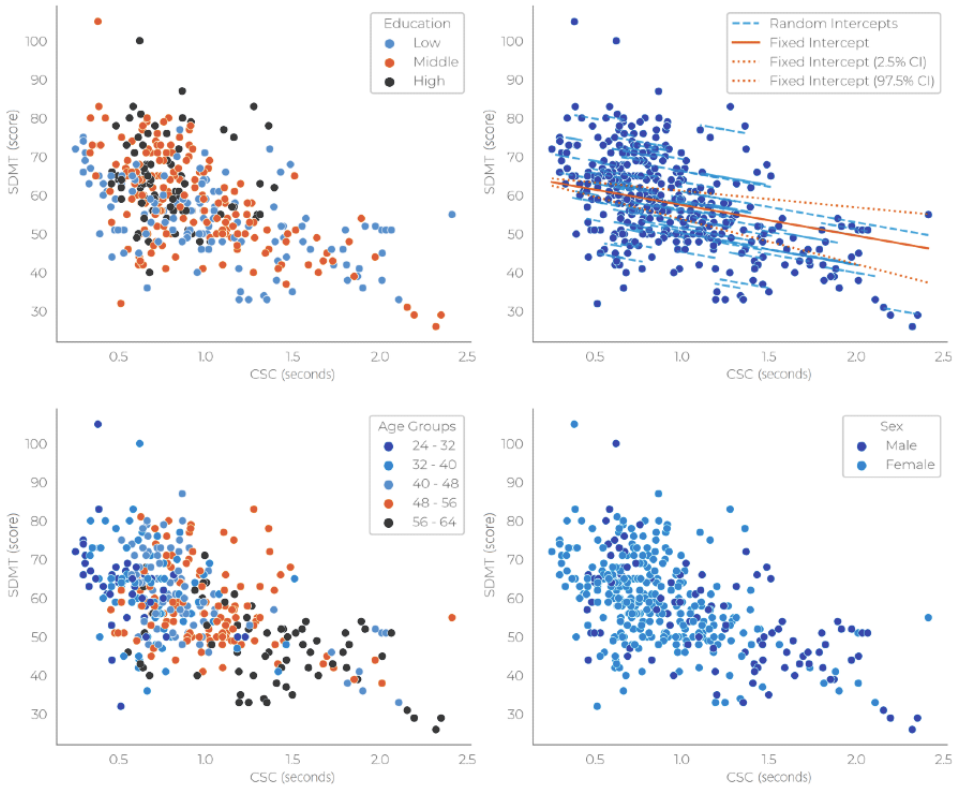
Table 4. Results of linear mixed model analyses of the Symbol Digit Modalities Test results over time with random intercepts on subject level and time (in days).

	β (95% CI)	*P* value	Random effect variance, %	Explained variance, %
Intercept only			110.9	N/A[a]
Cognition score cluster	–8.57 (–12.02 to –5.12)	<.001	82.7	25.4
Cognition score cluster and covariates[b]	–5.02 (–9.02 to –1.02)	.02	77.1	30.4
Hybrid model (including covariates[a])			74.4	32.9
Between subjects	–11.25 (–17.28 to –5.21)	<.001	N/A	N/A
Within subjects	–0.35 (–5.60 to 4.89)	.9	N/A	N/A

[a]N/A: not applicable.

[b]Covariates included age, sex, and level of education.

Figure 4. Scatter plots and linear mixed model fit for Symbol Digit Modalities Test and cognition score cluster by covariates (level of education, age, and sex) and random intercepts on subject level. SDMT: symbol digit modalities test; CSC: cognition score cluster.



Discussion

Principal Findings

This study investigated the longitudinal association between smartphone KD and commonly used clinical measures for upper limb function and information processing speed in patients with MS. In the overall model, the fine motor keystroke cluster was significantly associated with the NHPT (β=12.6, 95% CI 9.6-15.6); higher latency for presses and releases of alphanumeric keys during typing was related to a worse performance on the NHPT. When splitting the model for between-subject and within-subject effects, the association remained significant for both (β=15.9, 95% CI 12.2-19.6, and β=6.9, 95% CI 2.0-11.9, respectively). For the association between the cognitive keystroke cluster and SDMT, the time in days was included to account for practice effects on the SDMT and the imbalance in intervals between visits across subjects. CSC was found to be significantly negatively associated with SDMT; higher latency for backspace and punctuation mark keypresses was related to a worse SDMT score. This association had a β of –5.0 (95% CI –9.0 to –1.0) after adjusting for age, sex, and level of education. In the hybrid model for the cognitive keystroke cluster, the between-subject effect increased to β=–11.2 (95% CI –17.3 to –5.2), whereas the within-subject effect decreased to β=–0.4 (95% CI –5.6 to 4.9). To improve the interpretability of these associations, rather than considering 1-unit changes in keystroke clusters, the effect sizes can be recalculated to represent a change of 1 SD in keystroke clusters. In this distribution-based approach, a 1-SD change in FMSC corresponded with a change in NHPT of 1.1 seconds within patients or a 2.5-second difference between

XSL•FO
RenderX

patients. Likewise, a 1-SD change in CSC corresponded with a change in SDMT of –0.14 points (although this was not significant) within patients or a –4.6-point difference between patients. Therefore, in our current cohort, a 2-SD change in FMSC and a 1-SD change in CSC would correspond to clinically relevant changes, as a 20% change in NHPT and a 4-point change in SDMT are considered clinically relevant based on group studies [14,16].

Comparison With Prior Work

Measurements of task or activity performance are an integral part of assessing and monitoring chronic neurological disorders such as MS. Typing on a smartphone is a near-daily activity from which biometric information pertaining to physical or mental functions can be derived. Despite this, the use of KD in the assessment of diseases is relatively underutilized, especially considering that touchscreen typing has existed for over a decade. Hence, our objective was to validate the use of passive KD, measured with the Neurokeys app, to improve disease management in MS. To this end, earlier investigations by our research group reported on the clinimetric properties of reliability, validity, and 3-month responsiveness of KD in MS [10,11]. To the best of our knowledge, other applications of KD analysis in diseases are limited to the detection of early-stage Parkinson disease [23-28], upper limb dysfunction in amyotrophic lateral sclerosis [29], and severity in mood disorders [30,31]. The objective of the studies of Parkinson disease was to differentiate subjects with disease or early disease from those without, making the study endpoints not directly comparable to ours. Nevertheless, the study on amyotrophic lateral sclerosis found worse typing to be associated with progression of the disease, which is similar to our current findings. Last, the 2 studies on mood disorders found significant regression effects between severity of depressive symptoms and smartphone keyboard activity. This is in line with our findings, in which worse typing parameters corresponded to worse performance on the clinical tests. In addition, concurrent to our findings, these studies showed that KD can be utilized and can even outperform clinical standards in the detection and assessment of disease status through capitalizing on motor anomalies and, to some extent, cognitive dynamics that affect typing behavior.

Of note is that, besides our 3-month responsiveness interim analysis, there are currently no studies investigating KD in a longitudinal setting in patients with MS. While research investigating differences between subjects is of great importance, especially in early validation research, differences across subjects cannot directly be extrapolated to changes that occur within individuals over time. Therefore, analyzing change over time within patients is essential for monitoring or predictive modeling in MS. Splitting our model to separately determine between-subject and within-subject effects showed that the latter were stronger than the former. This suggests that in our sample, differences in upper limb function and information processing speed tended to be greater across patients than within patients, as shown by the SD of the outcomes across patients being much larger than the average change over time. This is not surprising, given that research in outcome measures in MS often struggles to achieve adequate sensitivity to change over time compared to correlations across patients [32]. For upper limb function, our model that separated the between-subject and within-subject effects still showed a strong, significant within-subject effect estimate, indicating that the fine motor keystroke cluster is sensitive to change within individuals.

For information processing speed, prior to adjustment for time, we also found a significant within-subject effect estimate for the cognitive keystroke cluster (data not shown). However, accounting for practice effects by adding the time point as a random effect to the model resulted in a lower, statistically nonsignificant within-subject effect. This suggests that the association between SDMT and the cognitive keystroke cluster in our current cohort was affected more strongly by the effect of learning than by changes in the keystroke cluster within patients. This explanation is supported by the results of modeling the effect of time as a fixed term instead of a random effect. In this model, the effect of time was stronger than the within-subject effect of the cognition keystroke cluster. In addition, when time as a fixed term was modeled categorically (such as M_0, M_3, or M_6), instead of being linear, the effect of time on the association between SDMT and cognition keystroke cluster was larger at later time points than earlier time points. As the amount of learning differs between patients, patients who are less severely affected by MS tend to have stronger practice effects than patients with more severe disability [33], and the larger positive slopes at later time points can be explained by practice effects causing a larger spread in SDMT data over time while the cognition keystroke cluster stays more or less stable.

Despite practice effects most likely diluting our findings on the within-subject association, the strong between-subject effect demonstrates the promise of the use of KD as a biomarker of information processing speed. Therefore, monitoring of cognition using KD needs further investigation with clinical cognitive outcomes that are more sensitive or less affected by practice effects and with a study population that allows a closer focus on cognitive function (ie, by including the presence of cognitive deficits as a selection criterion or using a longer follow-up duration) to demonstrate effects larger than measurement variability or learning effects. Similarly, a smartphone-based cognition test in the same cohort was found to be valid on the cross-sectional level, but lacked responsiveness when looking at change longitudinally, as changes within subjects are subtler than differences across subjects and measurements can be variable [34]. Additionally, more advanced analysis methods, such as nonlinear models, may increase sensitivity and allow higher frequency keystroke data and further investigation [12,35].

Limitations

A few limitations should be considered. First, despite modeling time in the analyses to take into account score changes over time, practice effects were not adjusted for, such by having healthy controls throughout the study. In addition, practice effects may have been exacerbated in the current cohort by their weekly performance of a smartphone variant of the SDMT concurrently with the digital biomarkers. Second, we investigated 2 commonly affected domains in MS that are

XSL•FO
RenderX

directly involved with typing on the smartphone: upper limb function and information processing speed. In reality, MS entails a much broader array of functional spheres and relevant treatment outcomes. We collected a broad scope of clinical outcomes in our current cohort, and these data should be examined in future work in order to incorporate KD as a complete tool for monitoring MS. Lastly, a significant number of patients had missing clinical data, most prominently at M_6 and M_9, due to the COVID-19 pandemic. This also created a bias, as patients with secondary or primary progressive MS missed their clinical visits more often than patients with relapsing-remitting MS.

Conclusions

Keystroke clusters constructed from passively acquired smartphone KD data were shown to reflect function in patients with MS in a longitudinal setting, as measured with commonly used clinical outcome measures of upper limb function and cognitive functioning. In the longitudinal repeated measures analysis, the fine motor keystroke cluster was found to be associated with upper limb function across and within patients. This attests to the use of KD for monitoring and predictive purposes in MS. Monitoring cognitive function with KD needs further investigation, as a significant association was found in the overall model, but this relied mostly on differences between patients rather than changes within patients, likely exacerbated by practice effects related to the clinical measures. Altogether, KD during typing provided detailed data on the temporal and granular level on everyday upper limb and cognitive function. Our current findings are the first to demonstrate associations between clinical outcomes in MS and smartphone typing performance. With the ongoing expansion of therapeutic interventions, KD as a remote passive biomarker may improve clinical assessment and patient-centered disease management in MS. Important steps for future research are investigating other highly relevant MS outcomes, such as disease activity, and the external validity of the current results by monitoring function in clinical practice on the individual-patient level.

Acknowledgments

The authors would like to thank all the patients for their participation in this study. The authors also disclose receipt of the following financial support for the research, authorship, and publication of this article: funding from the Public Private Partnership Allowance, made available by Health-Holland, Top Sector Life Sciences and Health (grant number LSHM16060-SGF), and Stichting Multiple Scleroris Research (grant number 16-946 MS) to stimulate public-private partnerships; unrestricted funding was also received from Biogen.

Data Availability

Anonymized data not published within the article are available upon request from a qualified investigator. Such requests must be submitted in writing and will be reviewed for researcher qualifications and the legitimacy of the research purpose.

Authors' Contributions

KHL contributed to designing and conceptualizing the study, had a major role in the acquisition, analysis, and interpretation of the data, and drafted and revised the manuscript for intellectual content. JT had a major role in the acquisition, analysis, and interpretation of the data and contributed to revising the manuscript for intellectual content. BLW contributed to analysis and interpretation of the data and revising the manuscript for intellectual content. GL contributed to analysis and interpretation of the data and revising the manuscript for intellectual content. KM contributed to revising the manuscript for intellectual content. BU contributed to designing and conceptualizing the study and revising the manuscript for intellectual content. VDG contributed to designing and conceptualizing the study, analyzing and interpreting the data, and revising the manuscript for intellectual content. JK contributed to designing and conceptualizing the study, analyzing and interpreting the data, and revising the manuscript for intellectual content.

Conflicts of Interest

KHL, BLW, and VDG declare no conflicts of interest. JT, GL, and KM are employees of Neurocast BV (an industry partner). BU received consultancy fees from Biogen Idec, Genzyme, Merck Serono, Novartis, Roche, and Teva. JK has accepted speaker and consultancy fees from Merck, Biogen, Teva, Genzyme, Roche, and Novartis.

Multimedia Appendix 1
Supplementary table 1. Principal component analysis of the timing-related and error-related/paralinguistic keystroke features.
[DOCX File , 13 KB - jmir_v24i11e37614_app1.docx]

Multimedia Appendix 2
Supplementary table 2. Correlation matrix of the timing-related keystroke features.
[DOCX File , 13 KB - jmir_v24i11e37614_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Supplementary table 3. Correlation matrix of the error-related/paralinguistic keystroke features.
[DOCX File , 13 KB - jmir_v24i11e37614_app3.docx]

Multimedia Appendix 4
Supplementary figure 1. Line graph of individual patients' NHPT scores during the study.
[PNG File , 176 KB - jmir_v24i11e37614_app4.png]

Multimedia Appendix 5
Supplementary figure 2. Line graph of individual patients' SDMT scores during the study.
[PNG File , 160 KB - jmir_v24i11e37614_app5.png]

References

1. Hauser SL, Cree BA. Treatment of multiple sclerosis: a review. Am J Med 2020 Dec;133(12):1380-1390.e2 [FREE Full text] [doi: 10.1016/j.amjmed.2020.05.049] [Medline: 32682869]
2. Faissner S, Plemel JR, Gold R, Yong VW. Progressive multiple sclerosis: from pathophysiology to therapeutic strategies. Nat Rev Drug Discov 2019 Dec 09;18(12):905-922. [doi: 10.1038/s41573-019-0035-2] [Medline: 31399729]
3. Lubetzki C, Zalc B, Williams A, Stadelmann C, Stankoff B. Remyelination in multiple sclerosis: from basic science to clinical translation. Lancet Neurol 2020 Aug;19(8):678-688. [doi: 10.1016/S1474-4422(20)30140-X] [Medline: 32702337]
4. Motl RW, Sandroff BM, Kwakkel G, Dalgas U, Feinstein A, Heesen C, et al. Exercise in patients with multiple sclerosis. Lancet Neurol 2017 Oct;16(10):848-856. [doi: 10.1016/S1474-4422(17)30281-8] [Medline: 28920890]
5. Goverover Y, Chiaravalloti ND, O'Brien AR, DeLuca J. Evidenced-based cognitive rehabilitation for persons with multiple sclerosis: an updated review of the literature from 2007 to 2016. Arch Phys Med Rehabil 2018 Feb;99(2):390-407. [doi: 10.1016/j.apmr.2017.07.021] [Medline: 28958607]
6. Lublin F, Reingold S, Cohen J, Cutter G, Sørensen PS, Thompson A, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology 2014 Jul 15;83(3):278-286 [FREE Full text] [doi: 10.1212/WNL.0000000000000560] [Medline: 24871874]
7. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983 Nov 01;33(11):1444-1452. [doi: 10.1212/wnl.33.11.1444] [Medline: 6685237]
8. Cutter GB, Baier ML, Rudick RA, Cookfair D, Fischer J, Petkau J, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain 1999 May;122 (Pt 5):871-882. [doi: 10.1093/brain/122.5.871] [Medline: 10355672]
9. Goldman MD, LaRocca NG, Rudick RA, Hudson LD, Chin PS, Francis GS, Multiple Sclerosis Outcome Assessments Consortium. Evaluation of multiple sclerosis disability outcome measures using pooled clinical trial data. Neurology 2019 Nov 19;93(21):e1921-e1931 [FREE Full text] [doi: 10.1212/WNL.0000000000008519] [Medline: 31641014]
10. Lam K, Meijer K, Loonstra F, Coerver E, Twose J, Redeman E, et al. Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis. Mult Scler 2021 Aug 05;27(9):1421-1431 [FREE Full text] [doi: 10.1177/1352458520968797] [Medline: 33150823]
11. Lam K, Twose J, McConchie H, Licitra G, Meijer K, de Ruiter L, et al. Smartphone-derived keystroke dynamics are sensitive to relevant changes in multiple sclerosis. Eur J Neurol 2022 Feb 14;29(2):522-534 [FREE Full text] [doi: 10.1111/ene.15162] [Medline: 34719076]
12. Twose J, Licitra G, McConchie H, Lam KH, Killestein J. Early-warning signals for disease activity in patients diagnosed with multiple sclerosis based on keystroke dynamics. Chaos 2020 Nov;30(11):113133. [doi: 10.1063/5.0022031] [Medline: 33261343]
13. Neurokeys. Neurocast BV. URL: https://neurokeys.app/ [accessed 2022-10-13]
14. Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, Multiple Sclerosis Outcome Assessments Consortium. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Mult Scler 2017 Apr 16;23(5):711-720 [FREE Full text] [doi: 10.1177/1352458517690824] [Medline: 28206826]
15. Denney DR, Lynch SG, Parmenter BA. A 3-year longitudinal study of cognitive impairment in patients with primary progressive multiple sclerosis: speed matters. J Neurol Sci 2008 Apr 15;267(1-2):129-136. [doi: 10.1016/j.jns.2007.10.007] [Medline: 17976653]
16. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R, Multiple Sclerosis Outcome Assessments Consortium. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult Scler 2017 Apr 16;23(5):721-733 [FREE Full text] [doi: 10.1177/1352458517690821] [Medline: 28206827]
17. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
18. Twisk JWR. Applied Mixed Model Analysis: A Practical Guide, 2nd Ed. Cambridge, UK: Cambridge University Press; 2019.

XSL•FO
RenderX

19. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. Annu Rev Psychol 2011 Jan 10;62(1):583-619 [FREE Full text] [doi: 10.1146/annurev.psych.093008.100356] [Medline: 19575624]
20. Twisk JW, de Vente W. Hybrid models were found to be very elegant to disentangle longitudinal within- and between-subject relationships. J Clin Epidemiol 2019 Mar;107:66-70. [doi: 10.1016/j.jclinepi.2018.11.021] [Medline: 30500406]
21. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. Am J Epidemiol 1993 Dec 01;138(11):923-936. [doi: 10.1093/oxfordjournals.aje.a116813] [Medline: 8256780]
22. Rijnen SJM, Meskal I, Emons WHM, Campman CAM, van der Linden SD, Gehring K, et al. Evaluation of normative data of a widely used computerized neuropsychological battery: applicability and effects of sociodemographic variables in a Dutch sample. Assessment 2020 Mar 12;27(2):373-383 [FREE Full text] [doi: 10.1177/1073191117727346] [Medline: 28895436]
23. Arroyo-Gallego T, Ledesma-Carbayo MJ, Sanchez-Ferro A, Butterworth I, Mendoza CS, Matarazzo M, et al. Detection of motor impairment in Parkinson's Disease via mobile touchscreen typing. IEEE Trans Biomed Eng 2017 Sep;64(9):1994-2002. [doi: 10.1109/TBME.2017.2664802] [Medline: 28237917]
24. Adams WR. High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. PLoS One 2017 Nov 30;12(11):e0188226 [FREE Full text] [doi: 10.1371/journal.pone.0188226] [Medline: 29190695]
25. Iakovakis D, Hadjidimitriou S, Charisis V, Bostantzopoulou S, Katsarou Z, Hadjileontiadis LJ. Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage Parkinson's disease. Sci Rep 2018 May 16;8(1):7663-7613 [FREE Full text] [doi: 10.1038/s41598-018-25999-0] [Medline: 29769594]
26. Iakovakis D, Chaudhuri KR, Klingelhoefer L, Bostantjopoulou S, Katsarou Z, Trivedi D, et al. Screening of Parkinsonian subtle fine-motor impairment from touchscreen typing via deep learning. Sci Rep 2020 Jul 28;10(1):12623-12613 [FREE Full text] [doi: 10.1038/s41598-020-69369-1] [Medline: 32724210]
27. Papadopoulos A, Iakovakis D, Klingelhoefer L, Bostantjopoulou S, Chaudhuri KR, Kyritsis K, et al. Unobtrusive detection of Parkinson's disease from multi-modal and in-the-wild sensor data using deep learning techniques. Sci Rep 2020 Dec 07;10(1):21370-21310 [FREE Full text] [doi: 10.1038/s41598-020-78418-8] [Medline: 33288807]
28. Pham TD. Pattern analysis of computer keystroke time series in healthy control and early-stage Parkinson's disease subjects using fuzzy recurrence and scalable recurrence network features. J Neurosci Methods 2018 Sep 01;307:194-202. [doi: 10.1016/j.jneumeth.2018.05.019] [Medline: 29859213]
29. Londral A, Pinto S, de Carvalho M. Markers for upper limb dysfunction in Amyotrophic Lateral Sclerosis using analysis of typing activity. Clin Neurophysiol 2016 Jan;127(1):925-931. [doi: 10.1016/j.clinph.2015.06.017] [Medline: 26160275]
30. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a BiAffect digital phenotyping study. J Med Internet Res 2018 Jul 20;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]
31. Vesel C, Rashidisabet H, Zulueta J, Stange J, Duffecy J, Hussain F, et al. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: A BiAffect iOS study. J Am Med Inform Assoc 2020 Jul 01;27(7):1007-1018 [FREE Full text] [doi: 10.1093/jamia/ocaa057] [Medline: 32467973]
32. de Groot V, Beckerman H, Uitdehaag BMJ, de Vet HCW, Lankhorst GJ, Polman CH, et al. The usefulness of evaluative outcome measures in patients with multiple sclerosis. Brain 2006 Oct 15;129(Pt 10):2648-2659. [doi: 10.1093/brain/awl223] [Medline: 16956906]
33. Roar M, Illes Z, Sejbaek T. Practice effect in Symbol Digit Modalities Test in multiple sclerosis patients treated with natalizumab. Mult Scler Relat Disord 2016 Nov;10:116-122. [doi: 10.1016/j.msard.2016.09.009] [Medline: 27919477]
34. Lam K, van Oirschot P, den Teuling B, Hulst H, de Jong B, Uitdehaag B, et al. Reliability, construct and concurrent validity of a smartphone-based cognition test in multiple sclerosis. Mult Scler 2022 Feb 26;28(2):300-308 [FREE Full text] [doi: 10.1177/13524585211018103] [Medline: 34037472]
35. Lam K, Bucur IG, Van Oirschot P, De Graaf F, Weda H, Strijbis E, et al. Towards individualized monitoring of cognition in multiple sclerosis in the digital era: A one-year cohort study. Mult Scler Relat Disord 2022 Apr;60:103692 [FREE Full text] [doi: 10.1016/j.msard.2022.103692] [Medline: 35219240]

Abbreviations

CSC: cognition score cluster
EDSS: Expanded Disability Status Scale
FMSC: fine motor score cluster
KD: keystroke dynamics
MRI: magnetic resonance imaging
MS: multiple sclerosis
MSFC: Multiple Sclerosis Functional Composite
NHPT: Nine-Hole Peg Test
SDMT: Symbol Digit Modalities Test

XSL•FO
RenderX

Edited by A Mavragani; submitted 28.02.22; peer-reviewed by Z Xie, D Murray; comments to author 19.05.22; revised version received 31.07.22; accepted 22.08.22; published 07.11.22.

Please cite as:
Lam KH, Twose J, Lissenberg-Witte B, Licitra G, Meijer K, Uitdehaag B, De Groot V, Killestein J
The Use of Smartphone Keystroke Dynamics to Passively Monitor Upper Limb and Cognitive Function in Multiple Sclerosis: Longitudinal Analysis
J Med Internet Res 2022;24(11):e37614
URL: https://www.jmir.org/2022/11/e37614
doi:10.2196/37614
PMID:36342763

©Ka-Hoo Lam, James Twose, Birgit Lissenberg-Witte, Giovanni Licitra, Kim Meijer, Bernard Uitdehaag, Vincent De Groot, Joep Killestein. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Papetar

Personalized Prediction of Response to Smartphone-Delivered Meditation Training: Randomized Controlled Trial

Christian A Webb[1,2], PhD; Matthew J Hirshberg[3], PhD; Richard J Davidson[3,4,5], PhD; Simon B Goldberg[3,6], PhD

[1]Harvard Medical School, Boston, MA, United States

[2]McLean Hospital, Belmont, MA, United States

[3]Center for Healthy Minds, University of Wisconsin – Madison, Madison, WI, United States

[4]Department of Psychology, University of Wisconsin – Madison, Madison, WI, United States

[5]Department of Psychiatry, University of Wisconsin – Madison, Madison, WI, United States

[6]Department of Counseling Psychology, University of Wisconsin – Madison, Madison, WI, United States

Corresponding Author:
Simon B Goldberg, PhD
Department of Counseling Psychology
University of Wisconsin – Madison
315 Education Building
1000 Bascom Mall
Madison, WI, 53706
United States
Phone: 1 608 265 8986
Email: sbgoldberg@wisc.edu

Abstract

Background: Meditation apps have surged in popularity in recent years, with an increasing number of individuals turning to these apps to cope with stress, including during the COVID-19 pandemic. Meditation apps are the most commonly used mental health apps for depression and anxiety. However, little is known about who is well suited to these apps.

Objective: This study aimed to develop and test a data-driven algorithm to predict which individuals are most likely to benefit from app-based meditation training.

Methods: Using randomized controlled trial data comparing a 4-week meditation app (Healthy Minds Program [HMP]) with an assessment-only control condition in school system employees (n=662), we developed an algorithm to predict who is most likely to benefit from HMP. Baseline clinical and demographic characteristics were submitted to a machine learning model to develop a "Personalized Advantage Index" (PAI) reflecting an individual's expected reduction in distress (primary outcome) from HMP versus control.

Results: A significant group × PAI interaction emerged (t_{658}=3.30; P=.001), indicating that PAI scores moderated group differences in outcomes. A regression model that included repetitive negative thinking as the sole baseline predictor performed comparably well. Finally, we demonstrate the translation of a predictive model into personalized recommendations of expected benefit.

Conclusions: Overall, the results revealed the potential of a data-driven algorithm to inform which individuals are most likely to benefit from a meditation app. Such an algorithm could be used to objectively communicate expected benefits to individuals, allowing them to make more informed decisions about whether a meditation app is appropriate for them.

Trial Registration: ClinicalTrials.gov NCT04426318; https://clinicaltrials.gov/ct2/show/NCT04426318

(J Med Internet Res 2022;24(11):e41566) doi:10.2196/41566

KEYWORDS

precision medicine; prediction; machine learning; meditation; mobile technology; smartphone app; mobile phone

XSL•FO
RenderX

Introduction

Background

Precision medicine, which involves the use of individual variability to guide prevention and treatment, has gained momentum in the health sciences over the past several years [1]. This approach aims to improve outcomes by matching patients with interventions most likely to yield success. In some medical specialties, precision medicine has led to impressive advances in personalized care. For example, research in oncology (eg, lung and breast cancer) has effectively matched patients to targeted cancer treatments based on the unique genetic characteristics of their tumors, which has been shown to improve outcomes [2-4].

Psychiatry and clinical psychology have long hoped to better match patients with interventions. Numerous studies have examined patient-level factors (eg, demographic, clinical, and neurobiological characteristics) as predictors of treatment response [5,6]. However, with many potential predictors and inconsistencies across studies in the presence, direction, and strength of associations with outcomes, empirically supported guidelines for optimal treatment matching remain elusive.

Machine learning has emerged as a promising analytical approach well suited for handling and integrating large numbers of predictor variables, including correlated predictors, that may individually only modestly predict outcomes of interest but can collectively predict significant variance in patient outcomes [7,8]. Specific machine learning approaches such as decision tree–based algorithms (eg, random forest) also effectively model nonlinear and higher-order interactions that may underlie predictive relationships [9]. In contrast to traditional statistical approaches that emphasize evaluating a specific hypothesis (ie, null hypothesis significance testing), machine learning models typically emphasize optimizing predictive performance, and evaluating the generalizability of models to new individuals (eg, via cross-validation [CV], hold-out samples, or external validation) [10]. Machine learning approaches are increasingly being applied with some success in psychiatry and clinical psychology, with a growing number of studies demonstrating their ability to predict response to various psychiatric treatments [10-12].

In pursuit of precision mental health, researchers have leveraged machine learning approaches to optimize treatment recommendations [13-15]. For example, DeRubeis et al [16] developed the *Personalized Advantage Index* (PAI) as an algorithm for guiding treatment recommendations based on pretreatment patient characteristics. These models attempt to predict the benefit that a specific patient would derive from treatment A versus treatment B. The PAI has been successfully used to predict response to cognitive behavioral therapy (CBT) versus an antidepressant medication [16], CBT versus interpersonal therapy [17], CBT versus psychodynamic therapy [18], and an antidepressant medication versus placebo [19].

Prior research using the PAI and related approaches [12] provides promising initial evidence that data-driven algorithms may improve patient outcomes by matching individuals to the most therapeutically beneficial treatment, as opposed to the current suboptimal trial and error approach to treatment selection, which results in protracted psychiatric illness until an effective treatment is found. However, the fact remains that a substantial proportion of individuals with psychiatric disorders go untreated [20,21]. Digital health technologies, such as internet-based CBT [22] and smartphone-delivered mental health apps [23], have the potential to substantially increase access to evidence-based treatments [24]. However, the availability of thousands of mental health apps leaves potential consumers faced with a dizzying number of choices, with essentially no way of knowing which specific app may best suit their needs [25]. Data-driven treatment recommendation algorithms, such as the PAI, offer promising tools for informing optimal patient-treatment fit. Such approaches may also be valuable for addressing the persistent limitations of mobile health (mHealth) approaches, including notoriously high and rapid disengagement [26,27]. Moreover, the scalability of mHealth makes it possible to collect adequately powered sample sizes for robust modeling [28].

A recent analysis of available mental health apps revealed that meditation and mindfulness training (along with journaling and mood tracking) are the most common features offered across apps [29]. The two most widely used meditation apps (Headspace and Calm, with 5 million and 9 million monthly active users, respectively) account for 96% of daily active users in a recent evaluation of the top 27 apps for depression and anxiety [30]. Despite the soaring popularity of meditation apps, a critical question remains unanswered: *For whom* is app-based meditation training well suited?

This Study

This study involved a secondary analysis of a large-scale randomized controlled trial (RCT) comparing a meditation-based smartphone app, the Healthy Minds Program (HMP), with an assessment-only control condition [31]. The RCT was conducted on a sample of school district employees (n=662) in the state of Wisconsin during the COVID-19 pandemic. Relative to prepandemic levels, the rates of emotional distress and depressive symptoms increased substantially during the COVID-19 pandemic [32]. Available evidence suggests that the emotional well-being of teachers also decreased during the pandemic [33,34], as they coped with COVID-19–related stressors, uncertainty, and risks with the return to in-person instruction. Using data from the above RCT, the overarching goal of this study was to develop and evaluate a data-driven (PAI) approach to inform personalized meditation app recommendations for school employees. Using readily gathered self-reported baseline demographic and clinical characteristics, we developed and tested a machine learning algorithm to identify which individuals are most likely to benefit from the HMP app.

Methods

Participants and Procedure

Wisconsin school district employees were recruited via email and other electronic media between mid-June 2020 and late August 2020 (for a full description of study procedures, refer

XSL•FO
RenderX

to the study by Hirshberg et al [31]). Eligible participants were adults (aged ≥18 years) currently employed by a Wisconsin school who owned a smartphone capable of downloading the HMP, were fluent in English, had limited exposure to meditation or the HMP app, and had depressive symptoms below the severe range (*t* score<70 on Patient-Reported Outcomes Information System [PROMIS] Depression [35]). The *t* scores are population normed at 50, with an SD of 10. On completing the pretest measures, 666 participants were randomly assigned to use the 4-week HMP or an assessment-only control condition (4 participants were removed for failing multiple attention checks; refer to Figure S2 in Multimedia Appendix 1 [31,35-49] for the CONSORT [Consolidated Standards of Reporting Trials] flow diagram). Participants completed weekly questionnaires during the intervention period (ie, weeks 1, 2, and 3) along with a posttreatment assessment (week 4) and follow-up assessment (3 months after the end of the intervention period). These measures were administered via the web-based REDCap (Research Electronic Data Capture) survey system.

The trial was preregistered at ClinicalTrials.gov (NCT04426318) and through the Open Science Framework [50]. However, the current prediction analyses were not planned a priori and were not included in the preregistered data analysis plan. All code (implemented in the R statistical software [51]) used to reproduce the analyses in the manuscript have been posted on Open Science Framework [52].

The HMP includes contemplative practices designed to build skills supportive of 4 pillars of well-being: awareness, connection, insight, and purpose [36,37]. Participants were encouraged to engage with content from each of the 4 modules for approximately 1 week (ie, 4 weeks total). The content included didactic instruction and guided meditation practices. For the guided practices, participants could select the length of practice from 5 to 30 minutes. The HMP app was used for a mean of 10.9 (SD 9) days during the 4-week trial. For additional trial and sample details, refer to the study by Hirshberg et al [31].

Ethics Approval

The study procedures were approved by the University of Wisconsin—Madison Institutional Review Board (number 2020-0533).

Measures

Demographic Characteristics

The participants reported their age, gender identity, race and ethnicity, marital status, and income at baseline.

Primary Outcome

The prespecified primary outcome in the parent RCT was psychological distress, which was a composite of the computer-adaptive versions of the PROMIS Anxiety and PROMIS Depression measures [35] and the 10-item Perceived Stress Scale [38]. All 3 are widely used measures with established reliability and validity [39,40]. Refer to Multimedia Appendix 1 for details. Consistent with the prespecified data analytic plan, multilevel models estimated changes in distress over the 4-week intervention period. Random slopes (representing individual changes in distress over the intervention period) were calculated for each participant and served as the primary outcome in our machine learning prediction models.

Predictors

Several additional self-report questionnaires assessed secondary outcomes and candidate mediators that were theoretically linked to the pillars of well-being trained within the HMP. The 15-item Perseverative Thinking Questionnaire (PTQ) [53] assessed worry and rumination (Cronbach α=.95). The 5-item World Health Organization [54] assessed global well-being (α=.85). The 8-item Act with Awareness subscale of the Five Facet Mindfulness Questionnaire [55] assessed mindful attention in daily life (α=.91). The 5-item National Institutes of Health Toolbox Loneliness Questionnaire [56] assessed perceived social disconnection (α=.90). The 12-item Self-Compassion Scale Short Form [57] assessed feelings of kindness toward oneself (α=0.86). The 10-item Drexel Defusion Scale [58] assessed the ability to experientially distance from internal experiences (α=.84). The 10-item Meaning in Life Questionnaire (MLQ [59]) assessed the presence and search for meaning (Cronbach α=.91 and Cronbach α=.93, respectively).

Analytic Strategy

Predictor variables included preintervention distress (composite measure), anxiety (PROMIS), depression (PROMIS), stress (Perceived Stress Scale), repetitive negative thinking (PTQ), the mindfulness facet of acting with awareness (Five Facet Mindfulness Questionnaire), loneliness (National Institutes of Health Toolbox Loneliness), defusion (Drexel Defusion Scale), presence (MLQ), search for meaning (MLQ), self-compassion (Self-Compassion Scale Short Form), well-being (5-item World Health Organization), age, gender, race, marital status, and income.

Missing Value Imputation

Missing data were imputed using a random forest–based imputation (MissForest package in R [60]). To avoid contamination between the predictor and outcome scores, which may optimistically bias predictive performance, the outcome variable (slope of change in distress) was excluded from the imputation procedure. The rate of missing data was very low, with no variable missing more than 6 values. Refer to Multimedia Appendix 1 for additional details.

Generating Predicted Outcomes

To predict outcomes, 2 prognostic models (using elastic net regularized regression [ENR]; glmnet package in R) were developed: one for participants who received HMP and one for those who received the assessment-only control condition. To minimize overfitting, which can occur with traditional *k*-fold CV, a nested CV procedure was used for each of these prognostic models (ie, incorporating an outer and inner CV loop [41-44]). Refer to Multimedia Appendix 1 for details of the nested CV procedure.

The steps mentioned earlier generated predicted HMP outcomes for HMP participants and predicted control condition outcomes for the control participants. To generate predicted outcomes for the counterfactual condition (ie, the treatment condition one did

XSL•FO
RenderX

not receive), an ENR model was developed for one group (ie, full HMP or control sample) and used to predict outcomes for participants in the other group.

Evaluation of Recommendations

As a final product of the prediction models mentioned earlier, every participant had 2 predicted outcome scores: one for HMP and one for the control condition. Consistent with previous similar studies [18,19,61], we computed a PAI score by subtracting these 2 predicted outcomes (ie, the predicted slope of change in distress for HMP *minus* control) for each individual. Thus, a negative PAI score indicates that a given participant is predicted to experience greater reductions (ie, a more negative slope) in distress in HMP relative to the assessment-only control condition (and vice versa for positive PAI scores). The PAI can be interpreted as a continuous indicator reflecting the expected magnitude of the advantage of one treatment condition relative to the other (eg, a large negative PAI value indicates that the model predicts a relatively large between-group difference in outcome favoring HMP). We tested whether PAI scores moderated treatment group differences in outcome (ie, slope of change in distress) via a group (ie, intervention condition) × PAI interaction. The latter test allowed us to answer the following question: Are more negative PAI scores (reflecting relatively greater predicted benefit from HMP relative to the control condition) in fact associated with larger *observed* differences in outcomes favoring HMP?

Comparison Model

We compared the abovementioned multivariable machine learning (ENR) model with a simple linear regression with baseline repetitive negative thinking (PTQ) scores as the sole predictor (ie, repeating the above steps to generate a PAI score for every participant) implemented via 10-fold CV (repeated 100 times to generate stable estimates). Repetitive negative thinking was selected as a predictor in this comparison model based on prior research, indicating that it predicts response to mindfulness apps [43,45]. Refer to Multimedia Appendix 1 for additional analyses with baseline distress as the sole predictor. Finally, we used the parameter estimates from the final models to demonstrate the translation of the predicted outcomes to *personalized* recommendations for app-based mindfulness training.

All analyses were conducted using R software (version 4.0.2) [62]. The sample size was originally determined for the purpose of the parent trial to detect between-group differences in the primary outcome (change in distress [50]). To estimate whether the current sample size was adequately powered for the analyses proposed in this study, a Monte Carlo simulation approach (InteractionPoweR package in R) was used. Informed by the effect sizes from a prior mindfulness app RCT [45] that tested similar group × PAI interactions, simulations revealed that a sample size of at least 153 was needed for group × PAI interaction tests (with Cronbach α=.05; power=80%; refer to Figure S1, including figure note, in Multimedia Appendix 1 for additional power analysis details).

Results

Sample Demographics

The majority (523/662, 79%) of the participants reported depression or anxiety symptoms at baseline that were above the clinical cutoff on the PROMIS Depression and PROMIS Anxiety measures (*t* score>55), and more than half of the sample (343/662, 51.8%) reported moderate or greater anxiety or depressive symptoms at baseline (*t* score>60).

The groups did not differ at baseline in terms of the demographic or clinical variables (Table 1). Of those assigned to HMP, 95.6% (329/344) downloaded the app and 78.8% (271/344) used the app for ≥1 day. The mean number of days of use was 10.88 (SD 9.08). The mean number of minutes of practice was 127.93 (SD 130.63).

XSL•FO
RenderX

Table 1. Descriptive statistics for Healthy Minds Program and assessment-only control at baseline.

Variable	Healthy Minds Program			Control			P[a] value
	Value, N	Value, n (%)	Value, mean (SD)	Value, N	Value, n (%)	Value, mean (SD)	
Age (years)	344	—	42.47 (11.06)	318	—	42.70 (10.23)	.78
Gender (female)	344	299 (86.9)	—	318	279 (87.7)	—	.75
Non-Hispanic White	344	304 (88.4)	—	318	268 (84.3)	—	.13
Married	344	243 (70.6)	—	318	216 (67.9)	—	.45
College education	343	308 (89.8)	—	316	281 (88.9)	—	.72
Income (US $)							
≤50,000	344	56 (16.3)	—	318	55 (17.3)	—	.73
50,000-100,000	344	141 (41.0)	—	318	129 (40.6)	—	.91
100,000-150,000	344	104 (30.2)	—	318	96 (30.2)	—	.99
≥150,000	344	40 (11.6)	—	318	37 (11.6)	—	.99
PROMIS[b]							
Depression	342	—	55.37 (6.20)	315	—	55.47 (6.43)	.85
Anxiety	342	—	59.83 (6.95)	315	—	60.00 (7.11)	.75
Perceived Stress Scale	342	—	2.89 (0.56)	315	—	2.87 (0.60)	.69
Distress[c] (composite)	342	—	0.00 (0.88)	315	—	0.00 (0.91)	.97
Perseverative Thinking Questionnaire	342	—	29.89 (10.43)	315	—	29.62 (11.29)	.76
Five Facet Mindfulness Questionnaire—Acting with Awareness Subscale	342	—	24.80 (5.93)	315	—	24.56 (6.12)	.62
National Institutes of Health Toolbox Loneliness	342	—	2.53 (0.77)	315	—	2.58 (0.77)	.45
Drexel Defusion Scale	342	—	24.83 (7.89)	315	—	24.50 (8.16)	.60
MLQ[d]							
Presence	342	—	26.20 (5.44)	315	—	25.81 (5.46)	.36
Search for meaning	342	—	21.63 (6.61)	315	—	22.09 (6.79)	.38
World Health Organization well-being	341	—	12.76 (4.71)	315	—	12.47 (4.33)	.42
Self-Compassion Scale	342	—	2.98 (0.69)	315	—	2.93 (0.70)	.37

[a]*P* value from independent samples *t* test comparing groups at baseline.

[b]PROMIS: Patient-Reported Outcomes Information System.

[c]Distress: composite of PROMIS Depression, PROMIS Anxiety, and Perceived Stress Scale.

[d]MLQ: Meaning in Life Questionnaire.

Outcome Prediction

Higher baseline levels of distress, depression, and stress predicted better outcomes (ie, greater reductions in distress) in HMP (Table 2). The zero-order correlations between outcome and these 3 predictors were $r=-0.30$ (for distress), $r=-0.30$ (depression), and $r=-0.26$ (stress). Predicted HMP outcomes were significantly correlated with observed outcomes for the HMP group ($r=0.27$; $P<.001$; root mean square error [RMSE]=0.10) but not with the control condition outcomes ($r=0.07$; $P=.21$; RMSE=0.12). Conversely, predicted control condition outcomes were significantly correlated with observed outcomes for the control group ($r=0.19$; $P<.001$; RMSE=0.10) but not with HMP outcomes ($r=0.10$; $P=.06$; RMSE=0.12). Higher baseline scores for the following variables predicted better outcomes in the control condition: distress, anxiety, depression, stress, loneliness, defusion, and presence. In addition, lower levels of repetitive negative thinking, higher self-compassion, and being married were associated with better control condition outcomes (Table 2).

XSL•FO
RenderX

Table 2. Baseline variables retained in elastic net models predicting outcomes for each condition[a].

Predictors	Healthy Minds Program model, coefficient	Control model, coefficient
Age (years)	—[b]	—
Gender	—	—
Race	—	—
Marital status	—	−0.006
Income	—	—
PROMIS[c]		
Depression	−0.012	−0.005
Anxiety	—	−0.007
Perceived Stress Scale	−0.003	−0.006
Distress[d] (composite)	−0.011	−0.008
Perseverative Thinking Questionnaire	—	0.012
Five Facet Mindfulness Questionnaire—Acting with Awareness Subscale	—	—
National Institutes of Health Toolbox Loneliness	—	−0.002
Drexel Defusion Scale	—	−0.011
MLQ[e]		
Presence	—	−0.008
Search for meaning	—	—
World Health Organization well-being	—	—
Self-Compassion Scale	—	−0.002

[a]The larger set of baseline predictors retained in the elastic net regularized regression model applied to the control participants relative to the Healthy Minds Program (HMP) group was because the best-fitting model in the former group had a lower α value (ie, closer to ridge than lasso regression) relative to the HMP group. Negative parameter estimates indicate that higher scores on the predictor variable are associated with better outcomes (ie, reductions in distress).

[b]Variables that were not retained in the elastic net model.

[c]PROMIS: Patient-Reported Outcomes Information System.

[d]Distress: composite of PROMIS Depression, PROMIS Anxiety, and Perceived Stress Scale.

[e]MLQ: Meaning in Life Questionnaire.

Meditation App Recommendations

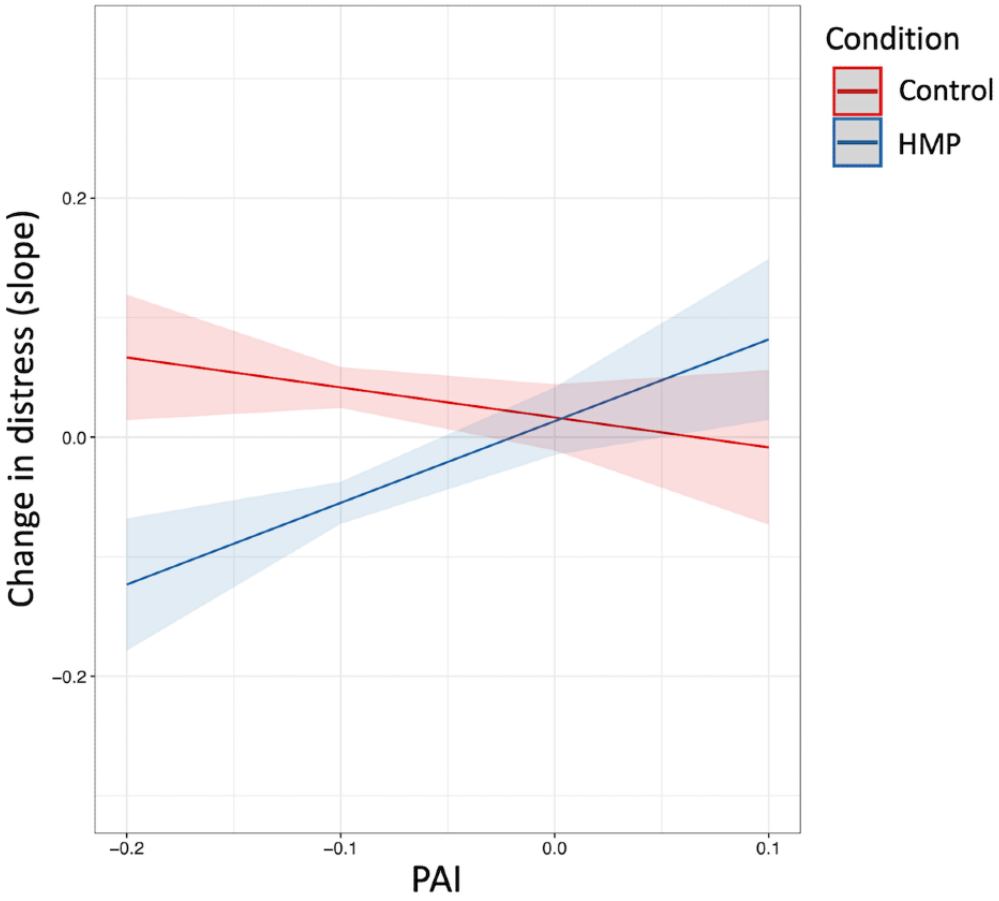
The mean PAI score was −0.07 (SD 0.03; range −0.17 to 0.03), indicating that the model predicted greater average symptom improvement for the HMP meditation app than for the assessment-only control condition. The model recommended HMP (PAI<0) for all participants except 5 (657/662, 99.2%).

Evaluation of Recommendations

A significant group × PAI interaction emerged in predicting outcome (t_{658}=3.30; P=.001; adjusted r^2=0.10), indicating that PAI scores moderated group differences in outcome. As displayed in Figure 1, as PAI scores decrease (ie, reflecting relatively stronger HMP recommendations), group differences in observed outcome increase, favoring HMP.

XSL•FO
RenderX

Figure 1. Group × Personalized Advantage Index (PAI) interaction. As PAI scores decrease (ie, reflecting relatively stronger recommendations for the Healthy Minds Program [HMP] app) group differences in observed outcome increase, favoring HMP.



Comparison Model

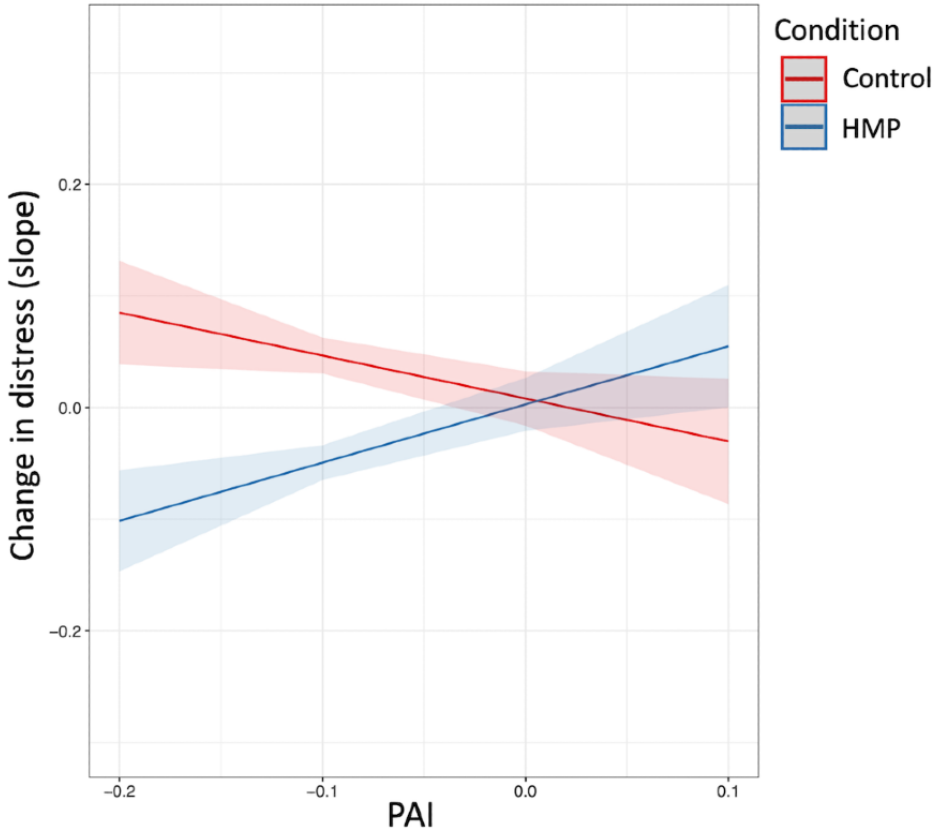
In the linear regression comparison model applied to the HMP group, higher levels of repetitive negative thinking were significantly associated with a *greater* reduction in distress from the mindfulness app (B=−0.02; t_{342}=−3.37; *P*<.001). The correlation between predicted HMP outcomes and observed outcomes was *r*=0.16 (*P*=.003; RMSE=0.10) for participants who received HMP and *r*=−0.14 (*P*=.02; RMSE=0.12) for the control group. In contrast to the pattern of findings for the HMP group, the linear regression model applied to the control sample revealed that higher levels of repetitive negative thinking were significantly associated with *poorer* outcomes than in the control condition (B=0.01; t_{316}=2.44; *P*=.02).

The correlation between predicted control condition outcomes and observed outcomes was *r*=0.11 (*P*=.049; RMSE=0.11) for the control group and *r*=−0.18 (*P*<.001; RMSE=0.12) for the HMP group.

A significant group × PAI interaction emerged in predicting changes in distress (t_{658}=3.81; *P*<.001; adjusted r^2=0.11), indicating that PAI scores moderated group differences in outcomes (Figure 2). Specifically, as PAI scores decreased (reflecting increasing repetitive negative thinking scores), group differences favoring the HMP condition also increased. Given the association between repetitive negative thinking and depressive symptoms [46,47], we also conducted additional sensitivity analyses controlling for baseline levels of depressive symptoms (as well as considering the number of days the app was used), which yielded the same pattern of findings (Multimedia Appendix 1). In summary, these results indicate that a simple linear regression including repetitive negative thinking as the sole predictor yields equivalent performance relative to a more complex multivariable ENR model (ie, adjusted r^2=0.11 vs r^2=0.10, respectively, for the group × PAI interaction).

XSL•FO
RenderX

Figure 2. Group × Personalized Advantage Index (PAI) interaction for the comparison model (ie, linear regression with baseline repetitive negative thinking [PTQ] scores as the sole predictor). As PAI scores decrease (ie, reflecting relatively stronger recommendations for the Healthy Minds Program [HMP] app) group differences in observed outcome increase, favoring HMP.



Translating a Predictive Model to Personalized Meditation App Recommendations

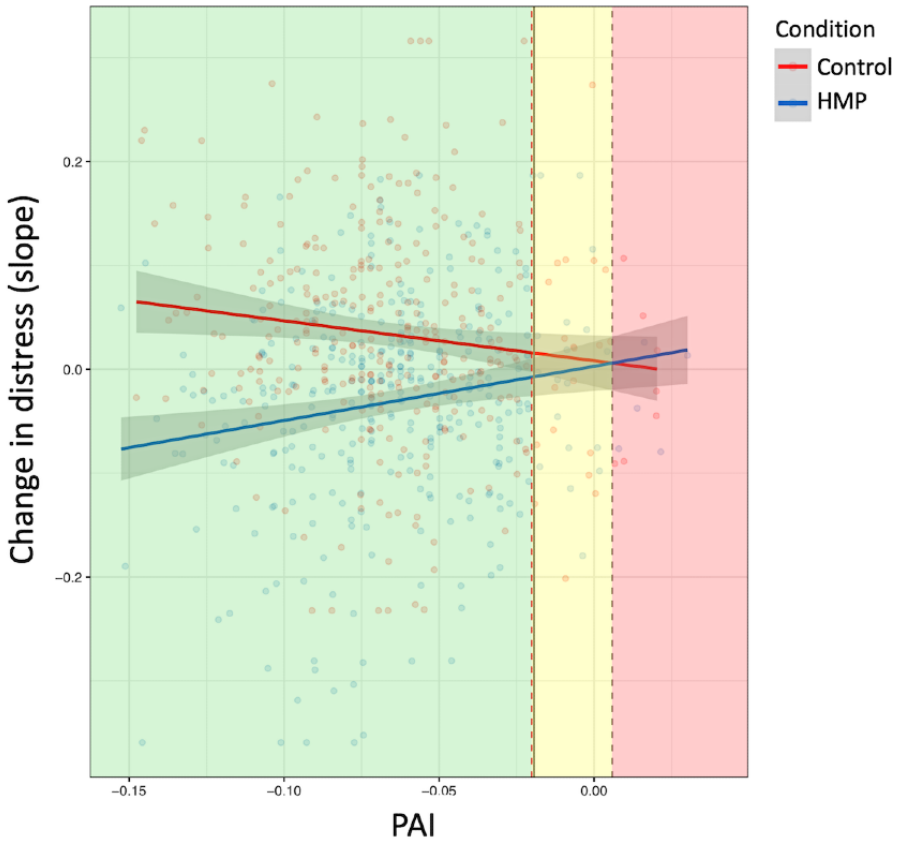
To demonstrate the translation of a predictive model to personalized recommendations, we used the parameter estimates from the above regression models to estimate predicted changes in distress in HMP versus the assessment-only condition for a new individual based on their preintervention repetitive negative thinking score. Given that the simpler regression model performed similarly to the more complex multivariable ENR models, we used the former model for this demonstration.

First, as shown in Figure 3, we plotted the relationship between PAI scores and outcomes for HMP (blue line) and the assessment-only control condition (red line). The dashed vertical gray line represents the point at which the 2 regression lines intersect. An individual with a PAI score to the left of this line was predicted to have a better outcome in HMP relative to the assessment-only control condition (and vice versa for individuals with PAI scores to the right of this line). The area to the left of this line is colored yellow, reflecting a "cautious recommendation" for app-based meditation training. Second, we computed a 95% CI via bootstrap resampling (Boot package in R) [63]. Specifically, we drew 1000 samples with replacement and recomputed the 2 regression lines and their intersection points in each of these samples. The dashed vertical red line represents the left margin of the 95% CI for this intersection point. In other words, if an individual's PAI score falls to the left of this line, our confidence in the predicted benefit of HMP relative to the assessment-only condition increases. Third, we also implemented the Johnson-Neyman technique [64] (Interactions package in R) to probe the group × PAI interaction and to estimate the value of the moderator (PAI) at which group differences in outcomes become statistically significant. This occurred at PAI<−0.02 (solid vertical gray line in Figure 3, immediately adjacent to the dashed red line). If a participant's PAI score falls to the left of both the 95% CI (dashed red line) and the Johnson-Neyman threshold (solid gray line), the plot area is colored green to reflect a more confident recommendation to use HMP.

To illustrate with a concrete example, an individual with a repetitive negative thinking (PTQ) score of 1 SD above the mean (ie, 41) would have a PAI score of −0.10 (within the "green zone" of Figure 3) and a predicted slope of change in distress of −0.049 (ie, expected reduction in distress) in HMP and 0.047 (ie, expected increase in distress) in the assessment-only condition over 4 weeks. Assuming that this individual had a preintervention level of distress at the 50th percentile, they would be predicted to be at the 41st percentile (relative to preintervention distress scores) following the 4-week mindfulness app course and at the 58th percentile if they only completed symptom assessments (ie, control condition). In summary, based on a brief assessment of perseverative negative thinking, our algorithm can provide individual users with useful information regarding their expected benefit before they decide to enroll in a multiweek course of app-based meditation training.

XSL•FO
RenderX

Figure 3. Plot of the relationship between Personalized Advantage Index (PAI) scores and outcome for each condition to inform personalized recommendations. The dashed vertical gray line indicates the point at which the 2 regression lines intersect (left margin of a bootstrapped 95% CI is shown with a dashed vertical red line). The solid vertical gray line (adjacent to the red line) is derived from the Johnson-Neyman technique and represents the value of the moderator (PAI) at which between-group differences in outcome become statistically significant. Refer to the detailed description in text, with an example for personalized Healthy Minds Program [HMP] recommendation.



Discussion

Principal Findings

An increasing number of individuals are turning to meditation apps to alleviate their emotional distress. Meditation apps represent the most commonly used mental health apps for depression and anxiety [30]. Despite their growing popularity, little is known about the benefits of these apps. In this study, we developed an algorithm to predict the benefit that an individual would be expected to experience from a smartphone-based meditation intervention (HMP) relative to an assessment-only control condition. We found evidence that a data-driven model can successfully predict differential response to a meditation app versus an assessment-only control condition using self-reported baseline demographic and clinical characteristics. Specifically, PAI scores significantly moderated group differences in outcomes. Individuals with more negative PAI scores, reflecting relatively stronger meditation app (ie, HMP) recommendations, had better outcomes if randomly assigned to the meditation app relative to the control condition. As expected, given overall group (ie, HMP > control) differences in outcome [31], the models typically predicted greater benefits from HMP versus the control condition. However, the predicted benefits of HMP were not always large, and in some cases, the PAI model predicted either relatively small between-group differences in outcome (“yellow zone” in Figure 3) or even better outcomes in the control condition (“red zone”). The former cases could be interpreted as instances in which the costs of engaging in a multiweek meditation app course (eg, time investment, delay in engaging with other, more helpful interventions) may not be worth the potential benefits.

Critically, a comparison linear regression model that only included information about baseline levels of repetitive negative thinking performed comparably well to a multivariable machine learning model (in contrast, refer to the studies by Webb et al [65] and Buckman et al [66]). Repetitive negative thinking moderated the outcome of app-based meditation training relative to the assessment-only control condition. Importantly, these findings reveal that higher repetitive negative thinking is not simply a general “prognostic” indicator of one’s likelihood of experiencing reductions in distress (eg, due to regression to the mean or the passage of time). In other words, greater repetitive negative thinking did not predict greater reductions in distress in *both* the meditation app and control conditions. Instead, and similar to prior research focused on a different mindfulness app and sample (adolescents with elevated rumination) [43,45], individuals with higher baseline levels of repetitive negative thinking derived greater relative benefit from a meditation app. One question is whether these findings are specific to repetitive negative thinking or instead may be driven by correlated clinical characteristics, in particular, depressive symptoms or distress. Sensitivity analyses revealed that repetitive negative thinking significantly moderated group differences in outcomes even when controlling for depressive symptom severity or distress

XSL•FO
RenderX

(Multimedia Appendix 1). In summary, these findings indicate that a brief self-report assessment of repetitive negative thinking can inform which individuals are most likely to benefit from app-based meditation training.

As illustrated in Figure 3, our predictive model can be readily applied for personalized meditation app recommendations for new individuals. First, the model provides a binary prediction of whether an individual is expected to experience greater reductions in distress from the meditation app relative to symptom assessment only (ie, based on whether PAI scores fall to the left or right of the intersection point [vertical dashed gray line]). Second, the model provides an estimate of the *magnitude* of the expected difference in outcomes between the meditation app and the control condition. Finally, the model also distinguishes between the strengths of recommendations to use the meditation app, demarcated by the green (confident recommendation) and yellow (cautious recommendations) zones of the figure (with boundaries defined by a bootstrapped CI and Johnson-Neyman interval). Collectively, this information can be used to provide individuals with objective metrics about expected outcomes to inform their decision on whether to enroll in a meditation app course. Such information could readily be implemented within mHealth interventions such as the HMP. Participants could first complete a brief self-report assessment of repetitive negative thinking and receive feedback on their predicted outcomes before deciding to use the app.

Although potentially useful in terms of encouraging the optimal use of users' time and attention, informing some individuals that engagement with a meditation app may not be beneficial to them is unlikely to be embraced by many intervention developers. However, these models can be readily extended to instances in which one or more mHealth interventions are being compared. Given the thousands of available mental health apps [25], which should be compared? One approach is to focus on the most popular (eg, most frequently downloaded) mental health apps, which include mindfulness, journaling, CBT, and mood tracking apps [29,30]. For example, future studies could develop algorithms for predicting response to various popular mental health apps, which differ substantially in intervention focus (eg, meditation apps vs CBT-based apps vs mood tracking) [29,67], or even compare a mental health app with conventional (in-person) psychotherapy or pharmacotherapy. Such studies could determine, for example, whether we can predict which individuals with depressive symptoms require conventional, face-to-face CBT (or an antidepressant prescription) versus those who would experience symptom remission from a brief app-based meditation or CBT course. In addition, future studies could compare different versions of a single app. For example, individuals may differ in the extent to which they benefit from different types of meditation (eg, cultivating focused attention on breath, open monitoring, or loving-kindness meditations) or different lengths or frequencies of guided meditation sessions.

In addition to informing consumer choice, the ability to predict who is most likely to benefit from a particular intervention could inform health care policy and decision-making. In contrast to a stepped care model in which treatment intensity is escalated based on the response to interventions, predictive models could be used to initially assign patients to the treatment expected to yield the best outcomes for that individual based on their baseline characteristics (ie, stratified care) [68]. In theory, the latter approach could minimize the delay in receiving an effective intervention.

Another important avenue for future research is to test the extent to which these findings can be generalized to other meditation apps (eg, Headspace and Calm). In many ways, HMP is similar to other meditation apps. It includes training in mindfulness and connection (eg, loving-kindness, compassion) practices that are also available in popular mindfulness apps such as Headspace and Calm. One difference is that HMP includes practices designed specifically to cultivate a healthy sense of self (Insight module) as well as purposes and meaning in life (Purpose module). The inclusion of these practices is derived from a neuroscience-based model of well-being on which HMP is based [36]. Thus, it is more accurate to view HMP as a meditation app that intentionally moves beyond mindfulness to place equal emphasis on other domains of well-being and contemplative practices designed to support these additional domains. Ultimately, additional research is needed to test whether the pattern of findings presented in this study generalize to other meditation and mindfulness apps.

Finally, given the lack of prior research predicting mental health app outcomes, further research is needed to test the impact of presenting predicted mindfulness app prognosis on patient outcomes. For example, before using a mindfulness app, patients could be randomly assigned to receive their predicted outcomes or not receive this information. Several relevant outcomes could be examined, including (1) between-group differences in symptom change, (2) the extent to which receiving these predictions influences expectancies of therapeutic benefit, (3) the relationship between expectancies and app outcome, and (4) the extent to which individuals use the algorithm-recommended intervention or disregard the recommendation.

Limitations

This study has several important limitations. First, although basing models exclusively on self-reported data is attractive from an implementation perspective, we may have excluded other patient characteristics that provide important additional predictive information to inform optimal treatment recommendations (eg, biomarkers and cognitive tasks) [12]. In addition, repetitive negative thinking, which emerged as a predictor of differential response, may be more validly assessed via methods other than conventional retrospective self-report questionnaires (eg, repeated, daily ecological momentary assessment [43,69]). Other relevant variables (eg, app use data, motivational variables, and involvement in other activities linked to better mental health) could be assessed in future studies. Second, our results emerged within a specific sample (school district employees), which did not have adequate representation of males, Black, Indigenous, and people of color, or individuals with low income. The sample is representative of Wisconsin in terms of race (83% of Wisconsinites are White) but includes a higher proportion of females. However, the gender difference in our sample is not surprising, given that females are more likely than males to (1) be employed as teachers [70] and (2)

XSL•FO
RenderX

experience and seek treatment for depressive and anxiety symptoms [71]. Third, we were unable to conduct external validation by evaluating model prediction performance in an entirely new sample (eg, from another RCT). Fourth, as is common in digital therapies [48], a sizable subset of participants used the app for relatively few days. However, the results remained significant when restricting our analyses to subsets of participants who used the app for a longer period (Multimedia Appendix 1). Fifth, we did not include an active comparison condition. Our assessment-only control condition was not designed to control for placebo-related processes [72]. The methods demonstrated here may ultimately be most relevant in helping patients and clinicians decide between competing interventions that are intended to be therapeutic.

Conclusions and Future Directions

This study demonstrated the potential utility of data-driven approaches for informing personalized meditation app recommendations. A natural extension of this study is to conduct a prospective test of our algorithm using a doubly randomized design. For example, participants could be randomized to either (1) random treatment assignment (ie, treatment A or treatment B) or (2) be assigned to their algorithm-indicated treatment. To the extent that patient outcomes are significantly (and clinically meaningfully) better in the latter condition, the results would support the clinical benefits of algorithm-informed treatment recommendations (for a recent example of a similar design testing predictive matching of patients to therapists, refer to the study by Constantino et al [73]). In addition to comparing treatment packages, this design could be readily used to evaluate other customizable elements of HMP or other mHealth interventions. This may include assignment to receive various components or ordering of components within HMP, assignment to HMP or an alternative commonly used mHealth intervention (eg, CBT, behavioral activation, journaling, or mood tracking apps), or assignment to varying treatment intensities (eg, meditation practice frequency).

Other potentially fruitful future directions include evaluating a broader set of patient characteristics previously shown or hypothesized to predict the likelihood of responding to different interventions [5]. In addition, prediction models could be developed using data drawn from large naturalistic data sets evaluating mHealth interventions, as has been done for in-person psychotherapy and pharmacotherapy [65,74-76]. In addition to testing the utility of these models in “real-world” settings, naturalistic settings often provide large data sets relative to RCTs and thus can increase statistical power [28]. Ultimately, these approaches may gradually help supplement our reliance on trial and error for treatment selection with empirically supported, data-driven algorithms to objectively communicate expected benefits to individuals, allowing them to make well-informed decisions about which interventions are best for their needs.

Acknowledgments

This research was supported by National Center for Complementary and Integrative Health grant K23AT010879 (SBG), National Institute for Mental Health grant R01MH43454 (RJD), Chan Zuckerberg Initiative Grant 2020-218037 (RJD), a National Academy of Education/Spencer Postdoctoral Fellowship (MJH), and funding from the Wisconsin Center for Education Research (SBG). The first author (CAW) was partially supported by National Institute for Mental Health R01MH116969, National Center for Complementary and Integrative Health R01AT011002, the Tommy Fuss Fund, and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. SBG was partially supported by the Hope for Depression Research Foundation, Defeating Depression Award.

Conflicts of Interest

RJD is the founder, president, and serves on the board of directors for the nonprofit organization Healthy Minds Innovations, Inc. MJH has been a paid consultant at Healthy Minds Innovations, Inc for work unrelated to this study.

Multimedia Appendix 1
Supplement.
[DOCX File , 440 KB - jmir_v24i11e41566_app1.docx]

Multimedia Appendix 2
CONSORT e-HEALTH checklist.
[PDF File (Adobe PDF File), 1068 KB - jmir_v24i11e41566_app2.pdf]

References

1. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795 [FREE Full text] [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
2. Paez J, Jänne PA, Lee J, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004 Jun 04;304(5676):1497-1500. [doi: 10.1126/science.1099314] [Medline: 15118125]

XSL•FO
RenderX

3. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Spanish Lung Cancer Group in collaboration with Groupe Français de Pneumo-CancérologieAssociazione Italiana Oncologia Toracica. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012 Mar;13(3):239-246. [doi: 10.1016/S1470-2045(11)70393-X] [Medline: 22285168]
4. Schwaederle M, Zhao M, Lee J, Eggermont A, Schilsky R, Mendelsohn J, et al. Impact of precision medicine in diverse cancers: a meta-analysis of phase II clinical trials. J Clin Oncol 2015 Nov 10;33(32):3817-3825. [doi: 10.1200/jco.2015.61.5997]
5. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Ebert DD, et al. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. Epidemiol Psychiatr Sci 2017 Feb;26(1):22-36 [FREE Full text] [doi: 10.1017/S2045796016000020] [Medline: 26810628]
6. Norcross JC, Wampold BE. A new therapy for each patient: evidence-based relationships and responsiveness. J Clin Psychol 2018 Nov 18;74(11):1889-1906. [doi: 10.1002/jclp.22678] [Medline: 30334258]
7. Hastie T, Tibshirani R, Friedman J. The Elements Of Statistical Learning: Data Mining, Inference, And Prediction, Second Edition. Cham: Springer; 2017.
8. Webb CA, Cohen ZD. Progress towards clinically informative data-driven decision support tools in psychotherapy. Lancet Digital Health 2021 Apr;3(4):e207-e208. [doi: 10.1016/s2589-7500(21)00042-x]
9. Boehmke B, Greenwell B. Hands-On Machine Learning with R. Boca Raton: Chapman and Hall/CRC; 2019.
10. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 2018 May 07;14(1):91-118. [doi: 10.1146/annurev-clinpsy-032816-045037] [Medline: 29401044]
11. Aafjes-van Doorn K, Kamsteeg C, Bate J, Aafjes M. A scoping review of machine learning in psychotherapy research. Psychother Res 2021 Jan 29;31(1):92-116. [doi: 10.1080/10503307.2020.1808729] [Medline: 32862761]
12. Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 2021 Jun 18;20(2):154-170 [FREE Full text] [doi: 10.1002/wps.20882] [Medline: 34002503]
13. Cohen ZD, DeRubeis RJ. Treatment selection in depression. Annu Rev Clin Psychol 2018 May 07;14:209-236. [doi: 10.1146/annurev-clinpsy-050817-084746] [Medline: 29494258]
14. Delgadillo J, Lutz W. A development pathway towards precision mental health care. JAMA Psychiatry 2020 Sep 01;77(9):889-890. [doi: 10.1001/jamapsychiatry.2020.1048] [Medline: 32459326]
15. DeRubeis RJ. The history, current status, and possible future of precision mental health. Behav Res Ther 2019 Dec;123:103506. [doi: 10.1016/j.brat.2019.103506] [Medline: 31706160]
16. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. PLoS One 2014 Jan 8;9(1):e83875 [FREE Full text] [doi: 10.1371/journal.pone.0083875] [Medline: 24416178]
17. Huibers MJ, Cohen ZD, Lemmens LH, Arntz A, Peeters FP, Cuijpers P, et al. Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. PLoS One 2015 Nov 10;10(11):e0140771 [FREE Full text] [doi: 10.1371/journal.pone.0140771] [Medline: 26554707]
18. Cohen ZD, Kim TT, Van HL, Dekker JJ, Driessen E. A demonstration of a multi-method variable selection approach for treatment selection: recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression. Psychother Res 2020 Feb 11;30(2):137-150. [doi: 10.1080/10503307.2018.1563312] [Medline: 30632922]
19. Webb CA, Trivedi MH, Cohen ZD, Dillon DG, Fournier JC, Goer F, et al. Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. Psychol Med 2018 Jul 2;49(07):1118-1127. [doi: 10.1017/s0033291718001708]
20. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. World Psychiatry 2017 Feb 26;16(1):90-99 [FREE Full text] [doi: 10.1002/wps.20388] [Medline: 28127925]
21. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
22. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
23. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct 09;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
24. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
25. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]

XSL•FO
RenderX

26. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 01;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
27. 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]
28. Luedtke A, Sadikova E, Kessler RC. Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. Clin Psychol Sci 2019 Jan 18;7(3):445-461. [doi: 10.1177/2167702618815466]
29. Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing mental health apps marketplaces with objective metrics from 29,190 data points from 278 apps. Acta Psychiatr Scand 2021 Aug 29;144(2):201-210. [doi: 10.1111/acps.13306] [Medline: 33835483]
30. Wasil AR, Gillespie S, Shingleton R, Wilks CR, Weisz JR. Examining the reach of smartphone apps for depression and anxiety. Am J Psychiatry 2020 May 01;177(5):464-465. [doi: 10.1176/appi.ajp.2019.19090905] [Medline: 32354266]
31. Hirshberg M, Frye C, Dahl C, Riordan K, Vack N, Sachs J, et al. A randomized controlled trial of a smartphone-based well-being training in public school system employees during the COVID-19 pandemic. J Educ Psychol 2022 Mar 17 [FREE Full text] [doi: 10.1037/edu0000739]
32. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
33. Markowitz A, Bassok D, Bellows L, Doromal J, Michie M, Smith A. Understanding early educators' responses to the coronavirus pandemic: lessons from large-scale surveys in Louisiana. In: Proceedings of the Association for Public Policy Analysis & Management Annual Meeting. 2020 Presented at: Association for Public Policy Analysis & Management Annual Meeting; Jul 23-24, 2020; Health Sciences Building.
34. Markowitz A, Ehrlich S, Gassman-Pines A. Panel - child care in the time of covid-19: new evidence from child care business owners, early educators, and parents to inform recovery efforts (family and child policy). In: Proceedings of the APPAM Virtual Fall Research Conference. 2020 Presented at: APPAM Virtual Fall Research Conference; Nov 13, 2020; Washington, DC.
35. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, PROMIS Cooperative Group. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment 2011 Sep;18(3):263-283 [FREE Full text] [doi: 10.1177/1073191111411667] [Medline: 21697139]
36. Dahl CJ, Wilson-Mendenhall CD, Davidson RJ. The plasticity of well-being: a training-based framework for the cultivation of human flourishing. Proc Natl Acad Sci U S A 2020 Dec 22;117(51):32197-32206 [FREE Full text] [doi: 10.1073/pnas.2014859117] [Medline: 33288719]
37. Goldberg SB, Imhoff-Smith T, Bolt DM, Wilson-Mendenhall CD, Dahl CJ, Davidson RJ, et al. Testing the efficacy of a multicomponent, self-guided, smartphone-based meditation app: three-armed randomized controlled trial. JMIR Ment Health 2020 Nov 27;7(11):e23825 [FREE Full text] [doi: 10.2196/23825] [Medline: 33245288]
38. Cohen S. Perceived stress in a probability sample of the United States. In: The Social Psychology of Health. Thousand Oaks, California: SAGE Publications; 1988.
39. Roberti J, Harrington L, Storch E. Further psychometric support for the 10-item version of the perceived stress scale. J College Counsel 2006;9(2):135-147.
40. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, et al. Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. J Clin Epidemiol 2016 May;73:119-127 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.036] [Medline: 26931289]
41. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med 2015 May 19;162(10):735-736. [doi: 10.7326/l15-5093-2]
42. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. Neuroimage 2017 Jan 15;145(Pt B):166-179. [doi: 10.1016/j.neuroimage.2016.10.038] [Medline: 27989847]
43. Webb C, Swords C, Murray L, Hilt L. App-based mindfulness training for adolescent rumination: predictors of immediate and cumulative benefit. Mindfulness (N Y) 2021 Oct;12(10):2498-2509 [FREE Full text] [doi: 10.1007/s12671-021-01719-0] [Medline: 35432625]
44. Wetherill RR, Rao H, Hager N, Wang J, Franklin TR, Fan Y. Classifying and characterizing nicotine use disorder with high accuracy using machine learning and resting-state fMRI. Addict Biol 2019 Jul 27;24(4):811-821 [FREE Full text] [doi: 10.1111/adb.12644] [Medline: 29949234]
45. Webb CA, Swords CM, Lawrence HR, Hilt LM. Which adolescents are well-suited to app-based mindfulness training? A randomized clinical trial and data-driven approach for personalized recommendations. J Consult Clin Psychol 2022 Sep;90(9):655-669 [FREE Full text] [doi: 10.1037/ccp0000763] [Medline: 36279218]
46. Watkins ER, Nolen-Hoeksema S. A habit-goal framework of depressive rumination. J Abnorm Psychol 2014 Feb;123(1):24-34. [doi: 10.1037/a0035540] [Medline: 24661156]

XSL•FO
RenderX

47. Watkins ER, Roberts H. Reflecting on rumination: consequences, causes, mechanisms and treatment of rumination. Behav Res Ther 2020 Apr;127:103573. [doi: 10.1016/j.brat.2020.103573] [Medline: 32087393]
48. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
49. Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating Second Edition. Cham: Springer; 2019.
50. Hirshberg MJ, Goldberg S, Davidson R. COVID and the Healthy Minds Program for Educators. Open Science Framework. 2020 Jun 10. URL: https://osf.io/eqgt7 [accessed 2022-10-24]
51. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. URL: https://www.R-project.org/ [accessed 2022-10-24]
52. Webb CA. R code for "Personalized prediction of response to smartphone-delivered meditation training" submission. Open Science Framework. 2021 Jul 27. URL: https://osf.io/94a6s [accessed 2022-10-24]
53. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): validation of a content-independent measure of repetitive negative thinking. J Behav Ther Exp Psychiatry 2011 Jun;42(2):225-232 [FREE Full text] [doi: 10.1016/j.jbtep.2010.12.003] [Medline: 21315886]
54. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
55. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, et al. Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. Assessment 2008 Sep 29;15(3):329-342. [doi: 10.1177/1073191107313003] [Medline: 18310597]
56. Cyranowski JM, Zill N, Bode R, Butt Z, Kelly MA, Pilkonis PA, et al. Assessing social support, companionship, and distress: National Institute of Health (NIH) toolbox adult social relationship scales. Health Psychol 2013 Mar;32(3):293-301 [FREE Full text] [doi: 10.1037/a0028586] [Medline: 23437856]
57. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the self-compassion scale. Clin Psychol Psychother 2011 Jun 08;18(3):250-255. [doi: 10.1002/cpp.702] [Medline: 21584907]
58. Forman EM, Herbert JD, Juarascio AS, Yeomans PD, Zebell JA, Goetter EM, et al. The Drexel defusion scale: a new measure of experiential distancing. J Contextual Behav Sci 2012 Dec;1(1-2):55-65. [doi: 10.1016/j.jcbs.2012.09.001]
59. Steger MF, Frazier P, Oishi S, Kaler M. The meaning in life questionnaire: assessing the presence of and search for meaning in life. J Counsel Psychol 2006 Jan;53(1):80-93. [doi: 10.1037/0022-0167.53.1.80]
60. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 2012 Jan 01;28(1):112-118. [doi: 10.1093/bioinformatics/btr597] [Medline: 22039212]
61. Schwartz B, Cohen Z, Rubel J, Zimmermann D, Wittmann W, Lutz W. Personalized treatment selection in routine care: integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. Psychother Res 2021 Jan;31(1):33-51. [doi: 10.1080/10503307.2020.1769219] [Medline: 32463342]
62. Team R. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
63. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. R package version. 2021. URL: https://www.bibguru.com/r/how-to-cite-r-package-boot/ [accessed 2022-10-21]
64. Hayes AF, Matthes J. Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. Behav Res Method 2009 Aug;41(3):924-936. [doi: 10.3758/brm.41.3.924]
65. Webb CA, Cohen ZD, Beard C, Forgeard M, Peckham AD, Björgvinsson T. Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: a comparison of machine learning approaches. J Consult Clin Psychol 2020 Jan;88(1):25-38 [FREE Full text] [doi: 10.1037/ccp0000451] [Medline: 31841022]
66. Buckman J, Cohen Z, O'Driscoll C, Fried E, Saunders R, Ambler G, et al. Predicting prognosis for adults with depression using individual symptom data: a comparison of modelling approaches. Psychol Med 2021 May 06:1-11. [doi: 10.1017/s0033291721001616]
67. Flett JA, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018 Oct 31;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
68. Lipton R, Stewart W, Stone A, Láinez MJ, Sawyer J, Disability in Strategies of Care Study group. Stratified care vs step care strategies for migraine: the Disability in Strategies of Care (DISC) Study: a randomized trial. JAMA 2000;284(20):2599-2605. [doi: 10.1001/jama.284.20.2599] [Medline: 11086366]
69. Webb CA, Israel ES, Belleau E, Appleman L, Forbes EE, Pizzagalli DA. Mind-wandering in adolescents predicts worse affect and is linked to aberrant default mode network-salience network connectivity. J Am Acad Child Adolesc Psychiatry 2021 Mar;60(3):377-387 [FREE Full text] [doi: 10.1016/j.jaac.2020.03.010] [Medline: 32553785]
70. Characteristics of public school teachers. National Center for Education Statistics. URL: https://nces.ed.gov/programs/coe/indicator/clr [accessed 2021-10-06]

XSL•FO
RenderX

71. Eaton NR, Keyes KM, Krueger RF, Balsis S, Skodol AE, Markon KE, et al. An invariant dimensional liability model of gender differences in mental disorder prevalence: evidence from a national sample. J Abnorm Psychol 2012 Feb;121(1):282-288 [FREE Full text] [doi: 10.1037/a0024780] [Medline: 21842958]
72. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9]
73. Constantino MJ, Boswell JF, Coyne AE, Swales TP, Kraus DR. Effect of matching therapists to patients vs assignment as usual on adult psychotherapy outcomes: a randomized clinical trial. JAMA Psychiatry 2021 Sep 01;78(9):960-969 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.1221] [Medline: 34106240]
74. Bone C, Simmonds-Buckley M, Thwaites R, Sandford D, Merzhvynska M, Rubel J, et al. Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data. Lancet Digital Health 2021 Apr;3(4):e231-e240. [doi: 10.1016/s2589-7500(21)00018-2]
75. Delgadillo J, Appleby S, Booth S, Burnett G, Carey A, Edmeade L, et al. The Leeds risk index: field-test of a stratified psychological treatment selection algorithm. Psychother Psychosom 2020;89(3):189-190. [doi: 10.1159/000505193] [Medline: 31940642]
76. Webb C, Forgeard M, Israel E, Lovell-Smith N, Beard C, Björgvinsson T. Personalized prescriptions of therapeutic skills from patient characteristics: an ecological momentary assessment approach. J Consult Clin Psychol 2022 Jan;90(1):51-60 [FREE Full text] [doi: 10.1037/ccp0000555] [Medline: 33829818]

Abbreviations

CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
CV: cross-validation
HMP: Healthy Minds Program
mHealth: mobile health
MLQ: Meaning in Life Questionnaire
PAI: Personalized Advantage Index
PROMIS: Patient-Reported Outcomes Information System
PTQ: Perseverative Thinking Questionnaire
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
RMSE: root mean square error

Edited by R Kukafka; submitted 31.07.22; peer-reviewed by Z Li, Y Wang, M Kapsetaki; comments to author 02.09.22; revised version received 03.09.22; accepted 26.09.22; published 08.11.22.

Please cite as:
Webb CA, Hirshberg MJ, Davidson RJ, Goldberg SB
Personalized Prediction of Response to Smartphone-Delivered Meditation Training: Randomized Controlled Trial
J Med Internet Res 2022;24(11):e41566
URL: https://www.jmir.org/2022/11/e41566
doi:10.2196/41566
PMID:36346668

©Christian A Webb, Matthew J Hirshberg, Richard J Davidson, Simon B Goldberg. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Addiction Symptom Network of Young Internet Users: Network Analysis

Jianxia Lu[1*], PhD; Qinhan Zhang[2*], BSc; Na Zhong[3*], MD, PhD; Jin Chen[1], MSc; Yujia Zhai[2], BSc; Lei Guo[3], MSc; Chunlei Lu[2], MSc; Tianzhen Chen[3], MD, PhD; Zhongli Jiang[4], PhD; Hui Zheng[3], MSc

[1]School of Rehabilitation, Jiangsu Vocational College of Medicine, Yancheng, China

[2]College of Teacher Education, Zhejiang Normal University, Jinhua, China

[3]Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[4]Department of Rehabilitation Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
Hui Zheng, MSc
Shanghai Key Laboratory of Psychotic Disorders
Shanghai Mental Health Center
Shanghai Jiao Tong University School of Medicine
600 Wan Ping Nan Road
Shanghai, 200030
China
Phone: 86 18267908575
Email: zh.dmtr@gmail.com

Abstract

Background: An increasing number of people are becoming addicted to the internet as a result of overuse. The Internet Addiction Test (IAT) is a popular tool for evaluating internet use behaviors. The interaction between different symptoms and the relationship between IAT and clinical diagnostic criteria are not well understood.

Objective: This study aimed to explore the core symptoms of internet addiction (IA) and the correlation between different symptoms of the IA symptom network. Network analysis was also conducted to explore the association between the IAT scale and the Diagnostic and Statistical Manual of Mental Disorders–5th edition (DSM-5) criteria for IA.

Methods: We recruited 4480 internet users (aged 14-24 years), and they completed the IAT. The final analysis included 63.50% (2845/4480) of the participants after screening the submitted questionnaires. Participants were classified into IA group and non-IA (NIA) group. By using partial correlation with Lasso regularization networks, we identified the core symptoms of IA in each group and compared the group differences in network properties (strength, closeness, and betweenness). Then, we analyzed the symptom networks of the DSM-5 diagnostic criteria and IAT scale for IA.

Results: A total of 12.47% (355/2845) of the patients were in the IA group and 87.52% (2490/2845) of the patients were in the NIA group, and both groups were evaluated for the following nodes: IAT_06 (school work suffers; strength=0.511), IAT_08 (job performance suffers; strength=0.531), IAT_15 (fantasize about being on the web; strength=0.474), IAT_17 (fail to stop being on the web; strength=0.526), and IAT_12 (fear about boredom if offline; strength=0.502). The IA groups had a stronger edge between IAT_09 (defensive or secretive about being on the web) and IAT_18 (hidden web time) than the NIA groups. The items in DSM-5 had a strong association with IAT_12 (weight=−0.066), IAT_15 (weight=−0.081), IAT_17 (weight=−0.106), IAT_09 (weight=−0.198), and IAT_18 (weight=−0.052).

Conclusions: The internet use symptom network of the IA group is significantly different from that of the NIA group. Nodes IAT_06 (school work affected) and IAT_08 (work performance affected) are the resulting symptoms affected by other symptoms, whereas nodes IAT_12 (fear about boredom if offline), IAT_17 (inability to stop being on the web), and IAT_15 (fantasize about being on the web) are key symptoms that activate other symptoms of IA and are strongly linked to the inability to control the intention to play games in the DSM-5.

(J Med Internet Res 2022;24(11):e38984) doi:10.2196/38984

XSL•FO
RenderX

KEYWORDS

internet addiction; Internet Addiction Test; network analysis; adolescents

Introduction

Background

Internet addiction (IA) refers to the inability of individuals to control their internet behaviors, which can also lead to dysfunction in their lives [1]. The prevalence of IA varies from 0.8% to 26.7% in different populations, with high prevalence in adolescents and young adults [2]. IA leads to serious dysfunction symptoms, which are often characterized by irritability, quarrels with people, increased lying, academic neglect, social withdrawal, and mental symptoms [3]. Currently, there is no unified standard for the diagnosis of IA. The recently proposed Assessment of Criteria for Specific Internet-use Disorders mainly refers to International Classification of Diseases 11th revision (ICD-11) [4]. The treatment of IA remains in the mapping stage [5]. However, the diagnosis and treatment of IA are complicated by the lack of clarity regarding the core symptoms and their interactions.

Addiction is generally considered to have symptoms such as abnormal craving desires, impaired inhibitory control, compulsive addictive behaviors, and abnormalities in negative emotion regulation [6]. By comparing ICD-11 and Diagnostic and Statistical Manual of Mental Disorders–5th edition (DSM-5) diagnostic criteria, we found 2 characteristics mentioned by both. First, people not only spend a lot of time and energy in playing games but, more importantly, they also ignore the realities of life and can no longer undertake the social roles they used to play and no longer participate in social life. Second, they lose control of their actions and let the games control their lives. The reason for both is that individuals cannot be satisfied in the real world but show their spiritual world on the internet to obtain a sense of achievement [7]. Such a perfect image, which is different from that of the real world, is brought about by the anonymity of the internet. Anonymity may amplify negative symptoms, thereby influencing other symptoms of IA [8]. The core symptoms are the most influential symptoms of a disorder, and they can activate other symptoms and promote the development of mental illness [9]. In general, identifying the core symptoms helps to target clinical interventions. Core symptoms are classified as key symptoms and outcome symptoms based on their different properties. Key symptoms are the core symptoms that have the greatest impact on other symptoms in the network, and outcome symptoms are the most affected by other symptoms [10].

In recent years, the emerging perspective of network analysis has provided new view and tools for understanding psychopathology [11]. In network analysis, which is different from the previous perspective, the existing models conceptualize psychiatric disease as a static structure based on underlying models, suggesting that observable clinical symptoms and signs are caused by underlying variables [12-16]. The symptom network perspective considers symptoms as the integrated component of mental disorders [9,17]. The network analysis approach can identify the most influential symptoms (core symptoms) in the symptom network, which are defined as the symptoms with high centrality. These influential symptoms can more actively affect other symptoms in the network, thus promoting the development of mental illness [9]; therefore, identifying the influential symptoms is helpful in identifying targets for clinical intervention. *Strength* and *edge* are two important indices in network analysis, and *strength* represents the total weight of the connections from other nodes to specific nodes. The association line (or *edge*) between the node is represented. *Edge* is a line between nodes that represents a regularization part. The wider the edge, the stronger the association.

Clinical symptoms of IA are often measured using questionnaires such as the Internet Addiction Test (IAT) [18,19]. The IAT is a widely used tool, and it shows adequate psychometric features [20]. There is evidence that the IAT is a valid and reliable scale for the screening of IA among Chinese adolescents [21]. The IAT scale contains a total of 20 items and can reflect IA from compulsive use behaviors, withdrawal symptoms, tolerance, interpersonal health, time management, and other aspects [21]. Although the total score of multiple items can reflect the degree of the disease, there is still no conclusion on how to treat the symptoms; one of the reasons is that the relationship between symptoms is not clear. IA affects adolescents' daily life; for instance, learning is an important part of life and is reflected in two items: IAT_06 (school work suffers) and IAT_08 (job performance suffers). In addition, IAT_09 (defensive or secretive about being on the web) and IAT_18 (hidden web time) involve the concealment of information related to internet access, which is related to internet anonymity; therefore, there may be a certain connection between them. Previous studies have explored the relationships between different symptoms from several perspectives [22], but there are only few conclusions [13]. A small sample of investigations on autism have found defensive and secretive behaviors and the concealment of internet use to be core symptoms [13], whereas a small sample of studies on depression have found that the most important bridge symptom is node IAT_11 ("Anticipation for future online activities"), followed by IAT_12 ("Fear that life is boring and empty without the internet") and IAT_19 ("Spend more time online over going out with others") [23]. However, studies of large samples of IA groups are still lacking.

Objectives

Therefore, the purpose of this study was to explore the core symptoms of IA using network analysis and the associations between different symptoms in the IA symptom network. We additionally included symptom measure entries from the DSM-5, similar to other 9-item Internet Gaming Disorder Scale–Short Form (IGDS9-SF) questionnaires [24]. On the basis of previous studies, we propose 2 hypotheses. Hypothesis 1 is that the symptom network is different between IA and non-IA (NIA) groups, and it is determined by comparing ICD-11 and DSM-5. Hypothesis 2 is that craving for the internet, loss of control, and negative emotional state are the main symptoms of IA [7].

XSL•FO
RenderX

Methods

Participants

Approximately 5900 college students—all with internet experience—from 2 universities in Jiangsu, China, were invited via advertisement to participate in this study. This study was a single-center sampling study. In September 2019, 15.18% (896/5900) of the participants completed the questionnaires. Then, in May 2020, 60.74% (3584/5900) of the participants completed the questionnaires. After we sifted through the 2 batches of questionnaires submitted, 36.49% (1635/4480) of the participants were excluded because they chose the same answer for ≥7 consecutive questions in the questionnaire. This was most likely because they did not fill in the questionnaire carefully, which will lead to inaccurate completion of the questionnaire and affect the accuracy of the final data analysis results. Therefore, 63.50% (2845/4480) of the participants were included in the final analysis. The first group of included participants reported their internet use history and IAT scale on a reliable, web-based data-collection survey platform in China [25]. The second batch had 9 DSM-5 questions about IA in addition to the original questionnaire. All available students (4480/4480, 100%) were fully informed of the purpose of the investigation and participated voluntarily.

Ethics Approval

This study was approved by the research ethics committee of Jiangsu Vocational College of Medicine (2020002). All participants in the survey and their parents and schools signed the informed consent forms.

Measures

The participants' demographic information, including gender, age, and internet use history (participants' common internet access methods and activities and internet access frequency) were collected using a homemade structured survey.

The IA severity of the participants was assessed using a self-evaluated instrument: the IAT scale [26]. The score for each item ranges from 1 (rarely) to 5 (always). The total score of the IAT scale ranges from 20 to 100, and a high score is indicative of severe IA. The IAT is a valid and reliable scale with satisfactory internal consistency (Cronbach α=.84). According to a previous study, IAT score ≥50 is indicative of IA [13]. Those with IAT scale score <50 were allocated into the NIA group. The abbreviations for the IAT scale are shown in Table 1.

The criteria for DSM-5 were revised with reference to the IGDS9-SF (which contains all 9 Internet Gaming Disorder criteria proposed by the American Psychological Association in the DSM-5) and were used to aid in the diagnosis of IA, along with the IAT score. The 9 questions used in this study were modified with reference to the IGDS9-SF, under the guidance of clinicians. Internal consistency in the present sample was good (Cronbach α=.74) [27]. IA diagnosis depends on nine diagnostic criteria, of which at least five must be met to be diagnosed with IA: (1) preoccupation with internet games, (2) withdrawal symptoms when not playing, (3) tolerance, (4) unsuccessful attempts to reduce or stop playing, (5) gives up other activities to play, (6) continues playing despite problems caused by it, (7) deceives or covers up playing, (8) plays to escape adverse moods, and (9) risks or losses in relationships or career opportunities because of excessive playing.

XSL•FO
RenderX

Table 1. Node abbreviations for IAT[a] scale.

Node	Item	Abbreviation
IAT_01	How often do you find that you stay online longer than you intended?	Stay on the web beyond schedule
IAT_02	How often do you neglect household chores to spend more time online?	Neglect household chores
IAT_03	How often do you prefer the excitement of the internet to intimacy with your partner?	Internet trumps intimacy
IAT_04	How often do you form new relationships with fellow online users?	Make new connections via the web
IAT_05	How often do others in your life complain to you about the amount of time you spend online?	Complained about being on the web for very long
IAT_06	How often do your grades or school work suffer because of the amount of time you spend online?	School work suffers
IAT_07	How often do you check your email before something else that you need to do?	Check email first
IAT_08	How often does your job performance or productivity suffer because of the internet?	Job performance suffers
IAT_09	How often do you become defensive or secretive when anyone asks you what you do online?	Defensive or secretive about being on the web
IAT_10	How often do you block out disturbing thoughts about your life with soothing thoughts of the internet?	Use the web to escape from emotion
IAT_11	How often do you find yourself anticipating when you will go online again?	Craving for next internet use
IAT_12	How often do you fear that life without the internet would be boring, empty, and joyless?	Fear about boredom if offline
IAT_13	How often do you snap, yell, or act annoyed if someone bothers you while you are online?	Annoyed at being interrupted
IAT_14	How often do you lose sleep due to being online?	Lose sleep
IAT_15	How often do you feel preoccupied with the internet when offline, or fantasize about being online?	Fantasize about being on the web
IAT_16	How often do you find yourself saying “just a few more minutes” when online?	Reluctant to be offline
IAT_17	How often do you try to cut down the amount of time you spend online and fail?	Fail to stop being on the web
IAT_18	How often do you try to hide how long you have been online?	Hidden web time
IAT_19	How often do you choose to spend more time online over going out with others?	Prefer using the web than going out
IAT_20	How often do you feel depressed, moody, or nervous when you are offline, which goes away once you are back online?	Web makes you feel better

[a]IAT: Internet Addiction Test.

Network Creation

The IAT symptom network model was constructed using all 2845 participants and analyzed with R software (version 3.6.1; R Foundation for Statistical Computing) using the *qgraph* package. In this study, the partial correlation network method was used to estimate all symptom networks, including the IAT network and DSM-5 internet-based symptom network. The edge of the network can be understood as partly related quantities. The estimation steps of each part of the related networks are as follows: to estimate the symptom network illustrating the relationship between IA and sleep disturbance symptoms, partial correlation-based sparse Gaussian graphical models with Lasso regularization were constructed. In this procedure, all edges in the network and sets of small edges were shrunk exactly to 0 (the Lasso regularization) [13]. This process of regularization was coupled with best-fit model selection by following an extended Bayesian information criterion, leading to a sparse network with explanatory power [13]. In the network, a circle represents an individual symptom (1 item from IAT or DSM-5). The association line (or *edge*) between the nodes is represented. An *edge* is a line between nodes that represents a regularization part. The existence of the edge indicates the dependence between the nodes of the IAT network. There is no indication of variables that cannot be independent (all other nodes in a given network). The wider the edge, the stronger the association.

To explore the network correlation between the IAT scale and participants’ DSM-5 items, we constructed the internet-based symptom network of these 2 scales using the same methods. For this internet-based symptom network, we further selected data from the second batch of 79.38% (2845/3584) participants.

Network Properties

We used the three common centrality measures of *strength*, *closeness*, and *betweenness* to quantify the features of the nodes in the IAT symptom network and IAT and DSM-5 internet-based symptom network [28]. *Strength* represents the total weight of the connections from other nodes to specific nodes. For example, in the Netherlands, a prospective longitudinal study of healthy adults, Boschloo et al [29] found that high-intensity minds lacking depressive symptoms (fatigue, depression, pleasure, and attention concentration) reached threshold levels of the

XSL•FO
RenderX

crowd; in the next 6 years, there was great risk of severe depressive episodes. *Closeness* is defined as the inverse of the sum of the shortest distances from a particular node to all other nodes in the network. The shortest distance is the minimum number of edges from one node to the next node. High internet closure indicates that the average distance between the given node and all other nodes in the network is short. In epidemiology, patients with high closeness are more likely to trigger the rapid development of an epidemic. *Betweenness* is the number of times the shortest path between any 2 symptoms passes through another symptom. Nodes with high betweenness can be considered as *bridges* to other symptoms; that is, if you remove a high node, the distance between the other nodes generally increases. *Expected influence* is a measurement of centrality that quantifies how strongly and directly a symptom node is associated with all the other nodes in the network. The results from a longitudinal study of older bereaved adults showed that decreased symptom severity with high expected influence predicted clinical improvements across the network compared with complex grief symptoms with low expected influence. Centrality is measured as the shortest path length of any 2 symptoms, and a highly mediated symptom can be considered as a bridge that connects other symptoms.

For network symptoms, nodes with symptoms with high strength can be considered as core symptoms because symptoms of nodes with high strength are more closely related to other symptoms. According to the psychopathological network theory, when a high-strength node is activated, the probability of its symptoms being further activated is relatively high. Symptoms with low node strength can be considered as marginal [28].

Network Stability Estimates

The accuracy of the edge and stability estimation of the network were calculated using a bootstrapping process of 1000 iterations. First, we estimated the accuracy of the edge of the 95% CI by bootstrapping the edge weight. The overlap between these CIs indicates less accuracy. Second, we tested the stability of the *node* center through a subset of the bootstrapping process. We estimated the central stability (CS) coefficient (CS factor) as a reference indicator. A CS coefficient weight that equals 0.25 indicates acceptable stability.

Network Comparison

We compared the IAT symptom networks of IA and NIA samples using the Network Comparison Test package in R to explore possible differences in the overall connectivity between IA and NIA. The comparison was based on a permutation procedure, and the number of permutations was 5000.

Results

Participant Characteristics

We included 63.50% (2845/4480) of the participants after screening, with 12.47% (355/2845) of the patients in the IA group (mean score 57.32, SD 7.70; mean age 19.36, SD 1.07 years; 89/355, 25.1% men) and 87.52% (2490/2845) of the patients in the NIA group (mean score 34.75, SD 7.87; mean age 18.59, SD 1.69 years; 938/2490, 37.67% men). The present sample was aged 14 to 24 years, and most of them (1203/2845, 42.28%) were aged 18 or 19 years. There were significant differences in age and IAT scores between the IA and NIA groups. There were also significant gender differences in IAT scores.

A descriptive analysis of the internet use habits of 69.49% (1977/2845) of the participants after the second screening (Table 2) indicated that mobile phones were the most dominant way of accessing the internet. Most participants (882/1977, 44.61%) spent an average of 3 to 6 hours a day on the web in the past month.

XSL•FO
RenderX

Table 2. Sample characteristics and internet use habits (n=1977[a]).

	IA[b] (n=344, 17.40%)	NIA[c] (n=1633, 82.59%)
Age in years, mean (SD)	19.6 (0.852)	19.7 (0.938)
Gender, n (%)		
Men	113 (32.8)	392 (24)
Women	231 (67.2)	1241 (75.99)
Main internet access modes, n (%)		
Mobile phone	338 (98.3)	1598 (97.85)
Computer	4 (1.2)	28 (1.71)
Tablet computer	2 (0.6)	5 (0.31)
Other[d]	0 (0)	2 (0.12)
Time spent on the web every day in the last month (hours), n (%)		
0-3	36 (10.5)	359 (21.98)
3-6	130 (37.8)	752 (46.05)
6-9	91 (26.5)	359 (21.98)
9-12	43 (12.5)	77 (4.72)
>12	44 (12.8)	86 (5.27)

[a]Of a total of 2845 participants included, 1977 (69.49%) completed self-reports of internet use.

[b]IA: internet addiction.

[c]NIA: non–internet addiction.

[d]In addition to the 3 modes mentioned, the participants' main mode of access to the internet.

Symptom Networks of IA

In Figure 1, the network diagram shows the conditional associations among the 20 questions in the IAT scale for the IA group. The network diagram shows the correlation (circles) and predictability estimates (outer circles) for the 20 items in the IAT scale in the IA group. Each circle in the figure represents an item of the named nodes in the network. The edges represent the strengths of the associations between the nodes. The outer circles represent the level of predictability. The meaning of the title represented by each node abbreviation is provided in Table 1. The most predictable node was IAT_06. Nodes IAT_06 and IAT_08 had the strongest connections.

Centrality measures for the IA symptom network represent the strength, closeness, betweenness, and expected influence value of each node. Higher values indicate that the item is more central in the network. Nodes IAT_06, IAT_08, IAT_17, IAT_12, and IAT_15 were the highest in the strength centrality indices. This indicates that the total weights of the connections from other nodes to nodes IAT_06, IAT_08, IAT_17, IAT_12, and IAT_15 were the highest. Node IAT_15 had the highest closeness, indicating that this cognitive symptom was influential in connecting other symptoms that were otherwise unrelated in the network. Nodes IAT_15 and IAT_17 were the highest in the betweenness centrality indices. This indicates that these 2 highly mediating nodes can be regarded as a bridge to connect other symptoms. Nodes IAT_06, IAT_08, IAT_17, IAT_12, and IAT_15 were the highest in expected influence, indicating that they had the strongest positive associations with other nodes; however, not all other centrality indices were the highest. The meaning of the title represented by each node abbreviation is provided in Table 1.

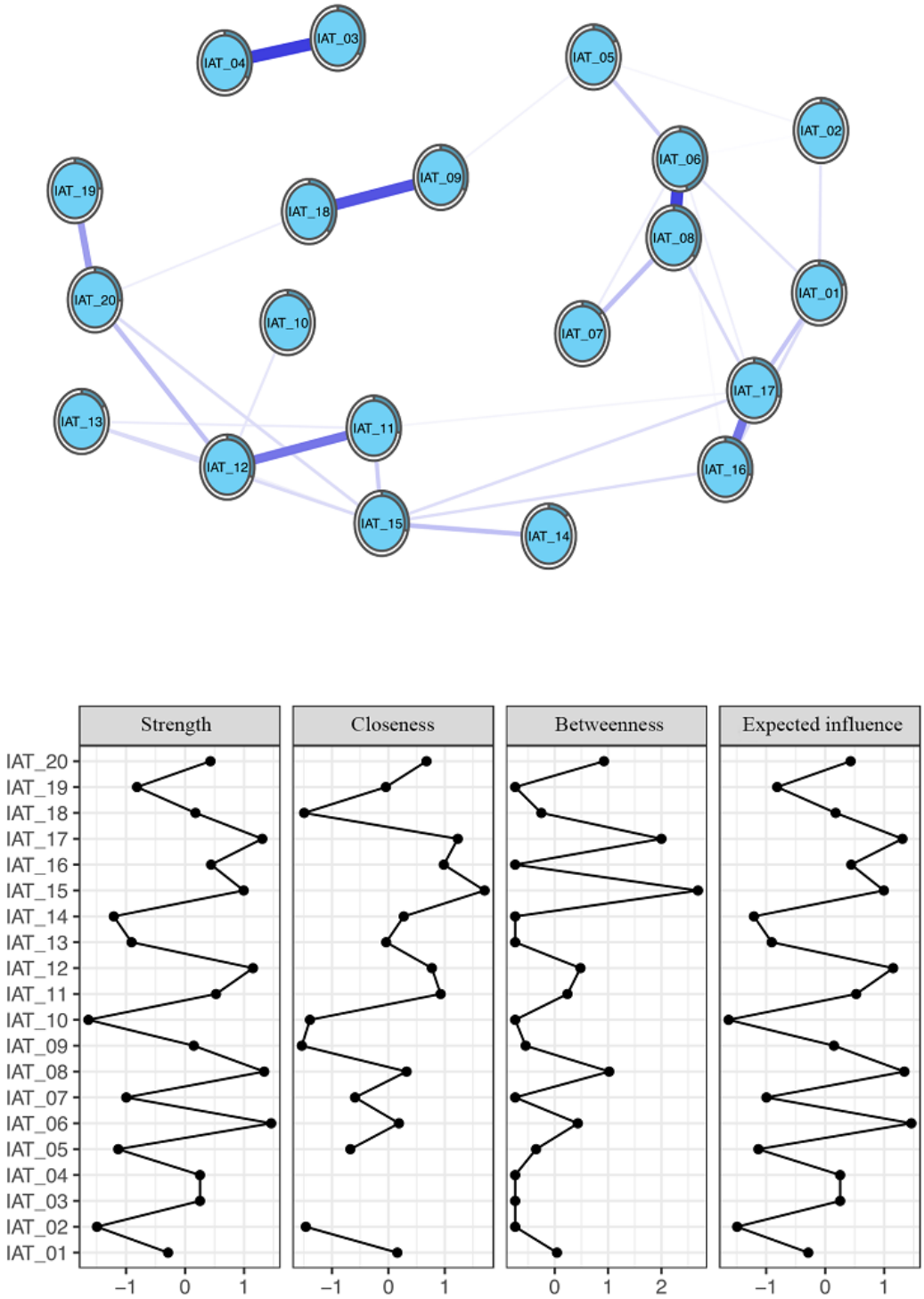
Figure 1 shows the constructed symptom network of the IAT scale for the addiction group. As illustrated in Figure 2, IAT_06 (school work suffers; strength=0.511; betweenness=48) and IAT_08 (job performance suffers; strength=0.531; betweenness=72) had high strength.

Moreover, IAT_12 (fear about boredom if offline; strength=0.500; betweenness=50), IAT_17 (fail to stop being on the web; strength=0.526; betweenness=112), and IAT_15 (fantasize about being on the web; strength=0.474; betweenness=140) also had high strength on the IAT scale. These 5 nodes were the 5 points with the highest strength. It is worth noting that these nodes had the highest value on the *strength* measures. These nodes were also significantly higher than those for all other symptoms when comparing the expected influence values.

Furthermore, we found that the values of the *closeness* of IAT_03 (internet trumps intimacy) and IAT_04 (make new connections via the web) were not shown. *Closeness* is defined as the inverse of the sum of the shortest distances from a particular node to all other nodes in the network. According to Figure 1, there is no connection with the other nodes, except for the connection between these 2 nodes; therefore, the total value of both nodes is very extreme and not shown in the figure.

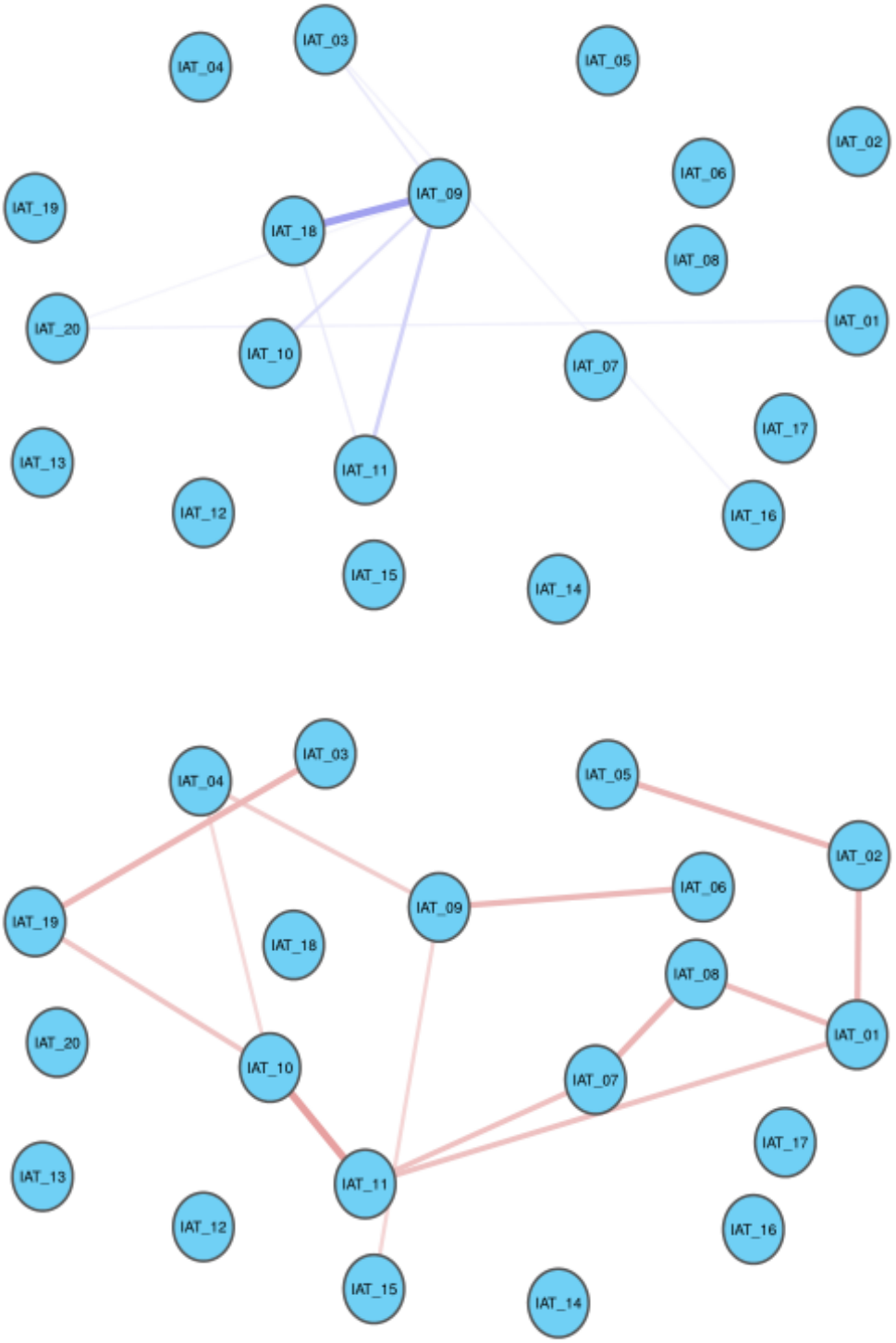
XSL•FO
RenderX

Figure 1. Internet addiction symptom network and centrality measures of the internet addiction group. IAT: Internet Addiction Test.



XSL•FO
RenderX

Figure 2. Edges exhibiting significant differences between internet addiction and non–internet addiction groups. IAT: Internet Addiction Test.



Comparison of Networks Between the IA and NIA Groups

In Figure 2, the blue edges denote the increased correlations between items in IA compared with those in the NIA group, and the red edges denote the decreased correlations. Edges exhibit significant differences between the IA group and NIA group. The correlation between IAT_18 and IAT_09 in the IA group was significantly stronger than that in the NIA group. The correlation between IAT_10 and IAT_11 in the IA group was significantly weaker than that in the NIA group. The meaning of the title represented by each node abbreviation is provided in Table 1.

We used a permutation-based method to compare edges exhibiting significant differences between individuals with IA and those without IA (NIA); significant positive and negative correlations are shown in Figure 2. Compared with individuals without IA (NIA), IAT_09 (defensive or secretive about being

XSL•FO
RenderX

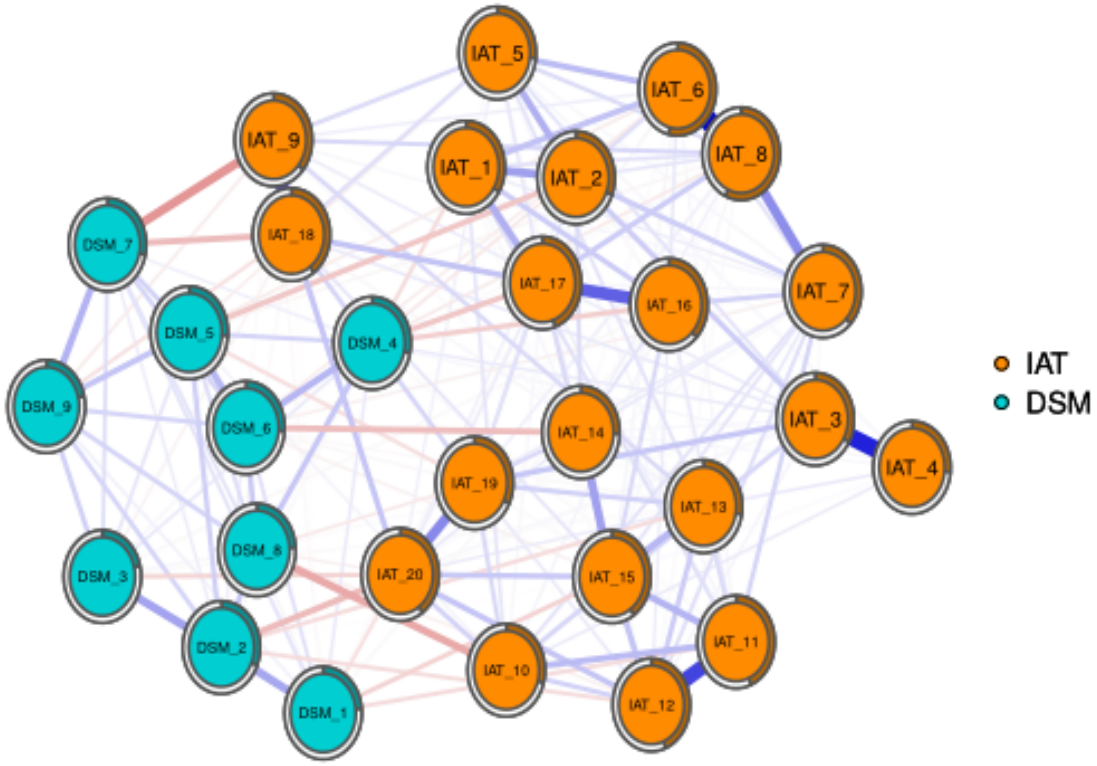
on the web) gained strong connections with IAT_18 (hidden web time; mean difference=0.172; *P*=.01).

In contrast, the symptoms of IAT_11 (craving for next internet use) showed decreased connection with IAT_10 (being on the web to escape from emotion; mean difference=−0.174; *P*=.003). Similarly, IAT_09 (defensive or secretive about being on the web) had weak connection with IAT_15 (fantasize about being on the web; mean difference=−0.073; *P*=.03).

Symptom Networks of IA and DSM-5

In Figure 3, the network graph shows associations and predictability estimates of 20 items in IAT and 9 items in the DSM-5 diagnostic criteria for IA. Each circle in the diagram represents a term of the named node in the network. The outer circle represents the predictable level. The most predictable node was IAT_08. The edges represent the strengths of the associations between the nodes. Nodes IAT_06 and IAT_08 had the strongest connections. The meaning of the title represented by each node abbreviation is provided in Table 1.

Figure 3. Internet addiction test (IAT) and Diagnostic and Statistical Manual of Mental Disorders–5th edition (DSM-5) symptom network.



The estimated network included 29 nodes, 20 items in IAT, and 9 items in the DSM-5 diagnostic criteria for IA. The estimated network yielded 406 edges (29×[29–1]/2), among which 199 edges had non-0 weights. The weight of the edge connecting IAT_06 (school work suffers) and IAT_08 (job performance suffers) was the strongest (0.479). Other strong associations included those between IAT_03 (internet trumps intimacy) and IAT_04 (make new connections via the web; 0.414) and between IAT_11 (craving for next internet use) and IAT_12 (fear about boredom if offline; 0.353).

Figure 3 shows that node DSM_1 in the DSM-5 diagnostic criteria for the IA community had the strongest weight with node IAT_15 (fantasize about being on the web) in the IAT community (−0.081). Node DSM_2 had a strong weight with node IAT_20 (being on the web makes you feel better; −0.125) and node IAT_12 (fear about boredom if offline; −0.066). Node DSM_4 had the strongest weight with node IAT_17 (failing to stop being on the web; −0.106). Node DSM_7 had the strongest weight with node IAT_09 (defensive or secretive about being on the web; −0.198). Node DSM_9 had the strongest weight with node IAT_18 (hidden web time; −0.052).

Discussion

Principal Findings

In this study, the core symptoms of IA were identified by conducting a descriptive analysis of the network symptoms of the IAT scale in the IA and NIA groups. In addition, the difference between the 2 networks was determined by network comparison. From this, we found 2 main results. First, the core symptoms of the IA group were IAT_06 (school work suffers), IAT_08 (job performance suffers), IAT_12 (fear about boredom if offline), IAT_17 (fail to stop being on the web), and IAT_15 (fantasize about being on the web). Second, there were differences in the network between the IA group and NIA group. Together, these results partially support our hypotheses. The results support hypothesis 1 and partially support hypothesis 2. In hypothesis 2, craving the internet and loss of control are the

XSL•FO
RenderX

core symptoms of IA, but the results show that a negative emotional state is not the core symptom of IA. Interestingly, the result showed that IAT_09 (defensive or secretive about being on the web) had a strong connection with IAT_18 (hidden web time). This result was also supported by the network connection with DSM-5 symptoms; for example, anonymity was associated with community. Thus, the anonymity of the internet is a potential factor affecting the behavioral performance of IA. The DSM-5 diagnostic criteria for IA have high correlation with key symptoms and low correlation with outcome symptoms.

Craving, losing control, and boredom appeared to play a key role in the development of IA. Our study found that IAT_06 (school work suffers), IAT_08 (job performance suffers), IAT_12 (fear about boredom if offline), IAT_17 (fail to stop being on the web), and IAT_15 (fantasize about being on the web) were the core symptoms of IA, which was consistent with the results of previous studies that found that two items—"Life is boring and empty without the internet" and "Anticipation for future online activities"—were central to the IA network in both groups [13]. Studying is an important part of college students' lives. Excessive use of the internet has an impact on study and life. This study showed that IAT_06 (school work suffers) and IAT_08 (job performance suffers) nodes were closely related to other symptoms, and other symptoms (eg, IAT_17—fail to stop being on the web) lead to IAT_06 and IAT_08, which affect daily learning and life. Therefore, the core symptoms, IAT_06 and IAT_08, were the most affected by other symptoms.

Interestingly, the other nodes IAT_12 (fear about boredom if offline), IAT_17 (fail to stop being on the web), and IAT_15 (fantasize about being on the web) were key symptoms that activated other IA symptoms. Previous studies have found that individuals with IA have low self-esteem, high feelings of loneliness, and poor social skills [30]. The anonymity of the internet makes it easy to socialize. Individuals with IA are more eager to obtain social satisfaction in the web-based world because their social needs cannot be met in real life. This also causes individuals with IA to rely on the internet and ignore the real world, which has a certain impact on people's daily lives [31]. After leaving the internet, their sense of estrangement in the real world makes people miss the internet more [32], they cannot control their web-based behavior, and they are dominated and occupied by games. Therefore, for individuals with IA, more attention should be given to building harmonious social circles for them, teaching them certain social skills, and enhancing their self-esteem. In addition, cognitive behavioral therapy may be beneficial because the fear that life without web-based activity will become boring and empty is a core belief of individuals who have difficulty in controlling their internet use. Previous studies have reported the efficacy of cognitive behavioral therapy in reducing IA, reducing internet use, and improving time management skills and emotional stability [13].

Furthermore, connection analysis showed that IAT_09 (defensive or secretive about being on the web) gained strong connections with IAT_18 (hidden web time) when comparing individuals with IA and those without IA (NIA). Both nodes involve the hiding of information related to internet access and anonymity of the internet. Studies have shown that the special nature of web-based communication (anonymity, lack of visual indicators for social discomfort, lack of physical presence, etc) can promote people's self-disclosure [33,34]. In a web-based environment, where there is no need to be seen, people can change their identities on the web and act as if they are someone else [35]. Therefore, they do not want to let the people around them know, and they want to hide information about the network. The anonymity of the internet is a potential factor affecting the behavioral performance of individuals with IA. This may be related to adolescents' desire to avoid discipline and stigma. It also suggests that we need to pay attention to the reasons adolescents mask their IA time, especially because Chinese adolescents are generally under academic pressure, playing games is seen as bad behavior, and excessive game use may also make them feel shame and fear about being blamed. Parents and teachers should pay more attention to the balanced development of adolescents and establish good communication relationships with them so that mental health problems can be identified earlier. In addition, for adolescents who are at high risk for addiction, diagnostic assessments should focus on reliable information provided by those around them, rather than on just the individual's representations, so that people addicted to the internet can be identified more accurately and early [36]. The symptom network of IA and DSM-5 shows that the DSM-5 diagnostic criteria for IA are associated with key symptoms (IAT_12, IAT_15, and IAT_17) and have low correlation with outcome symptoms (IAT_06 and IAT_08). Our study results have found that the symptom networks of IA and DSM-5 are stable. Every criterion of the DSM-5 has a connection with a node of the IAT. On the basis of the IAT scale, the analysis of the core symptoms of IA shows that the diagnostic criteria of DSM-5 have a strong connection with IAT_09 (defensive or secretive about being on the web), IAT_12 (fear about boredom if offline), IAT_15 (fantasize about being on the web), IAT_17 (fail to stop being on the web), and IAT_18 (hidden web time), whereas the connections with IAT_06 (school work suffers) and IAT_08 (job performance suffers) are weak. IAT_06 and IAT_08 mainly emphasize the influence of IA on daily life and learning, whereas the diagnostic criteria of IA in DSM-5 may not lead to judgments of the direct influence of IA on daily learning and life. In addition, the concealment of addictive behavior is a common manifestation of the social stigma associated with addiction in general [37], and in adolescents, this concern can exacerbate other neuropsychiatric symptoms [38]. These findings may provide a practical basis for the diagnostic criteria of DSM-5 IA.

Limitations and Strengths

This study has several limitations. First, the sample size of the IA group was significantly smaller than that of the NIA group. Second, this study did not explore the effect of other intriguing variables, such as emotion regulation ability, social function, and personality traits. Third, as man and woman students may have different web-based behaviors, future studies can include gender as an independent variable to compare the differences between genders in the IA symptom network. Several points of strength should also be considered. To the best of our knowledge, so far, few studies have used the network analysis

XSL•FO
RenderX

model and IAT scale to analyze IA. Traditional game addiction interventions often lack targeted goals. This study suggests that there are core symptoms of game addiction that are closely related to other symptoms. Intervention for these core symptoms may lead to better efficacy. Previous studies have suggested this, and a network analysis of the Sequenced Treatment Alternatives to Relieve Depression study has revealed the importance of *sad mood* and *anhedonia* in nonpsychotic depressive disorder [39]. However, further studies are needed to confirm this finding. In addition, we analyzed the network differences between the IA samples and NIA samples to make the study's conclusions more reliable for the development of new intervention measures.

Conclusions

In conclusion, the network is different between IA and NIA. The influence on daily life and learning is the outcome symptom of IA. Craving, losing control, and boredom are key symptoms that activate other symptoms of IA. Withholding web-based information is a key symptom of IA. The anonymity of the internet is a potential factor affecting IA. The DSM-5 diagnostic criteria for IA are associated with key symptoms and have low correlation with outcome symptoms. Determining the core symptoms of IA and the link between IAT and DSM-5 diagnostic criteria are helpful in improving the pertinence and effectiveness of IA treatment.

Acknowledgments

The authors would like to thank the school that provided the study sites and the students who participated in the study. This study was supported by the guiding scientific research projects of the Department of Health of Jiangsu Province (grant Z201319), Jiangsu Higher Vocational Education High-level Specialty Group Construction Project (2021; number 1), National Nature Science Foundation (82171485), Qihang Project of Shanghai Mental Health Center (2019-QH-05), Sailing Project of Shanghai Rising-star Program (22YF1439200), Provincial Foreign Specialists Hundred Talents Program Project (BX2021081), and High-end Foreign Experts Introduction Program Project (G2021014067L). The funding agencies did not contribute to the experimental design or conclusions.

Data Availability

The analysis code and original data are available at the Open Science Framework [40].

Authors' Contributions

JL and QZ conducted the analysis and wrote the original manuscript. JC and YZ collected the data and participated in the literature collecting and organizing process. NZ, LG, and CL modified the manuscript. HZ, ZJ, and TC designed the study and revised the manuscript. All authors contributed to the manuscript and approved the submitted version.

Conflicts of Interest

None declared.

References

1. Vadher SB, Panchal BN, Vala AU, Ratnani IJ, Vasava KJ, Desai RS, et al. Predictors of problematic internet use in school going adolescents of Bhavnagar, India. Int J Soc Psychiatry 2019 Mar 11;65(2):151-157. [doi: 10.1177/0020764019827985] [Medline: 30744449]
2. Shadzi MR, Salehi A, Vardanjani HM. Problematic internet use, mental health, and sleep quality among medical students: a path-analytic model. Indian J Psychol Med 2020 Apr 01;42(2):128-135. [doi: 10.4103/ijpsym.ijpsym_238_19]
3. Zhai Q, Feng L, Zhang G, Liu M, Wang J. Research status of internet addiction in adolescents. Chinese General Practice 2020;23(13):1687-1694. [doi: 10.12114/j.issn.1007-9572.2019.00.558]
4. Müller SM, Wegmann E, Oelker A, Stark R, Müller A, Montag C, et al. Assessment of criteria for specific internet-use disorders (ACSID-11): introduction of a new screening instrument capturing ICD-11 criteria for gaming disorder and other potential internet-use disorders. J Behav Addict 2022 Apr 07;11(2):427-450 [FREE Full text] [doi: 10.1556/2006.2022.00013] [Medline: 35394924]
5. Zhang MX, Wu AM. Effects of smartphone addiction on sleep quality among Chinese university students: the mediating role of self-regulation and bedtime procrastination. Addict Behav 2020 Dec;111:106552. [doi: 10.1016/j.addbeh.2020.106552] [Medline: 32717501]
6. Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. Lancet Psychiatry 2016 Aug;3(8):760-773 [FREE Full text] [doi: 10.1016/S2215-0366(16)00104-8] [Medline: 27475769]
7. Jo YS, Bhang SY, Choi JS, Lee HK, Lee SY, Kweon Y. Clinical characteristics of diagnosis for internet gaming disorder: comparison of DSM-5 IGD and ICD-11 GD diagnosis. J Clin Med 2019 Jun 28;8(7):945 [FREE Full text] [doi: 10.3390/jcm8070945] [Medline: 31261841]
8. Misoch S. Stranger on the internet: online self-disclosure and the role of visual anonymity. Comput Human Behav 2015 Jul;48:535-541. [doi: 10.1016/j.chb.2015.02.027]

XSL•FO
RenderX

9. Borsboom D, Cramer AO. Network analysis: an integrative approach to the structure of psychopathology. Annu Rev Clin Psychol 2013 Mar 28;9(1):91-121. [doi: 10.1146/annurev-clinpsy-050212-185608] [Medline: 23537483]
10. Fried EI, Cramer AO. Moving forward: challenges and directions for psychopathological network theory and methodology. Perspect Psychol Sci 2017 Nov 05;12(6):999-1020. [doi: 10.1177/1745691617705892] [Medline: 28873325]
11. McNally RJ. Can network analysis transform psychopathology? Behav Res Ther 2016 Nov;86:95-104. [doi: 10.1016/j.brat.2016.06.006] [Medline: 27424882]
12. Chen L, Zhou H, Gu Y, Wang S, Wang J, Tian L, et al. The neural correlates of implicit cognitive bias toward internet-related cues in internet addiction: an ERP study. Front Psychiatry 2018;9:421 [FREE Full text] [doi: 10.3389/fpsyt.2018.00421] [Medline: 30245642]
13. Hirota T, McElroy E, So R. Network analysis of internet addiction symptoms among a clinical sample of Japanese adolescents with autism spectrum disorder. J Autism Dev Disord 2021 Aug 10;51(8):2764-2772. [doi: 10.1007/s10803-020-04714-x] [Medline: 33040268]
14. Pan N, Yang Y, Du X, Qi X, Du G, Zhang Y, et al. Brain structures associated with internet addiction tendency in adolescent online game players. Front Psychiatry 2018;9:67 [FREE Full text] [doi: 10.3389/fpsyt.2018.00067] [Medline: 29559931]
15. So R, Makino K, Fujiwara M, Hirota T, Ohcho K, Ikeda S, et al. The prevalence of internet addiction among a Japanese adolescent psychiatric clinic sample with autism spectrum disorder and/or attention-deficit hyperactivity disorder: a cross-sectional study. J Autism Dev Disord 2017 Jul 4;47(7):2217-2224. [doi: 10.1007/s10803-017-3148-7] [Medline: 28474226]
16. Tang J, Yu Y, Du Y, Ma Y, Zhang D, Wang J. Prevalence of internet addiction and its association with stressful life events and psychological symptoms among adolescent internet users. Addict Behav 2014 Mar;39(3):744-747. [doi: 10.1016/j.addbeh.2013.12.010] [Medline: 24388433]
17. Borsboom D. A network theory of mental disorders. World Psychiatry 2017 Feb 26;16(1):5-13 [FREE Full text] [doi: 10.1002/wps.20375] [Medline: 28127906]
18. Li J, Lau JT, Mo PK, Su X, Tang J, Qin Z, et al. Insomnia partially mediated the association between problematic internet use and depression among secondary school students in China. J Behav Addict 2017 Dec 01;6(4):554-563 [FREE Full text] [doi: 10.1556/2006.6.2017.085] [Medline: 29280394]
19. Li D, Liau A, Khoo A. Examining the influence of actual-ideal self-discrepancies, depression, and escapism, on pathological gaming among massively multiplayer online adolescent gamers. Cyberpsychol Behav Soc Netw 2011 Sep;14(9):535-539. [doi: 10.1089/cyber.2010.0463] [Medline: 21332374]
20. Stojković I, Vukosavljević-Gvozden T, David I, Purić D. Internet addiction in adolescents: a rational emotive and cognitive behavior therapy perspective. J Rat-Emo Cognitive-Behav Ther 2022 Mar 04:1-24. [doi: 10.1007/s10942-022-00445-3]
21. Lai C, Mak K, Watanabe H, Ang RP, Pang JS, Ho RC. Psychometric properties of the internet addiction test in Chinese adolescents. J Pediatr Psychol 2013 Aug;38(7):794-807. [doi: 10.1093/jpepsy/jst022] [Medline: 23671059]
22. Brand M, Rumpf H, King DL, Potenza MN, Wegmann E. Clarifying terminologies in research on gaming disorder and other addictive behaviors: distinctions between core symptoms and underlying psychological processes. Curr Opin Psychol 2020 Dec;36:49-54 [FREE Full text] [doi: 10.1016/j.copsyc.2020.04.006] [Medline: 32480020]
23. Cai H, Bai W, Sha S, Zhang L, Chow IH, Lei S, et al. Identification of central symptoms in internet addictions and depression among adolescents in Macau: a network analysis. J Affect Disord 2022 Apr 01;302:415-423. [doi: 10.1016/j.jad.2022.01.068] [Medline: 35065088]
24. Evren C, Dalbudak E, Topcu M, Kutlu N, Evren B, Pontes HM. Psychometric validation of the Turkish nine-item Internet Gaming Disorder Scale-Short Form (IGDS9-SF). Psychiatry Res 2018 Jul;265:349-354. [doi: 10.1016/j.psychres.2018.05.002] [Medline: 29793049]
25. Ranxing Information Technology. URL: https://www.wjx.cn/ [accessed 2022-10-17]
26. Young KS. The research and controversy surrounding internet addiction. Cyberpsychol Behav 1999 Oct;2(5):381-383. [doi: 10.1089/cpb.1999.2.381] [Medline: 19178209]
27. Boudard M, Alexandre J, Kervran C, Jakubiec L, Shmulewitz D, Hasin D, et al. Item response theory analyses of diagnostic and statistical manual of mental disorders, fifth edition (DSM-5) criteria adapted to screen use disorder: exploratory survey. J Med Internet Res 2022 Jul 27;24(7):e31803 [FREE Full text] [doi: 10.2196/31803] [Medline: 35896018]
28. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths. Social Networks 2010 Jul;32(3):245-251. [doi: 10.1016/j.socnet.2010.03.006]
29. Boschloo L, Schoevers RA, van Borkulo CD, Borsboom D, Oldehinkel AJ. The network structure of psychopathology in a community sample of preadolescents. J Abnorm Psychol 2016 May;125(4):599-606. [doi: 10.1037/abn0000150] [Medline: 27030994]
30. Ghassemzadeh L, Shahraray M, Moradi A. Prevalence of internet addiction and comparison of internet addicts and non-addicts in Iranian high schools. Cyberpsychol Behav 2008 Dec;11(6):731-733. [doi: 10.1089/cpb.2007.0243] [Medline: 18954277]
31. Bagir A, Emre O, Birgul C, Ulutas A. The relationship between social exclusion (ostracism) and internet addiction of adolescent girls. Istraživanja u pedagogiji 2020;10(1):50-65. [doi: 10.5937/istrped2001050b]
32. Nalwa K, Anand AP. Internet addiction in students: a cause of concern. Cyberpsychol Behav 2003 Dec;6(6):653-656. [doi: 10.1089/109493103322725441] [Medline: 14756932]

XSL•FO
RenderX

33. Joinson AN. Self-disclosure in computer-mediated communication: the role of self-awareness and visual anonymity. Eur J Soc Psychol 2001 Mar;31(2):177-192. [doi: 10.1002/ejsp.36]
34. McKenna KY, Green AS, Gleason ME. Relationship formation on the internet: what's the big attraction? J Social Isssues 2002 Jan;58(1):9-31. [doi: 10.1111/1540-4560.00246]
35. Leung L. Impacts of net-generation attributes, seductive properties of the internet, and gratifications-obtained on internet use. Telematics Informatics 2003 May;20(2):107-129. [doi: 10.1016/s0736-5853(02)00019-9]
36. Li N, Li D. A preliminary study on the cognitive basis of teenagers' internet addiction. J Gannan Normal University 2006(02). [doi: 10.13698/j.cnki.cn36-1037/c.2006.02.005]
37. Su B, Zheng M. Effects of substance-related cues on response inhibition in addicts. Adv Psychol Sci 2022 Jul 13;27(11):1863-1874. [doi: 10.3724/SP.J.1042.2019.01863]
38. Wu PH, Yizhi, Yongfeng Z, Peng, Shoulong SC. A study on personality characteristics of adolescents with internet addiction. CNKI 2022. [doi: 10.16719/j.cnki.1671-6981.2010.01.063]
39. Fried EI, Epskamp S, Nesse RM, Tuerlinckx F, Borsboom D. What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. J Affect Disord 2016 Jan 01;189:314-320. [doi: 10.1016/j.jad.2015.09.005] [Medline: 26458184]
40. Open Source Framework. URL: https://osf.io/k6mdq/ [accessed 2022-10-17]

Abbreviations

CS: central stability
DSM-5: Diagnostic and Statistical Manual of Mental Disorders–5th edition
IA: internet addiction
IAT: Internet Addiction Test
ICD-11: International Classification of Diseases–11th revision
IGDS9-SF: 9-item Internet Gaming Disorder Scale–Short Form
NIA: non–internet addiction

Edited by R Kukafka; submitted 24.04.22; peer-reviewed by SC Lu, W Ceron, Q Hu; comments to author 05.07.22; revised version received 30.08.22; accepted 16.09.22; published 10.11.22.

Please cite as:
Lu J, Zhang Q, Zhong N, Chen J, Zhai Y, Guo L, Lu C, Chen T, Jiang Z, Zheng H
Addiction Symptom Network of Young Internet Users: Network Analysis
J Med Internet Res 2022;24(11):e38984
URL: https://www.jmir.org/2022/11/e38984
doi:10.2196/38984
PMID:36355402

©Jianxia Lu, Qinhan Zhang, Na Zhong, Jin Chen, Yujia Zhai, Lei Guo, Chunlei Lu, Tianzhen Chen, Zhongli Jiang, Hui Zheng. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Connectedness in the Jackson Heart Study: Cross-sectional Study

Pramod Anugu[1], MSc; Md Abu Yusuf Ansari[2], MSc; Yuan-I Min[1], PhD; Emelia J Benjamin[3], MD; Joanne Murabito[3], MD; Karen Winters[1], PhD; Erica Turner[1], PhD; Adolfo Correa[1], MD

[1]University of Mississippi Medical Center, Jackson, MS, United States

[2]Department of Data Science, School of Population Health, University of Mississippi Medical Center, Jackson, MS, United States

[3]Boston University, Boston, MA, United States

Corresponding Author:
Pramod Anugu, MSc
University of Mississippi Medical Center
2500 N State St
Jackson, MS, 39216
United States
Phone: 1 6018155771
Email: panugu@umc.edu

Abstract

Background: Although new approaches for data collection, such as mobile technology and teleresearch, have demonstrated new opportunities for the conduct of more timely and less costly surveys in community-based studies, literature on the feasibility of conducing cardiovascular disease research using mobile health (mHealth) platforms among middle-aged and older African Americans has been limited.

Objective: The purpose of this study was to contribute to the knowledge regarding the penetrance of internet and mobile technologies, such as cellphones or smartphones in existing large cohort studies of cardiovascular disease.

Methods: A digital connectedness survey was conducted in the Jackson Heart Study (JHS), a Mississippi-based African American cohort study, as part of the annual follow-up calls with participants from July 2017 to February 2019.

Results: Of the 4024 participants contacted, 2564 (63.7%) completed the survey. Among survey respondents, 2262 (88.2%) reported use of internet or cellphone, and 1593 (62.1%) had a smartphone. Compared to nonusers (n=302), internet or cellphone users (n=2262) were younger (mean age 80.1, SD 8.0 vs 68.2, SD 11.3 years), more likely to be affluent (n=778, 40.1% vs n=39, 15.4%), and had greater than high school education (n=1636, 72.5% vs n=85, 28.1%). Internet or cellphone users were less likely to have cardiovascular disease history compared to nonusers (136/2262, 6.6% vs 41/302, 15.8%). The prevalence of current smoking and average BMI were similar between internet or cellphone users and nonusers. Among internet or cellphone users, 1316 (58.3%) reported use of email, 504 (22.3%) reported use of apps to track or manage health, and 1269 (56.1%) expressed interest in using JHS-developed apps.

Conclusions: Our findings suggest that it is feasible to use mHealth technologies to collect survey data among African Americans already enrolled in a longitudinal study. Our findings also highlight the need for more efforts to reduce the age and education divide in access and use of internet and smartphones for tracking health and research in African American communities.

(J Med Internet Res 2022;24(11):e37501) doi:10.2196/37501

KEYWORDS

teleresearch; mobile technology; cardiovascular disease; Jackson Heart Study; mobile phone

Introduction

Over decades, data collection in epidemiological longitudinal research into cardiovascular disease (CVD) has relied in part on costly and time-consuming surveys administered by phone or in person [1-4]. However, with the recent rapid growth in the use of digital or mobile technologies in the general population, there is now an opportunity to adopt eHealth or mobile health (mHealth) apps that are less expensive and more practical in research settings. mHealth as a part of eHealth can be defined as medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices,

XSL•FO
RenderX

personal digital assistants, and other wireless devices [5]. Over the past decades, mHealth app design and its role on the health behavior change among study participants has been discussed in several studies [6-8]. A qualitative study conducted on adults with type 1 diabetes showed that a well-designed mHealth app could serve to inform lifestyle choices, diabetes self-management tasks, preemptive self-care actions, and improved discussions with clinicians [9]. These evidences suggest that well-designed mHealth app can assist individuals especially older adults in independent living and self-management of (chronic) illnesses. mHealth apps have the prospect of participant convenience, collection of 'real-world' data, flexibility to pivot research agendas to focus on timely health issues (such as disparities in CVD and the long-term effects of COVID-19 pandemic), as well as the potential to be conducted in large and diverse populations. Although there have been recent efforts to examine mHealth apps for data collection in community-based studies [10,11], the literature on the feasibility of using mHealth applications in populations of middle-aged and older African Americans is limited.

The purpose of our study was to contribute to the knowledge regarding the penetrance of internet and mobile technologies, such as cellphones or smartphones in existing large cohort studies of CVD. Our study was modeled after a survey conducted in the Framingham Heart Study (FHS). The FHS cohort is a cohort of predominately White participants, largely based in New England; as such the findings from FHS may not be generalizable to other racial and ethnic groups or geographic regions. In this study, we conducted a digital connectedness survey in the Jackson Heart Study (JHS), a large cohort of exclusively middle-aged and older African Americans. We also examined the sociodemographic characteristics and CVD risk factors associated with digital connectedness. We have hypothesized that digitally connected individuals would be relatively younger, have higher educational attainment, and have an inverse association with CVD risk factor profiles compared to their counterparts who are not digitally connected.

Methods

The JHS is a large prospective community-based observational study designed to investigate risk factors for CVD in African Americans. Details of the JHS study design, recruitment, and data collection have been described previously [12,13]. Briefly, 5306 African American participants residing in the Jackson metropolitan area in Mississippi were recruited for the study between 2000 and 2004. Three research center visits were conducted to date (visit 1 [baseline]: 2000-2004; visit 2: 2005-2008; and visit 3: 2009-2012). Additionally, the participants were contacted once a year by telephone (annual follow-up) to ascertain events and vital status as well as update health data and contact information.

Ethical Considerations

All study participants provided written informed consent. The JHS was approved by the institutional review boards (Protocol 1998-6004) of Jackson State University, Tougaloo College, the Mississippi State Department of Health, and the University of Mississippi Medical Center.

Digital Connectedness Survey

All living JHS participants contacted for the annual follow-up calls between July 2017 and February 2019 were invited to participate in a digital connectedness survey. The survey collected data on the use of internet, cellphone, smartphone, email, health apps, social media, and computer games. Additionally, smartphone users were asked whether they had any software apps that helped them track or manage health and about the types of health apps they had. They were also asked about any technology that was used to store health readings digitally.

Sociodemographic Characteristics and CVD Risk Factors

Sociodemographic characteristics examined in this study included sex, age, smoking status, BMI, education, and income, collected at the baseline (visit 1). CVD risk factors examined included diabetes, hypertension, hyperlipidemia, CVD history, collected at visit 3. Participants' age was calculated at the time of the survey. Age was grouped into the following 4 levels: <65; ≥65 to < 75; ≥75 to <85; and ≥85 years. Education was categorized as follows: "less than high school"; "high school graduate or Graduate Education Development certificate"; and "attended vocational school, trade school, or college." Income was categorized into 4 groups according to the US census poverty levels based on household income and family size. The categories were assigned as "poor," representing income lower than poverty level; "lower-middle," representing income 1 to 1.5 times the poverty level; "upper-middle," representing income >1.5 but <3.5 times the poverty level; and "affluent," representing income ≥3.5 times the poverty level [14,15]. Diabetes was defined based on the following: fasting glucose ≥126 mg/dL; HbA1c ≥6.5%; use of diabetes medication; or self-reported diabetes [16]. Hypertension was defined as blood pressure ≥140/90 mm Hg (per the seventh report of the Joint National Committee) or use of blood pressure lowering medication [17]. Hyperlipidemia was defined as total fasting cholesterol ≥200 mg/dL or taking any statin medication.

Classification of Digital Connectedness and Participant Characteristics

For the purpose of this study, we defined digital connectedness as use of internet and cellphone; specifically, respondents who gave affirmative responses to the questions "Do you use the internet at least occasionally?" or "Do you use a cell phone?" were considered digitally connected. A negative response to both questions classified the respondents as not digitally connected. Instead of smartphone users, combining cellphone users and the internet users would allow us to connect with a larger cohort of African American population.

Statistical Analysis

All categorical data were reported as counts and relative frequencies as percentages. Continuous data were presented as mean (SD). We compared the characteristics between respondents and nonrespondents of the digital connectedness survey using logistic regression models adjusted for age and sex. We compared the sociodemographic characteristics and CVD risk factors between users and nonusers of internet or

XSL•FO
RenderX

cellphone, using logistic regression models adjusted for age and sex. The associations between CVD risk factors and internet and cellphone use were reported as adjusted odds ratios (AORs) and 95% CIs. A 2-sided *P* value <.05 was considered statistically significant for all analyses. All statistical analyses were performed using R statistical software (version 4.0.2; The R Foundation for Statistical Computing).

Results

Of the 4024 JHS participants contacted, 2564 (63.7%) completed the survey. The mean age of the respondents was 69.6 (SD 11.6) years, and 64.1% (1644/2564) of them were female. Compared to the nonrespondents, respondents were older (mean age 69.6, SD11.6 years vs mean age 64.8, SD 12.3 years; *P*<.001); more likely to have completed less than high school or attended vocational school, trade school, or college (1721/2564, 67.3% vs 972/1460, 66.8%; overall *P*<.001); more affluent (817/2194, 37.2% vs 324/1216, 26.6%; overall *P*<.001); and less likely to be a current smoker (267/2545, 10.5% vs 207/1445, 14.3%; *P*=.002; Table S1 in Multimedia Appendix 1). Among the 2564 respondents, 2262 (88.2%) were users of internet or cellphone (internet users 1507, 58.9%; cellphone users 2230, 87.1%). Use of a smartphone was reported in 1593 (62.1%) respondents.

Comparison of Sociodemographic Characteristics and CVD Risk Factors Between Internet or Cellphone Users and Nonusers

Compared to nonusers, internet and cellphone users were younger (mean age 68.2, SD 11.3 years vs mean age 80.1, SD 8.0 years; *P*<.001); they were more likely to have attended vocational school, trade school, or college (1636/2262, 72.5% vs 85/302, 28.1%; overall *P*<.001); have higher income (affluent 778/2262, 40.1% vs 39/302, 15.4%; overall *P*<.001); and be employed (716/2262, 31.7% vs 7/302, 2.3%; overall *P*<.001; Table 1). Internet or cellphone users were less likely to have a history of CVD (136/2262, 6.6% vs 41/302, 15.8%; AOR 0.58, 95% CI 0.39-0.89) and diabetes (624//2262, 30.3% vs 108/302, 41.7%; AOR 0.75; 95% CI 0.56-0.99) compared to nonusers (Table 1; Figure 1). There was no difference in sex distributions between users and nonusers of internet or cellphone. The prevalence of current smoking, hypertension, hyperlipidemia, and average BMI were similar in internet or cellphone users and nonusers. Similar patterns were observed comparing users and nonusers of internet, cellphone, and smartphone (Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Table 1. Sociodemographic characteristics and cardiovascular disease (CVD) risk factors by internet or cellphone use among digital connectedness survey respondents.

Characteristics	Internet or cellphone use, n (%)		
	Nonuser, n=302 (11.8)	User, n=2262 (88.2)	*P* value[a]
Sex, n (%)			.11
Female	220 (72.8)	1424 (63)	
Male	82 (27.2)	838 (37)	
Age (years)[b], mean (SD)	80.1 (8.0)	68.2 (11.3)	<.001
Age (years), n (%)			<.001
<65	13 (4.3)	864 (38.2)	
≥65 to <75	51 (16.9)	666 (29.4)	
≥75 to <85	150 (49.7)	601 (26.6)	
≥85	88 (29.1)	131 (5.8)	
Education, n (%)			<.001
Less than high school	88 (29.1)	389 (17.2)	
High school graduate or GED[c]	129 (42.7)	232 (10.3)	
Attended vocational school or trade school	85 (28.1)	1636 (72.5)	
Income, n (%)			<.001
Poor	55 (21.7)	175 (9)	
Lower-middle	95 (37.5)	357 (18.4)	
Upper-middle	64 (25.3)	631 (32.5)	
Affluent	39 (15.4)	778 (40.1)	
Diabetes, n (%)	108 (41.7)	624 (30.3)	.05
Hypertension, n (%)	219 (84.6)	1504 (72.6)	.73
Hyperlipidemia, n (%)	207 (79.9)	1416 (68.3)	.22
Current smoker, n (%)	31 (10.3)	236 (10.5)	.99
BMI (kg/m^2), mean (SD)	31.2 (6.8)	32.1 (6.7)	.48
Cardiovascular disease history, n (%)	41 (15.8)	136 (6.6)	.01
Employed, n (%)			<.001
Employed	7 (2.3)	716 (31.7)	
Homemaking	1 (0.3)	7 (0.3)	
Retired	285 (94.4)	1423 (63)	
Unemployed	9 (2.9)	114 (5)	

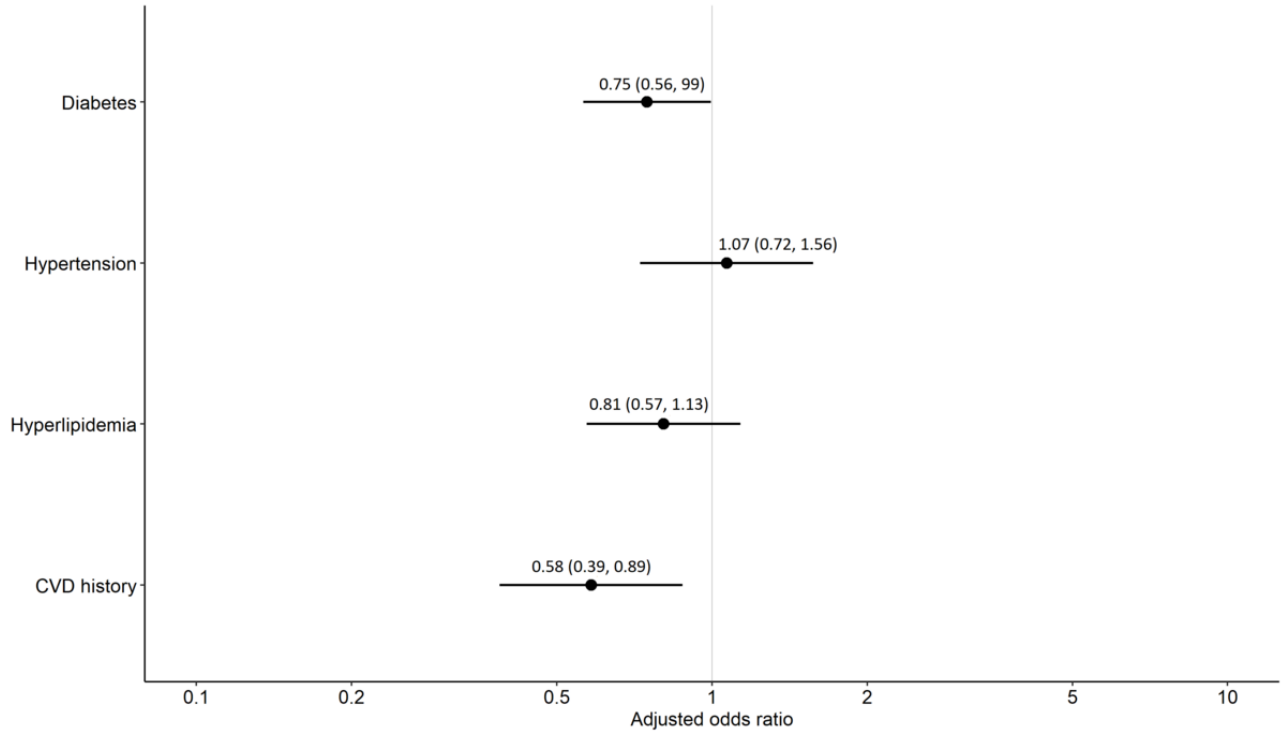
[a]*P* values are age- and sex-adjusted using logistic regression models.

[b]Age at the time of digital connectedness survey.

[c]GED: Graduate Education Development.

XSL•FO
RenderX

Figure 1. Adjusted odds ratio (95% CI) of cardiovascular disease (CVD) risk factors by internet or cellphone users versus nonusers. Odd ratios are age- and sex-adjusted using logistic regression models. Vertical line represents null effect.



Experience with the Use of Technologies (Apps) Among Internet or Cellphone Users

Among the 2262 internet or cellphone users, 1316 (58.3%) reported use of email; 504 (22.3%) reported use of software applications to track or manage health; and 1565 (69.2%) reported use of digital health technology, which stores health readings digitally (Table 2). Using social media to keep in touch with friends and family was reported among 830 (36.7%) internet or cellphone users, and 797 (35.5%) internet or cellphone users reported playing games on their devices.

When asked about whether they would be interested in using a JHS-developed mobile phone app to respond to health questions, 1269/2262 (56.1%) internet or cellphone users expressed interest, 353 (15.6%) expressed possible interest (by answering “not sure” or “maybe”), and 640 (28.3%) expressed no interest. Internet or cellphone users were significantly more inclined toward using JHS-developed apps as compared to the nonusers (1269/2262, 56.1% vs 14/302, 4.6%; $P<.001$).

XSL•FO
RenderX

Table 2. Experiences with technology or apps use among internet or cellphone users (N=2262).

Internet or cellphone users	Values, n (%)
Email, n=1316 (58.3)	
Can read new email	1316 (58.3)
Can use the reply feature	981 (43.4)
Can send an email	1001 (44.3)
Can open a file attached to an email	842 (37.2)
Apps to track health or manage health, n=504 (22.3)	
Exercise, fitness, pedometer, or heart rate monitor	484 (21.4)
Diet, food, calorie counter	46 (2)
Blood pressure	16 (0.7)
Weight	23 (1)
Blood sugar or diabetes	7 (0.3)
Medication management	7 (0.3)
Sleep	4 (0.2)
Mood	7 (0.3)
Other	6 (0.3)
Digital health technology, n=1565 (69.2)	
Digital blood pressure cuff	1387 (61.3)
Digital scale	549 (24.3)
Digital glucometer	418 (18.5)
Other	3 (0.1)
Interest in using the JHS[a] app in the future, n=2262 (100)	
Yes	1269 (56.1)
No	640 (28.3)
Not sure or maybe	353 (15.6)
Social media, n=830 (36.7)	
Facebook	784 (34.7)
LinkedIn	94 (4.2)
Google Plus	46 (2.0)
Twitter	83 (3.7)
Other	61 (2.7)
Playing games, n=797 (35.5)	
On computer	270 (11.9)
On tablet	337 (14.9)
On smartphone	517 (22.9)
On video game console	37 (1.6)
On other devices	5 (0.2)

[a]JHS: Jackson Health Study.

Discussion

Principal Findings

Overall, 63.7% of the JHS participants surveyed responded to our digital connectedness survey. Compared to nonrespondents, survey respondents were older, more affluent, less likely to be current smokers, and less likely to have a history of CVD. Among survey respondents, close to 90% were users of internet and cellphone, and over half of them had a smartphone. Among users of internet or cellphone, close to 60% reported use of email, over 20% had used software apps to track or manage

XSL•FO
RenderX

health and reported use of social media or played games on their devices, and over 70% expressed interest or possible interest in using an app to answer health questions. Participants interest for using a JHS-developed app was encouraging. We can integrate the app as an additional alternative tool for data collection. This could lead us to lower the participants' burden. We can receive ratings and reviews within the app from the app users to determine the success and possible places for improvement.

Strengths and Limitations

An important strength of our survey is that it is based on a community-based cohort of middle-aged and older African American men and women residents of Jackson, Mississippi who have been active participants in annual follow-up interviews in a longitudinal research study (ie, the JHS). Another strength is the high level of participation achieved, which probably reflects a high level of trust and commitment of participants with the JHS. Our survey also has some limitations. First, the survey was conducted on middle-aged to older adult African American residents of Jackson, Mississippi, and participants in the JHS. As such, the generalizability of our survey findings to other age groups, geographic regions, and race or ethnicities is uncertain. Second, the study was completed in 2019; it may not reflect the current prevalence of internet or cellphone use, which may have increased over time and thus may limit generalizations to more contemporary settings. Also, the sociodemographic characteristics used from visit 1 may not necessarily represent the characteristics of the participants in 2019. Third, although most JHS participants reported access to internet or cellphone, the extent to which they would use internet or cellphone to respond to research surveys or for ongoing health surveillance remains to be determined. Moreover, using cellphones for research requires participants' contact information to be updated regularly. In addition to that, the extent to which African Americans enrolled in a new community-based research project would participate in mHealth remains to be determined.

Lastly, there are potential risks to African American community arising from the promotion of secondary (and often commercial) uses of participant data; discriminatory profiling and inaccurate health status notifications because of algorithmic issues may undermine the trust in this new form of research [18,19].

Digital collection, monitoring, and transmission of data always involve some challenges related to data security and privacy that are often different from paper-based surveys. To avoid any security risk, a rigorous encryption policy needs to be employed. In JHS, the storage devices are bitLocker encrypted, with advanced encryption standard 256 enabled. Secure Sockets Layer (Transport Layer Security 1.2) certificate is installed and enforced. A natural barrier to using mHealth technologies is linked to physical limitation, which relates to the older age groups and comorbidities among the participants. Any attempt to design mHealth apps for older age groups needs to address motivational and cognitive barriers to older adults [20]. More research needs to be done to identify the reasons why certain populations such as older age groups in African American communities do not use mHealth technologies and if there are any improvements on using mHealth technologies that lead to ease of data collection and better CVD outcomes.

Comparisons With Prior Work

Although a high proportion of our survey participants reported use of internet or cellphone, the proportion observed is somewhat lower than national estimates. According to a 2021 survey conducted by the Pew Research Center [21], the vast majority of Americans (97%) own a cellphone and 85% own a smartphone, with no substantial differences among Black and White populations. However, among Americans who are 65 years and older, these percentages were lower (92% reported owning a cellphone and 61% owned a smartphone). Thus, one possible explanation for a lower proportion of our survey participants reporting use of internet and cell phone compared to national estimates may be due to the older age of our survey participants. Two other surveys conducted in 2019 by Pew Research Center [22,23] found that the digital divide persisted between rural and nonrural America, impacted by income levels [22,23]; therefore, the lower proportion of our survey participants reporting use of internet or cellphone may also reflect some differences in employment and income levels. Use of a smartphone may bridge this digital divide, as it offers a less expensive yet diverse alternative to the in-home Wi-Fi connection. It also provides more internet access to the undeserved communities compared to in-home Wi-Fi connection [24,25].

The digital health trend has the potential to transform health care [26-30]. The eHealth and mHealth technologies have come along as convenient and efficient ways to collect daily or frequent health data from individual persons to inform clinical care and research. For instance, several eHealth and mHealth studies using digital technologies had been piloted and tested in the FHS [31-36]. Nevertheless, the collection of mHealth information is relatively new and remains less tested in established population-based African American cohorts and clinical research. Some of the challenges for collecting health information remotely that need to be addressed include high attrition rates and lack of representativeness of the target populations [37]. Additionally, African Americans have not been well represented in these clinical and research studies, raising the question as to the feasibility of mHealth research in African American communities. Our survey findings among participants in the JHS suggest that it is feasible to conduct CVD research using mHealth platforms in an African American community in the southern part of the United States.

During the COVID-19 pandemic, in-person interviews in cohort studies was difficult to conduct, and that highlighted the need for more practical remote data collection methods. Although our survey was completed well before the pandemic had started (ie, February 2019), we were able to connect with a large group of respondents among which almost 90% were digitally connected. This experience suggests that JHS cohort participants can be reached digitally in the future when a face-to-face interview may not be possible. Use of cellphone and smartphone technology may increase the efficiency of research studies, may provide more frequent and 'real-world' data on participants,

XSL•FO
RenderX

and may allow participants greater choice and convenience regarding when and where to respond to research surveillance.

Conclusions

Our survey findings suggest that it is feasible to collect survey data using eHealth and mHealth technologies among middle-aged and older African Americans already enrolled in an ongoing research study in Jackson, Mississippi. Our findings also highlight the need for more efforts to reduce the age and education divide in access and use of internet and smartphones for tracking health and research in African American communities.

Acknowledgments

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I), and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I, and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS. The funding agency did not play a role in the design of the study; collection, analysis, and interpretation of data; and writing the manuscript.

EJB is supported in part by the National Institute of Aging (1R01AG076757, 1R01AG066914) and the NHLBI(5R01HL141434, R01HL092577).

The views expressed in this paper are those of the authors and do not necessarily represent the views of the NHLBI; the National Institutes of Health; or the US Department of Health and Human Services.

Conflicts of Interest

JM was a guest lecturer for Merck Laboratories unrelated to this work.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 24 KB - jmir_v24i11e37501_app1.docx]

Multimedia Appendix 2
Questionnaire Form.
[PDF File (Adobe PDF File), 173 KB - jmir_v24i11e37501_app2.pdf]

References

1. Khoury M, Lam T, Ioannidis J. Transforming epidemiology for 21st century medicine and public health. Cancer Epidemiol Biomarkers Prev 2013;22(4):508-516. [doi: 10.1158/1055-9965.EPI-13-0146]
2. Lauer MS. Time for a creative transformation of epidemiology in the United States. JAMA 2012 Nov 07;308(17):1804-1805. [doi: 10.1001/jama.2012.14838] [Medline: 23117782]
3. Rosbash M. A threat to medical innovation. Science 2011 Jul 08;333(6039):136-136. [doi: 10.1126/science.1210374] [Medline: 21737707]
4. Sorlie PD, Bild DE, Lauer MS. Cardiovascular epidemiology in a changing world--challenges to investigators and the National Heart, Lung, and Blood Institute. Am J Epidemiol 2012 Apr 01;175(7):597-601 [FREE Full text] [doi: 10.1093/aje/kws138] [Medline: 22415032]
5. WHO. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. URL: http://apps.who.int/iris/handle/10665/44607 [accessed 2022-05-20]
6. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
7. Vollmer DD, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. JMIR Mhealth Uhealth 2015;3(1):e31 [FREE Full text] [doi: 10.2196/mhealth.3861] [Medline: 25830810]
8. 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:49-56 [FREE Full text] [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
9. Ritholz MD, Henn O, Atakov Castillo A, Wolpert H, Edwards S, Fisher L, et al. Experiences of adults with type 1 diabetes using glucose sensor–based mobile technology for glycemic variability: qualitative study. JMIR Diabetes 2019 Jul 08;4(3):e14032 [FREE Full text] [doi: 10.2196/14032] [Medline: 31287065]
10. Fox CS, Hwang S, Nieto K, Valentino M, Mutalik K, Massaro JM, et al. Digital Connectedness in the Framingham Heart Study. JAHA 2016 Apr 03;5(4):e003193. [doi: 10.1161/jaha.116.003193]

XSL•FO
RenderX

11. Schrauben SJ, Appel L, Rivera E, Lora CM, Lash JP, Chen J, CRIC Study Investigators. Mobile health (mHealth) technology: assessment of availability, acceptability, and use in CKD. Am J Kidney Dis 2021 Jun;77(6):941-950.e1 [FREE Full text] [doi: 10.1053/j.ajkd.2020.10.013] [Medline: 33309860]
12. Sempos CT, Bild DE, Manolio TA. Overview of the Jackson Heart Study: a study of cardiovascular diseases in African American men and women. Am J Med Sci 1999 Mar;317(3):142-146. [doi: 10.1097/00000441-199903000-00002] [Medline: 10100686]
13. Taylor H. The Jackson Heart Study: an overview. Ethn Dis 2005;15(4 Suppl 6):S6-S1. [Medline: 16317981]
14. Carpenter MA, Crow R, Steffes M, Rock W, Heilbraun J, Evans G, et al. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. Am J Med Sci 2004 Sep;328(3):131-144. [doi: 10.1097/00000441-200409000-00001] [Medline: 15367870]
15. Min Y, Anugu P, Butler KR, Hartley TA, Mwasongwe S, Norwood AF, et al. Cardiovascular disease burden and socioeconomic correlates: findings from the Jackson Heart Study. JAHA 2017 Aug 02;6(8):e004416. [doi: 10.1161/jaha.116.004416]
16. American DA. Diagnosis and classification of diabetes mellitus. Diabetes Care 2010 Jan;33 Suppl 1:S62-S69 [FREE Full text] [doi: 10.2337/dc10-S062] [Medline: 20042775]
17. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, et al. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. Hypertension 2003 Dec;42(6):1206-1252 [FREE Full text] [doi: 10.1161/01.HYP.0000107251.49515.c2] [Medline: 14656957]
18. Callier S, Fullerton SM. Diversity and inclusion in unregulated mhealth research: addressing the risks. J Law Med Ethics 2020 Mar 01;48(1_suppl):115-121. [doi: 10.1177/1073110520917036] [Medline: 32342751]
19. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]
20. Wildenbos GA, Jaspers MWM, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: using an aging barriers framework to classify usability problems. Int J Med Inform 2019 Dec;124:68-77 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
21. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-08-06]
22. Center P. Digital gap between rural and nonrural America persists. Some digital divides persist between rural, urban and suburban America. 2019. URL: https://www.pewresearch.org/fact-tank/2019/05/31/digital-gap-between-rural-and-nonrural-america-persists/ [accessed 2021-06-08]
23. Vogels EA. Digital divide persists even as lower-income Americans make gains in tech adoption. Pew Research Center. 2021. URL: https://www.pewresearch.org/fact-tank/2019/05/07/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/ [accessed 2021-08-06]
24. Chapter one: a portrait of smartphone ownership. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/01/chapter-one-a-portrait-of-smartphone-ownership/ [accessed 2022-06-22]
25. Smith A, Zickhur K. Digital differences. Pew Research Center. 2012. URL: https://www.pewresearch.org/internet/2012/04/13/digital-differences/ [accessed 2022-06-08]
26. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
27. Powell AC, Landman AB, Bates DW. In search of a few good apps. JAMA 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
28. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
29. Topol EJ, Steinhubl SR, Torkamani A. Digital medical tools and sensors. JAMA 2015 Jan 27;313(4):353-354. [doi: 10.1001/jama.2014.17125] [Medline: 25626031]
30. Smith A. U.S. Smartphone use in 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2021-06-08]
31. McManus DD, Trinquart L, Benjamin EJ, Manders ES, Fusco K, Jung LS, et al. Design and preliminary findings from a new electronic cohort embedded in the framingham heart study. J Med Internet Res 2019 Mar 01;21(3):e12143 [FREE Full text] [doi: 10.2196/12143] [Medline: 30821691]
32. Lin H, Sardana M, Zhang Y, Liu C, Trinquart L, Benjamin EJ, et al. Association of habitual physical activity with cardiovascular disease risk. Circ Res 2020 Oct 23;127(10):1253-1260. [doi: 10.1161/circresaha.120.317578]
33. Pathiravasan CH, Zhang Y, Trinquart L, Benjamin EJ, Borrelli B, McManus DD, et al. Adherence of mobile app-based surveys and comparison with traditional surveys: ecohort study. J Med Internet Res 2021 Jan 20;23(1):e24773 [FREE Full text] [doi: 10.2196/24773] [Medline: 33470944]
34. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the be fit randomized clinical trial. JAMA Intern Med 2017 Nov 01;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]

XSL•FO
RenderX

35. Spartano NL, Lin H, Sun F, Lunetta KL, Trinquart L, Valentino M, et al. Comparison of on-site versus remote mobile device support in the framingham heart study using the health eheart study for digital follow-up: randomized pilot study set within an observational study design. JMIR Mhealth Uhealth 2019 Sep 30;7(9):e13238 [FREE Full text] [doi: 10.2196/13238] [Medline: 31573928]
36. Trinquart L, McManus D, Nowak C. Abstract 13084: increasing engagement in the electronic Framingham Heart Study (eFHS): a factorial randomized trial. Circulation 2019;140(Suppl_1):A13084 [FREE Full text]
37. 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 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]

Abbreviations

AOR: adjusted odds ratio
CVD: cardiovascular disease
FHS: Framingham Heart Study
JHS: Jackson Heart Study
mHealth: mobile health

Edited by A Mavragani; submitted 23.02.22; peer-reviewed by L Glover, C Diamantidis; comments to author 08.05.22; revised version received 26.07.22; accepted 18.10.22; published 21.11.22.

Please cite as:
Anugu P, Ansari MAY, Min YI, Benjamin EJ, Murabito J, Winters K, Turner E, Correa A
Digital Connectedness in the Jackson Heart Study: Cross-sectional Study
J Med Internet Res 2022;24(11):e37501
URL: https://www.jmir.org/2022/11/e37501
doi:10.2196/37501
PMID:36409531

©Pramod Anugu, Md Abu Yusuf Ansari, Yuan-I Min, Emelia J Benjamin, Joanne Murabito, Karen Winters, Erica Turner, Adolfo Correa. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Relative Validation of an Artificial Intelligence–Enhanced, Image-Assisted Mobile App for Dietary Assessment in Adults: Randomized Crossover Study

Audrey Moyen[1,2], BSc; Aviva Ilysse Rappaport[1], BSc, MSc; Chloé Fleurent-Grégoire[1,2], BSc; Anne-Julie Tessier[1,2], BSc, PhD; Anne-Sophie Brazeau[1], BSc, MSc, PhD; Stéphanie Chevalier[1,2,3], BSc, MSc, PhD

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

[2]Research Institute of the McGill University Health Centre, Montreal, QC, Canada

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

Corresponding Author:
Stéphanie Chevalier, BSc, MSc, PhD
School of Human Nutrition
McGill University
21111 Lakeshore Rd
Sainte-Anne-de-Bellevue, QC, H9X 3V9
Canada
Phone: 1 514 398 8603
Email: Stephanie.chevalier@mcgill.ca

Abstract

Background: Thorough dietary assessment is essential to obtain accurate food and nutrient intake data yet challenging because of the limitations of current methods. Image-based methods may decrease energy underreporting and increase the validity of self-reported dietary intake. Keenoa is an image-assisted food diary that integrates artificial intelligence food recognition. We hypothesized that Keenoa is as valid for dietary assessment as the automated self-administered 24-hour recall (ASA24)–Canada and better appreciated by users.

Objective: We aimed to evaluate the relative validity of Keenoa against a 24-hour validated web-based food recall platform (ASA24) in both healthy individuals and those living with diabetes. Secondary objectives were to compare the proportion of under- and overreporters between tools and to assess the user's appreciation of the tools.

Methods: We used a randomized crossover design, and participants completed 4 days of Keenoa food tracking and 4 days of ASA24 food recalls. The System Usability Scale was used to assess perceived ease of use. Differences in reported intakes were analyzed using 2-tailed paired *t* tests or Wilcoxon signed-rank test and deattenuated correlations by Spearman coefficient. Agreement and bias were determined using the Bland-Altman test. Weighted Cohen κ was used for cross-classification analysis. Energy underreporting was defined as a ratio of reported energy intake to estimated resting energy expenditure <0.9.

Results: A total of 136 participants were included (mean 46.1, SD 14.6 years; 49/136, 36% men; 31/136, 22.8% with diabetes). The average reported energy intakes (kcal/d) were 2171 (SD 553) in men with Keenoa and 2118 (SD 566) in men with ASA24 (P=.38) and, in women, 1804 (SD 404) with Keenoa and 1784 (SD 389) with ASA24 (P=.61). The overall mean difference (kcal/d) was −32 (95% CI −97 to 33), with limits of agreement of −789 to 725, indicating acceptable agreement between tools without bias. Mean reported macronutrient, calcium, potassium, and folate intakes did not significantly differ between tools. Reported fiber and iron intakes were higher, and sodium intake lower, with Keenoa than ASA24. Intakes in all macronutrients (r=0.48-0.73) and micronutrients analyzed (r=0.40-0.74) were correlated (all P<.05) between tools. Weighted Cohen κ scores ranged from 0.30 to 0.52 (all P<.001). The underreporting rate was 8.8% (12/136) with both tools. Mean System Usability Scale scores were higher for Keenoa than ASA24 (77/100, 77% vs 53/100, 53%; P<.001); 74.8% (101/135) of participants preferred Keenoa.

Conclusions: The Keenoa app showed moderate to strong relative validity against ASA24 for energy, macronutrient, and most micronutrient intakes analyzed in healthy adults and those with diabetes. Keenoa is a new, alternative tool that may facilitate the work of dietitians and nutrition researchers. The perceived ease of use may improve food-tracking adherence over longer periods.

(J Med Internet Res 2022;24(11):e40449) doi:10.2196/40449

XSL•FO
RenderX

KEYWORDS

dietary intake; dietary assessment; food diary; food records; automated self-administered 24-hour recall; ASA24; Keenoa

Introduction

Background

Associations among diet, health, and disease have been made, but findings are often criticized based on the unreliability of the collected dietary data. Indeed, current dietary assessment methods used in research are either burdensome to participants, resulting in a lower rate of compliance, or lack accuracy and precision [1]. They are also expensive because it is time-consuming for researchers to collect and analyze data. These methods include (1) written food diaries, which require participants to write every food and beverage consumed, usually during 3 to 7 consecutive days [2,3]; (2) 24-hour recalls where participants report all foods and beverages consumed for the previous day in the presence of a trained interviewer; and (3) food frequency questionnaires, which ask the frequency of consumption of specific food items or categories of food consumed during a defined time period [4]. Food recalls rely on memory for both the food items consumed and their quantity [4], which may impact the validity of collected data. One of the most common methods for dietary assessment is the food diary that is often incomplete and therefore requires researchers to make assumptions to evaluate dietary intake. Food diaries are more prone to reactivity bias; that is, when participants change their food intake owing to the awareness of their intake being analyzed [4]. Lack of participant motivation may also be a source of error, limiting the number of valid tracking days [4,5]. Furthermore, research personnel must be trained on standardized methods to administer, review, and analyze food diaries. This is of particular importance in the context of multisite studies. In addition, these methods require participants to estimate their food portion sizes accurately and are demanding on research personnel who must enter collected dietary data into computer software for data analysis. Finally, underestimation of total calorie intake is frequent, in particular among individuals living with overweight and obesity [3,6,7].

New tools have been developed using web and mobile technologies [8-11]. It is now possible to collect food diaries and recalls on the web, to facilitate the recording process and data analysis [9]. Only a few image-assisted food record methods exist, and few studies have investigated their use. Evidence points to potential improvement of accuracy with such methods compared with traditional self-reported dietary assessment approaches by providing through pictures of meals, food items, and additional details that could be otherwise omitted [12-14]. The use of pictures may decrease underreporting and increase the accuracy of portion size estimation. Keenoa is a newly designed, image-assisted food-tracking mobile app that integrates artificial intelligence for food recognition. It was developed to introduce user-friendly technologies facilitating the work of researchers and dietitians. Given the lower burden on participants, it has the potential to overcome some of the limitations of traditional methods.

Objectives

This study was conducted to evaluate the relative validity of Keenoa against a 24-hour validated web-based food recall platform automated self-administered 24-hour recall (ASA24) in both healthy individuals and those living with type 1 and type 2 diabetes. People with diabetes, namely type 1, are generally more aware of their food intake to match with insulin injections, hence we aimed to validate the tool in both a healthy and diabetic population. Secondary objectives were to compare the proportion of under and overreporters between tools and to assess the user's appreciation of the tools. We hypothesized that dietary assessment using Keenoa is comparable with that of ASA24-Canada and better appreciated by users.

Methods

Study Population

Participants were recruited between February and November 2021 through social media, email, and word of mouth. Participants with diabetes were recruited through the Behaviors, Therapies, Technologies, and Hypoglycemic Risk in Type 1 Diabetes (BETTER) registry [15] and through emails from a diabetes association (ie, Diabète Québec). Inclusion criteria were adults aged 18 to 70 years; owning a smartphone, computer, or tablet; having access to the internet; being able to read English or French; and having a self-reported BMI between 18 and 35 kg/m^2 to minimize under- and overreporting [16]. Exclusion criteria were living outside of Canada, history of frequent dieting, following a weight loss diet, having a disordered eating pattern, having any active and uncontrolled acute or chronic disease, having gained or lost a significant amount of weight (>5 kg) within the past 3 months, and being pregnant or breastfeeding. Participants with diabetes were excluded if their diagnosis dated ≤1 year, if their diabetes medication was changed within the past 3 months or if they had celiac disease. Electronic informed consent was obtained from all participants.

Ethics Approval

The study was approved by the McGill Research Ethics Board (REB 20-09-035) and registered on the Dietary Assessment Calibration or Validation Register from the National Cancer Institute.

Study Design

We used a randomized crossover design for a 2-week duration, and participants completed a 4-consecutive-day tracking period (3 weekdays and 1 weekend day) with both dietary assessment methods, ASA24 and Keenoa, in a random order, on the same days of the week. The study was performed entirely on the web. Participants were also asked to answer a web-based questionnaire about their sex, gender, height, weight, years of education, occupation, chronic diseases, weight history, body image satisfaction, medications, and vitamin and mineral supplements (Multimedia Appendix 1). The questionnaire also included questions on physical activity levels following the

XSL•FO
RenderX

International Physical Activity Questionnaire–Short Form [17]. Participants were given written instructions (1 page each) on the use of each tool, and assistance was available as needed by email or phone to ensure accurate data entry. Upon completion of the questionnaire, they were assigned a set of 4 consecutive days to track their food intake. Those 4 days were determined to begin on the Wednesday or Sunday following questionnaire completion to allow for the 4 consecutive days to include 3 weekdays. If patients omitted to enter their daily food intake, reminders were sent at the end of the day, before dinner time. With ASA24, participants could still log in until midnight to fill their recall. If the recall had not been completed by the next day, participants were allotted another day of tracking (weekday or weekend day, depending on the missing day). With Keenoa, missing days were also replaced with additional tracking days. Each participant was allotted a maximum of 2 reminders per tool, after which they were excluded from the study. An encouragement message was sent after the completion of the first day with each tool to maximize adherence. Finally, participants were asked to complete the System Usability Scale (SUS) [18], a validated usability questionnaire that consists of 10 questions with 5 options each, from strongly disagree to strongly agree. The SUS asked participants about both tools in the same order in which they tracked their intake (ie, if participants used Keenoa first and ASA24 second, the tool asked participants to answer questions about Keenoa followed by the same set of questions on ASA24). After completing the tracking period with both dietary assessment tools, if food items present in the pictures of participants were omitted or entered incorrectly, entries from the Keenoa app were adjusted by a dietitian. For instance, when condiments such as ketchup or mayonnaise were present in a picture but not in the participants' entries, the item was added to the food diary.

Automated Self-administered 24-Hour Recall

The ASA24 is a web-based dietary recall or food diary tool developed by the National Cancer Institute of the National Health Institute designed for epidemiological and clinical research purposes and is free to use. The ASA24 recall tool has been extensively validated notably against true intake [19] and recovery biomarkers [20], hence we used the recall tool as a comparator. The recall tool was designed to mimic a 24-hour recall without requiring an interviewer. As such, it uses the multiple-pass method which includes multiple reminders on frequently forgotten foods or ingredients and asks for details regarding portion sizes using pictures as models. Once completed, researchers access the data analysis, which is based on the Canadian Nutrient File (CNF; version 2015) for the ASA24-Canada 2018. The ASA24 tool allows to enter supplements, with an option to include them in the analytic files. In this study, supplements were not included in the nutrient comparisons.

Keenoa

Keenoa (version 1.0.3) is a newly designed, intelligent food-tracking mobile app (participant's end) linked to a web platform (researcher's end). With the mobile app, participants take pictures of their meals and snacks, which are recognized or partly recognized by an artificial intelligence–based algorithm. Prompted by a few questions, users specify the foods and beverages consumed and estimate portion sizes using dynamic pictograms. From the web app, the researcher has access to the meal pictures and their associated detailed nutritional analysis, in real time. Dietary data are obtained from the CNF database [21], the US Department of Agriculture database [22], and frequently imported food items from the South Korean Food Composition Database [23], the Hong Kong Nutrient Information Inquiry System [24], the Indian Food Composition Tables [25], and the Australian Food Composition Database [26]. Bar codes from grocery store items can be scanned from the app. Relevant information such as time of consumption and time between meals is also available. The Keenoa app has been validated in the past [27,28]. However, those were performed using a prototype of the app. The validity and reliability of the updated Keenoa food diary remain to be established for both research and evidence-based dietetic practice.

Sample Size

Considering that correlations between reference and test methods are expected to range between 0.5 and 0.7 in the context of validity, based on a 0.4 null-hypothesis correlation coefficient, 0.6 effect size (alternative hypothesis), Cronbach α=.05, and 80% power, the sample size required was 111 participants. This sample size for dietary intake validation studies is supported by Serra-Majem et al [29] and Willett [4]. Thus, we aimed to recruit 120 healthy participants within 3 age strata of 18 to 35, 36 to 54, and 55 to 70 years with an equal gender ratio of 20 (50%) men and 20 (50%) women per stratum. In addition, we aimed to recruit 40 patients (20 men and 20 women, 18-70 years) with diagnosed type 1 diabetes and 40 patients (20 men and 20 women, 18-70 years) with diagnosed type 2 diabetes for this study.

Statistical Analysis

Energy and macronutrient intakes were averaged to reflect habitual intake. Mean results from Keenoa were compared with those from ASA24 using paired sample 2-tailed *t* tests and Wilcoxon rank tests, by gender, with and without adjustment for caloric intake as per the residuals method [30]. The normality of data distributions was evaluated using the Shapiro-Wilk test. Correlations were evaluated using Spearman correlation coefficients and were deattenuated, as suggested [31]. Deattenuation involves adjusting for reliability, by dividing a correlation by the product of the 2 tools' daily intraindividual variability [32] from intraclass correlations. Weighted Cohen κ values were computed to assess interrater agreement between tools and were interpreted as per Landis and Koch [33]. The Bland-Altman method was used to assess the mean difference and agreement between the 2 tools. We established our acceptable limits of agreement (LOAs) as the mean intraindividual SD between reported daily intakes multiplied by 1.96, given that LOAs in similar studies are often large and no acceptable LOAs have been established in the literature for nutrition validation studies. Pearson correlations were used to assess associations between differences and average intake with Bland-Altman plots, as suggested by Lombard et al [31]. Energy underreporting was defined as a ratio between mean 4-day

XSL•FO
RenderX

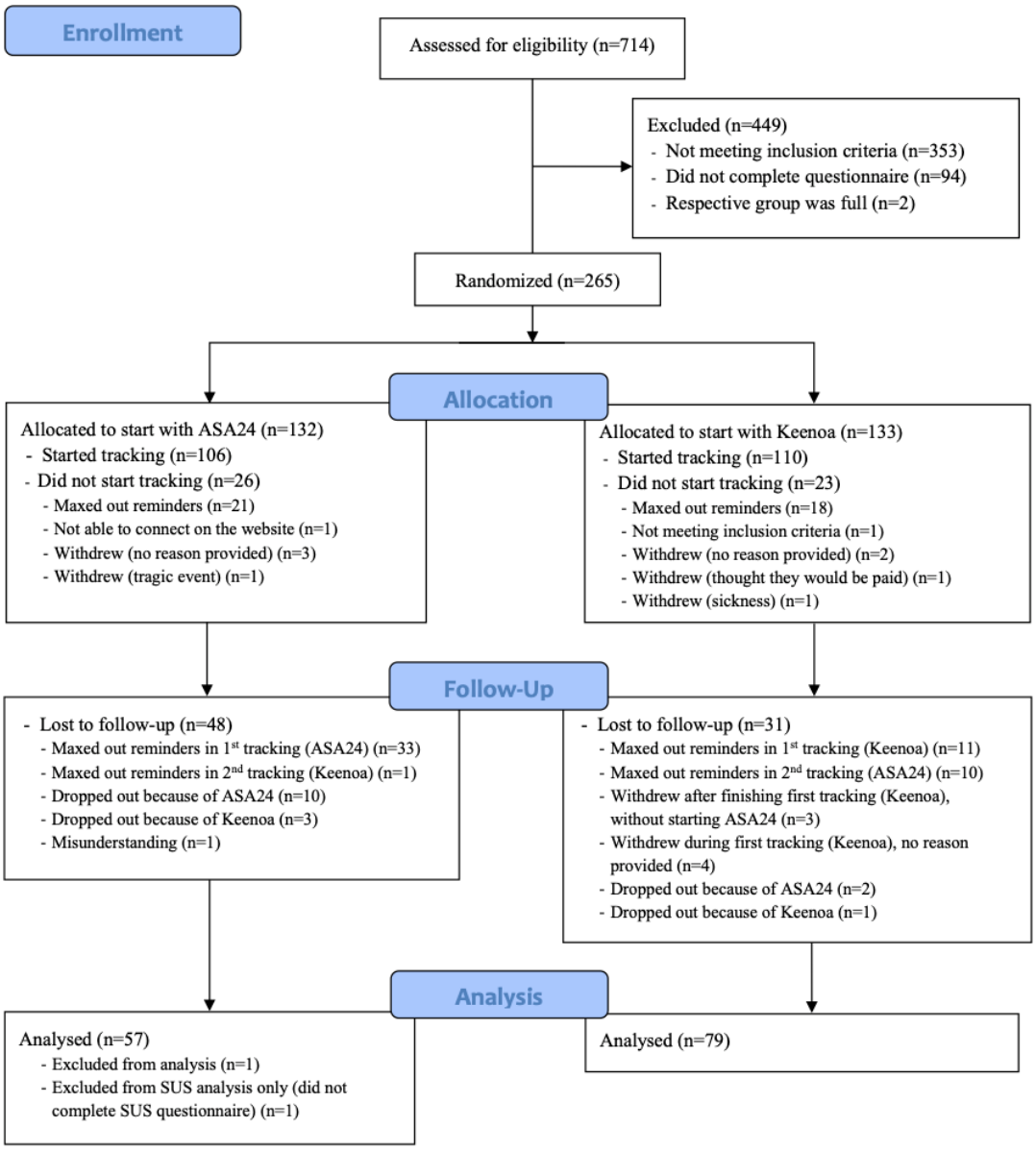
energy intake and estimated resting energy expenditure ratio below 0.9 [34] and overreporting as a ratio >2.4, as previously established [35]. Resting energy expenditure was estimated using the Mifflin St-Jeor equation [36]. The significance of differences in under- and overreporting was measured using χ^2 tests. All analyses were performed using SPSS (version 28; IBM Corp). *P* values <.05 were considered to be statistically significant.

Results

Participant Characteristics

A total of 361 respondents were eligible to participate in the study. Of the 361 participants, 92 (25.5%) participants did not fill the first questionnaire, 2 (0.5%) were screened when their respective stratum was full and were therefore excluded, 128 (35.4%) dropped out before completing both dietary tracking periods, and 1 (0.3%) participant was excluded owing to outstanding differences between both tracking periods (2.25-fold difference in energy intake). Completion rates and dropouts varied by tool with more dropouts and exclusions during ASA24 tracking (55/128, 43%) as opposed to Keenoa tracking (16/128, 12.5%; Figure 1). In total, 136 participants were included in the analyses; 11% (15/136) of participants were trained in nutrition or dietetics. Baseline participant characteristics are summarized in Table 1, by gender. The mean age of participants was 45.4 (SD 14.5) years in women and 47.2 (SD 15.2) years in men. The mean BMI was 24.0 (SD 3.9) kg/m^2 in women and 26.8 (SD 4.4) kg/m^2 in men. Most participants were of White ethnicity (113/136, 83.1%) and had a bachelor's degree (101/136, 74.3%).

Figure 1. Participant flow. ASA24: automated self-administered 24-hour recall; SUS: System Usability Scale.



XSL•FO
RenderX

Table 1. Characteristics of study participants (N=136)[a].

Variable	Women (n=87)	Men (n=49)
Age (years), mean (SD)	45.4 (14.5)	47.2 (15.2)
BMI (kg/m^2), mean (SD)	24.0 (3.9)	26.8 (4.4)
Diabetes, n (%)		
Yes	22 (25.3)	9 (18.4)
No	65 (74.7)	40 (81.6)
Chronic disease, n (%)		
Yes	30 (34.5)	12 (24.5)
No	56 (64.4)	37 (75.5)
Level of education, n (%)		
Some high school completed	1 (1.1)	1 (2)
High school diploma	2 (2.3)	2 (4.1)
Vocational school	1 (1.1)	4 (8.2)
Completed college	18 (20.7)	3 (6.1)
Bachelor's degree	35 (40.2)	18 (36.7)
Graduate degree (ie, MSc, MA, or PhD)	26 (29.9)	18 (36.7)
Professional degree (ie, MD)	1 (1.1)	3 (6.1)
Ethnicity, n (%)		
White	74 (85.1)	39 (79.6)
Aboriginal (first nations)	1 (1.1)	1 (2)
Asian	7 (8)	5 (10.2)
Black	0 (0)	2 (4.1)
Hispanic	3 (3.4)	2 (4.1)
Multiethnic	1 (1.1)	0 (0)
Body weight satisfaction , n (%)		
Yes	41 (47.1)	33 (67.3)
No	38 (43.7)	15 (30.6)
I do not know or I prefer not to answer	8 (9.1)	1 (2)
Medication , n (%)		
Yes	52 (59.8)	25 (51)
No	35 (40.2)	24 (49)
Taking vitamin or mineral supplement , n (%)		
Yes	28 (32.2)	20 (40.8)
No	59 (67.8)	29 (59.2)
Special diets , n (%)		
None	67 (77)	41 (83.7)
Vegetarian, vegan	8 (9.2)	3 (6.1)
Intermittent fasting	1 (1.1)	0 (0)
Gluten free	1 (1.1)	0 (0)
Other (eg, Mediterranean or low carb)	10 (21.7)	5 (10.2)
Cooking responsibility , n (%)		
Myself	78 (89.7)	32 (65.3)

XSL•FO
RenderX

Variable	Women (n=87)	Men (n=49)
Another family or household member	7 (8)	17 (34.7)
Grocery shopping responsibility , n (%)		
Myself	76 (87.4)	35 (71.4)
Another family or household member	9 (10.3)	14 (28.6)
Self-reported physical activity level , n (%)		
Sedentary	16 (18.4)	7 (14.3)
Low active	34 (39.1)	20 (40.8)
Active	35 (40.2)	21 (42.9)
Very active	2 (2.3)	1 (2)
Used diet tracking app before , n (%)		
Yes	32 (36.8)	17 (34.7)
No	55 (63.2)	32 (65.3)

[a]All missing data not adding up to the total sample population are "I don't know" or "I prefer not to answer."

Mean Differences and Agreement Between Tools

Mean reported intakes of energy and all selected nutrients are reported by the tool, presented in Table 2 for women and Table 3 for men. Mean reported energy intakes (kcal/d) with Keenoa and ASA24 were 2171 (SD 553) and 2118 (SD 566; *P*=.38) in men, and 1804 (SD 404) and 1784 (SD 389; *P*=.61) in women, respectively. There were no statistically significant differences between tools in macronutrient intake (all *P*>.05) except for fibers, which was higher with Keenoa in both women (*P*<.001) and men (*P*=.02). Differences in macronutrient distribution were also nonsignificant in both genders. From the micronutrients analyzed, reported potassium and folate intakes did not differ between tools. Reported fiber intakes were higher with Keenoa in both women (*P*=.002) and men (*P*=.05) in both genders. Mean reported calcium intake was not different between tools in women (*P*=.67) but men had significantly lower calcium with ASA24 with a mean difference of −90 mg/d (95% CI −175 to −5; *P*=.04). Reported sodium intake was higher with ASA24 in both genders (*P*<.001). Differences or the absence of differences remains when adjusting nutrients for energy intake (data not shown).

Results from Bland-Altman analyses are shown in Figure 2 and Table 4. Pooling participants of both genders resulted in mean differences of −32 kcal (LOAs: −789.2 to 725.2) for energy, −7.9 g (LOAs: −104.8 to 89.0) for carbohydrate, −1.9 g (LOAs: −45.0 to 41.3) for protein, and 0.0 g (LOAs: −44.8 to 44.9) for fat intakes, as shown in Bland-Altman plots (Figure 2); negative values indicated that reporting was higher with Keenoa compared with ASA24. Bland-Altman plots for energy and macronutrients by gender are presented in Figure 2. LOAs of all nutrients (Table 4) were within acceptable LOAs except for calcium, sodium, and folate, which had larger LOAs for the difference between tools as opposed to the intraindividual daily variability LOAs.

Table 2. Reported nutrient intakes and mean differences between automated self-administered 24-hour recall (ASA24) and Keenoa in women (n=87)[a].

	ASA24, mean (SD)	Keenoa, mean (SD)	Mean difference (95% CI)	*P* value
Energy (kcal)	1784 (389)	1804 (404)	−20 (−99 to 58)	.61
Carbohydrates (g)	195.3 (55.8)	202.0 (63.5)	−6.6 (−16.5 to 3.2)	.19
Protein (g)	76.9 (17.6)	79.0 (19.7)	−2.1 (−6.8 to 2.5)	.37
Fat (g)	76.1 (22.6)	74.8 (21.6)	1.2 (−3.6 to 6.1)	.61
Fiber (g)	19.8 (7.6)	22.2 (9.0)	−2.4 (−3.7 to −1.1)	<.001
Calcium (mg)	860 (309)	875 (384)	−15 (−85 to 55)	.67
Iron (mg)	12.0 (3.2)	13.6 (5.1)	−1.6 (−2.5 to −0.6)	.002[b]
Sodium (mg)	2950 (739)	2442 (799)	508 (302 to 715)	<.001[b]
Potassium (mg)	2852 (755)	2929 (896)	−77 (−202 to 48)	.23
Folate (µg)	434 (129)	406 (135)	28 (−0.4 to 56)	.05[b]

[a]Paired 2-tailed *t* tests unless otherwise indicated.

[b]Wilcoxon signed-rank test.

XSL•FO
RenderX

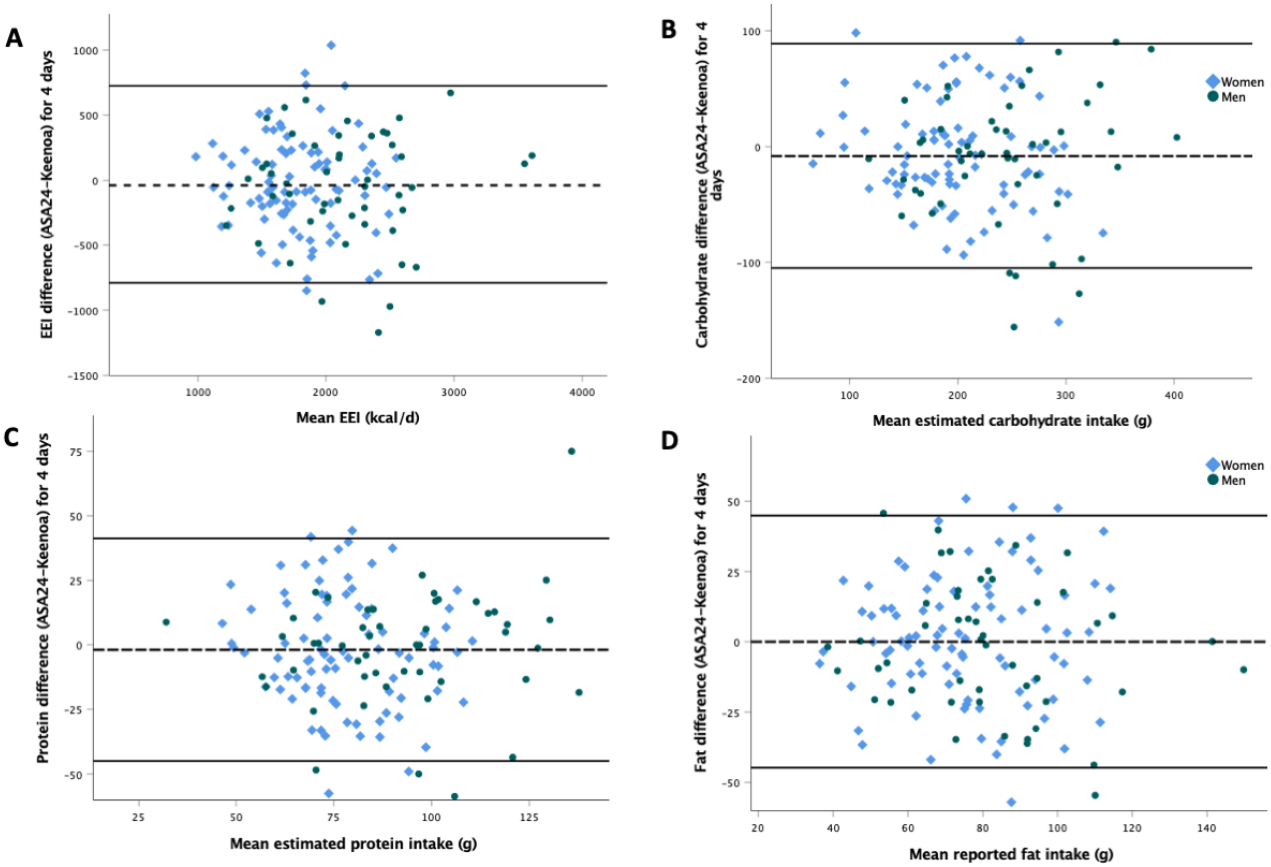
Table 3. Reported nutrient intakes and mean differences between automated self-administered 24-hour recall (ASA24) and Keenoa in men (n=49)[a].

	ASA24, mean (SD)	Keenoa, mean (SD)	Mean difference (95% CI)	*P* value
Energy (kcal)	2118 (566)	2171 (553)	−53 (−173 to 68)	.38
Carbohydrates (g)	239.1 (74.2)	249.3 (67.1)	−10.2 (26.0 to 5.6)	.20
Protein (g)	92.0 (27.3)	93.3 (24.0)	−1.4 (−7.8 to 5.1)	.67
Fat (g)	80.6 (24.1)	82.8 (27.5)	−2.1 (−8.8 to 4.5)	.52
Fiber (g)	21.7 (10.2)	24.3 (8.7)	−2.5 (−4.6 to −0.5)	.02
Calcium (mg)	874 (323)	964 (351)	−90 (−175 to −5)	.04
Iron (mg)	13.8 (4.2)	14.8 (4.7)	−1.0 (−2.0 to −0.0)	.05
Sodium (mg)	3534 (1146)	2880 (930)	654 (383 to 924)	<.001[b]
Potassium (mg)	3188 (1111)	3323 (1206)	−136 (−492 to 220)	.26[b]
Folate (µg)	475 (195)	511 (237)	−37 (−99 to 26)	.10

[a]Paired 2-tailed *t* tests unless otherwise indicated.

[b]Wilcoxon signed-rank test.

Figure 2. Bland-Altman plots. ASA24: automated self-administered 24-hour recall; EEI: estimated energy intake.



XSL•FO
RenderX

Table 4. Bland-Altman agreement analysis (N=136).

	Mean difference (SD)[a]	LOAs[b]	Acceptable LOAs[c]	Pearson *r*[d]	*P* value[d]
Energy (kcal)	−32.0 (386.3)	−789.2 to 725.2	−849.1 to 785.1	−0.022	.80
Carbohydrates (g)	−7.9 (49.4)	−104.8 to 89.0	−110.1 to 94.3	−0.048	.58
Protein (g)	−1.9 (22.0)	−45.0 to 41.3	−46.7 to 42.9	0.017	.84
Fat (g)	0.0 (22.9)	−44.8 to 44.9	−46.2 to 46.2	−0.047	.59
Fiber (g)	−2.4 (6.5)	−15.2 to 10.3	−16.6 to 11.8	−0.045	.60
Calcium (mg)	−42.1 (317.4)	−664.2 to 580.0	−588.4 to 504.2	−0.214	.01
Iron (mg)	−1.4 (4.1)	−9.5 to 6.7	−9.5 to 6.7	−0.341	<.001
Sodium (mg)	560.7 (959.0)	−1315.9 to 2440.4	−1126.6 to 2248.0	0.091	.29
Potassium (mg)	−97.9 (875.9)	−1814.8 to 1618.9	−1884.8 to 1689.0	−0.155	.07
Folate (µg)	4.7 (170.8)	−330.0 to 339.5	−321.2 to 330.6	−0.233	.006

[a]Automated self-administered 24-hour recall (ASA24)—Keenoa.

[b]LOA: limit of agreement.

[c]Established as the mean intraindividual SD between reported daily intakes multiplied by 1.96.

[d]Refers to the Pearson correlation between the mean and mean difference between the tools. Significant *P* values indicate systematic bias.

For energy and macronutrient intakes, deattenuated correlation coefficients ranged from 0.69 to 1.0 in men and from 0.42 to 0.84 in women (Multimedia Appendix 1). Micronutrients' deattenuated correlation coefficients ranged from 0.61 to 1.0 in men and 0.31 to 1.0 in women. Sodium was the nutrient with the lowest correlation coefficient value in women.

Cross-Classification Analysis

Results from the cross-classification analysis are shown in Table 5. The weighted κ score associated with energy reporting with both tools was 0.45 (95% CI 0.34-0.56). Weighted κ scores ranged from 0.29 to 0.52 for macronutrients and from 0.31 to 0.51 for the micronutrients analyzed (all *P*<.001).

Table 5. Cross-classification agreement by quartiles.

	Cohen κ[a] (95% CI)	*P* value[b]
Energy (kcal)	0.45 (0.34-0.56)	<.001
Carbohydrates (g)	0.52 (0.42-0.62)	<.001
Protein (g)	0.29 (0.17-0.42)	<.001
Fat (g)	0.32 (0.20-0.43)	<.001
Fiber (g)	0.51 (0.41-0.61)	<.001
Calcium (mg)	0.38 (0.27-0.49)	<.001
Iron (mg)	0.38 (0.26-0.49)	<.001
Sodium (mg)	0.31 (0.19-0.43)	<.001
Potassium (mg)	0.49 (0.39-0.60)	<.001
Folate (µg)	0.37 (0.25-0.48)	<.001

[a]Weighted Cohen κ.

[b]Significant *P* values indicate agreement.

Usability

Usability measured by the SUS indicated a mean score of 52.9 (SD 21.3) points out of 100 for ASA24 and 77.0 (SD 16.2) points for Keenoa (*P*<.001). Using the star rating (out of 5) after the completion of both tracking periods, of the 136 participants, 101 (74.8%) participants rated Keenoa higher compared with ASA24, whereas 11 (8.1%) gave a higher rating to ASA24. Among the 136 participants included in the analysis, 33 (24.3%) needed 2 reminders before they completed ASA24, while 11 (8.1%) needed 2 reminders with Keenoa. More reminders were needed during the first tracking with either tool. Finally, the frequency of energy underreporting was the same for each tool (12/136, 8.8%), and no overreporting was observed.

Discussion

Principal Findings

This study assessed the relative validity of the Keenoa mobile app against the validated ASA24 web-based recall method. No significant differences in estimated energy intake were found between tools that had similar rates of energy underreporting.

XSL•FO
RenderX

Mean carbohydrate, protein, fat, potassium, and folate intakes were also similar between the 2 tools, whereas fiber and iron reporting was higher and sodium was lower with Keenoa than ASA24, in both genders.

Interpretation of Findings

Bland-Altman analyses and paired 2-tailed *t* tests showed no systematic bias and agreement for all macronutrients and energy, with LOA falling within acceptable values, and without significant trends in the differences per intake. Weighted Cohen κ scores, which represent the ability of the tools to classify participants similarly based on their nutrient intake, indicated fair (fat and protein) to moderate (energy and carbohydrates) agreement at the individual level.

In addition to the main macronutrients and energy, fiber, calcium, iron, sodium, potassium, and folate were selected for analysis because they typically show little variability in day-to-day consumption [37,38] while generally of interest to clinicians and researchers. No systematic bias was detected for potassium and folate and calcium at a group level. Fiber and iron were consistently higher with Keenoa compared with ASA24, although the magnitude of the difference was relatively small: 2.4 g and 1.4 mg, respectively, both 11% of the mean estimated intake. On the other hand, reported sodium intake was consistently higher with ASA24, with a larger difference of 561 mg (19% of mean estimated intake). Fiber, iron, and potassium had LOAs lower than intraindividual daily variation LOAs suggesting an acceptable level of uncertainty between the tools. LOA for calcium, sodium, and folate were wider than the acceptable LOAs, which could indicate inadequate agreement. Bland-Altman analyses showed that, for most nutrients analyzed, the difference in reporting between the tools was not affected by the mean value of the intake. However, in addition to systematic bias, a significant trend was detected in iron reporting, where higher overall intakes correlated with higher Keenoa reporting. A slightly similar trend was also seen with folate and calcium, without systematic bias at the group level. Weighted Cohen κ scores confirmed fair (calcium, iron, sodium, and folate) to moderate (fiber and potassium) agreement at an individual level.

The difference between tools in estimating sodium intake may be explained by a combination of possible sodium overreporting with ASA24 and underreporting with Keenoa. First, neither ASA24 nor Keenoa asks participants about added salt. Although Keenoa does provide low- and no-sodium options for food items when applicable, ASA24 does not provide questions for salt adjustment and assumes that regular-sodium foods are consumed and that salt is added during food preparation [39]. Previous dietary intake surveys used to query on food items' salt content and the addition of salt at the table. However, in a 2012 report by the United States Department of Agriculture, taking out those salt adjustment questions was shown to lower sodium intake estimates by about 4%. Thus, those questions were not included in the development of ASA24 [39]. ASA24 was found to overreport sodium intake in 1 study [19] but not in others [20,40]. Alternatively, the absence of questions related to added salt within Keenoa might have led to an underestimation of sodium intake. These differences may have exaggerated the difference in sodium estimation between the 2 tools.

Regarding fiber intake reporting, higher values with Keenoa seem to be associated with the inclusion of mixed meals in the food diaries of our participants. The Keenoa database includes meal builders which allow participants to select different items used to prepare their mixed meal. Each item is linked to their unique CNF code in a proportion that is representative of popular recipes for this specific meal in North America. For example, chili can be logged by selecting the specific protein items, vegetables, and toppings that constituted the meal and that were really consumed. In contrast, ASA24 uses the predefined mixed dishes included in the CNF database, which are often fast-food versions, containing less vegetables than homemade versions, or versions in which the meal content is standard and cannot be modified. Finally, the difference in iron reporting between the 2 tools was higher in frequent consumers of breakfast cereals, for which bar codes could be used with Keenoa to specify the brand and type of cereals consumed. For instance, entering "cereals" in ASA24 yields "cereal (cold, other kind)" and "cereal (unknown kind)" as first options, leading participants to be less specific about cereals consumed, whereas Keenoa allows these items to be easily entered while not providing the possibility for vague entries.

In our study, no gender difference was seen in the reporting agreement between tools, except for calcium intake, which was significantly different in men. The difference is relatively small (90 mg; 10% of mean estimated intake). This difference is not because of outlier data and remains unexplained. Upon inspection of data distribution, no significant outliers were detected, and differences are well distributed around the mean. The difference was not statistically significant when grouping all participants. In addition, when performing the analyses by age group, the same findings were obtained.

Current findings of overall agreement between tools differ from those of a previous relative validation study published in 2020 [27] in which Keenoa showed significantly lower reporting for energy and most nutrients when compared with a 3-day pen-and-paper food diary analyzed with the Food Processor software (version 11.1; ESHA Research), using the CNF food database. Contrasting findings between that study [27] and this study may be because of the use of a different reference method (3-day written food diary vs ASA24) but most likely related to upgrades made to the Keenoa app. The study by Ji et al [27] was conducted using a prototype (version 0.3.7) of the Keenoa app, which was restricted to a search for food items available within the CNF. This database is limited, most notably in its cultural food content, and thus limits participants from selecting some of their consumed food items. The version of Keenoa used in this study (version 1.0.3) had a larger database including cultural foods from other national databases and had integrated meal builders which potentially facilitated participants' data entry. In addition, a multiple-pass function had also been implemented within Keenoa to maximize validity, prompting participants to review their diary at the end of the day and enter any omitted item. However, a conclusion as to which tool performs better cannot be verified as none of the 3-day food

XSL•FO
RenderX

diary or ASA24 reference methods is a gold standard to estimate dietary intake.

Other mobile apps are available to estimate nutritional intake. Most of them are not adapted for research purposes, as they allow the public to freely add food items to the database [41] and thus databases are not validated [42]. Some are aimed to induce weight loss and are not adapted for dietary assessment as the energy and nutrient intakes are inevitably shown to the participant, possibly leading to changes in behavior and intake [41-46]. Some mobile apps used for research have not been adapted to the Canadian food market [47-49]. Image-assisted mobile dietary assessment methods also exist but some are not connected to a nutrient composition database and thus require researchers to review all entries [50], providing limited advantage compared with traditional methods. Others are not available to the public [51]. In addition, other relative validation studies on mobile tools are often performed on the same day of intake (food diary followed by a 24-hour recall the next day), which induces a major training bias. Results from those studies typically show limited systematic bias with wide LOAs [52], which is comparable with our study's results even though our tracking periods were different between tools. Finally, some image-assisted methods involve wearable camera devices [53,54], which can also be time-consuming for researchers to collect and analyze data, in addition to potential ethical concerns and issues with participant acceptance because of their invasiveness.

Other web-based tools also exist to estimate nutritional intake in a recall format [20,55-58]. Most of them have been validated against recovery biomarkers and are currently used in research settings. Although most are adapted to European populations, 2 bilingual tools were available in Canada [20,57]. We selected ASA24 as a comparator to Keenoa because, in addition to being widely used in research, it has been extensively validated, notably against true intake [19], doubly labeled water, and urinary biomarkers [20]. However, ASA24 does have some limitations that might have limited adherence. First, it takes considerable time to complete the food recalls, between 41 and 58 minutes [40], given the automated multiple-pass method used to maximize validity. Second, because of the recall format, some participants were worried about forgetting their intake from the previous day and wrote down everything that was consumed in real time to properly enter it in ASA24 the next day, thus increasing the participant burden. The lower perceived usability with this tool led to the lower adherence that was detected with ASA24 in this study. This may lead to attrition bias where only the most motivated participants will report their intake or remain in longitudinal studies, thus limiting generalizability. It is noteworthy that participants who were assigned to start with ASA24 were more likely to drop out or be excluded from the study follow-up, consistent with the lower usability score from ASA24.

On the other hand, some participants preferred tracking once per day with ASA24 instead of continuously throughout the day with Keenoa. An advantage to ASA24 is the ability for prescription supplements to be entered and added to the nutritional analysis. The Keenoa app also has some limitations. For instance, it does not query participants on added salt, which might lead to sodium underreporting. The current Keenoa database does not include prescription vitamin and mineral supplements—those can only be added by hand by researchers or clinicians. However, the presence of a more extensive database, as well as the ability to copy previous meals, and the shorter time associated with tracking were preferred by participants, leading to a higher usability score. It is known that tracking periods over 4 days decreases the validity of filled diaries [5,59], likely secondary to a decrease in participant motivation. The perceived ease of use with the Keenoa app may increase food-tracking adherence over longer periods, thus leading to better estimations of one's usual intake. This remains to be shown.

Limitations

Our study has its limitations. Ideally, both measurement instruments should be used on the same days of food recording, but this approach would introduce training bias, as the use of one tool would influence the information provided with the other tool. Thus, we tested each instrument on 4 consecutive days, randomly, assuming a fairly constant intake of nutrients of interest. However, this approach increased intraindividual variability related to food intake, independently of the tracking tool. Furthermore, our sample was more educated and included a higher proportion of participants who self-identified as White than Canadian averages, which might limit the generalizability of our results. In addition, 11% (15/136) of our sample was trained in nutrition or dietetics and 19.9% (27/136) had type 1 diabetes, which may have improved the quality of food tracking. However, comparing a food recall to a food diary could have been leading to some differences in results between tools. We also aimed to recruit more participants, specifically men, and participants with diabetes. Despite numerous efforts to complete our group quotas, we experienced a plateau in recruitment. As the group having diabetes was inadequately powered to detect significant correlations and the results from this group did not differ from those of the healthy group, we pooled both groups together to ensure an adequate sample size. Finally, although extensively validated, ASA24 is not a gold standard measure. It is thus not possible to identify which tool provided results that were most representative of true intake. This supports the continuation of the validation process into a validation with recovery biomarkers or true intake measures.

Conclusions

The Keenoa app showed moderate to strong relative validity against ASA24 for energy, macronutrient, and most micronutrient intakes analyzed in healthy adults and those living with diabetes. Keenoa is a new, alternative tool that may facilitate the work of dietitians and researchers in nutrition. The perceived ease of use may improve food-tracking adherence over longer periods and minimize attrition bias.

XSL•FO
RenderX

Acknowledgments

SC, ASB, AJT, AIR, and AM developed the study design. SC, ASB, AIR, AM, and CFG contributed to participant recruitment. AM and AIR conducted the study and data collection. AM conducted all data analysis and drafted the manuscript. SC obtained ethics approval and funding and supervised all activities. All authors have read, edited, and approved the final manuscript.

Conflicts of Interest

AJT participated in the development of the Keenoa app and is the cofounder. The author did not receive any financial compensation for this study.

Multimedia Appendix 1
Baseline questionnaire and deattenuated correlation coefficients.
[DOC File , 78 KB - jmir_v24i11e40449_app1.doc]

References

1. Shim JS, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014 Jul 22;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]
2. Coulston A, Boushey C, Ferruzzi M. Nutrition in the Prevention and Treatment of Disease. 3rd edition. Cambridge, MA, USA: Elsevier Academic Press; 2012.
3. Ortega RM, Pérez-Rodrigo C, López-Sobaler AM. Dietary assessment methods: dietary records. Nutr Hosp 2015 Feb 26;31 Suppl 3:38-45 [FREE Full text] [doi: 10.3305/nh.2015.31.sup3.8749] [Medline: 25719769]
4. Willett W. Nutritional Epidemiology. 3rd edition. Oxford, UK: Oxford University Press; 2012.
5. Gersovitz M, Madden JP, Smiciklas-Wright H. Validity of the 24-hr. dietary recall and seven-day record for group comparisons. J Am Diet Assoc 1978 Jul;73(1):48-55. [Medline: 659761]
6. Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SA, Caan B, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. Am J Epidemiol 2011 Sep 01;174(5):591-603 [FREE Full text] [doi: 10.1093/aje/kwr140] [Medline: 21765003]
7. Briefel RR, Sempos CT, McDowell MA, Chien S, Alaimo K. Dietary methods research in the third National Health and Nutrition Examination Survey: underreporting of energy intake. Am J Clin Nutr 1997 Apr;65(4 Suppl):1203S-1209S. [doi: 10.1093/ajcn/65.4.1203S] [Medline: 9094923]
8. Ahmad Z, Kerr DA, Bosch M, Boushey CJ, Delp EJ, Khanna N, et al. A mobile food record for integrated dietary assessment. MADiMa16 (2016) 2016 Oct;2016:53-62 [FREE Full text] [doi: 10.1145/2986035.2986038] [Medline: 28691119]
9. Subar AF, Kirkpatrick SI, Mittl B, Zimmerman TP, Thompson FE, Bingley C, et al. The automated self-administered 24-hour dietary recall (ASA24): a resource for researchers, clinicians, and educators from the National Cancer Institute. J Acad Nutr Diet 2012 Aug;112(8):1134-1137 [FREE Full text] [doi: 10.1016/j.jand.2012.04.016] [Medline: 22704899]
10. Kong F, Tan J. DietCam: automatic dietary assessment with mobile camera phones. Pervasive Mob Comput 2012 Feb;8(1):147-163. [doi: 10.1016/j.pmcj.2011.07.003]
11. Labonté M, Cyr A, Baril-Gravel L, Royer MM, Lamarche B. Validity and reproducibility of a Web-based, self-administered food frequency questionnaire. Eur J Clin Nutr 2012 Feb;66(2):166-173. [doi: 10.1038/ejcn.2011.163] [Medline: 21934698]
12. Boushey CJ, Spoden M, Zhu FM, Delp EJ, Kerr DA. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. Proc Nutr Soc 2017 Aug;76(3):283-294. [doi: 10.1017/S0029665116002913] [Medline: 27938425]
13. Burrows TL, Ho YY, Rollo ME, Collins CE. Validity of dietary assessment methods when compared to the method of doubly labeled water: a systematic review in adults. Front Endocrinol (Lausanne) 2019 Dec 17;10:850 [FREE Full text] [doi: 10.3389/fendo.2019.00850] [Medline: 31920966]
14. Gemming L, Utter J, Ni Mhurchu C. Image-assisted dietary assessment: a systematic review of the evidence. J Acad Nutr Diet 2015 Jan;115(1):64-77. [doi: 10.1016/j.jand.2014.09.015] [Medline: 25441955]
15. Brazeau A, Messier V, Talbo MK, Gagnon C, Taleb N, Fortier I, BETTER Study Group. Self-reported severe and nonsevere hypoglycemia in type 1 diabetes: population surveillance through the BETTER patient engagement registry: development and baseline characteristics. Can J Diabetes 2022 May 28. [doi: 10.1016/j.jcjd.2022.05.010] [Medline: 35835670]
16. Wehling H, Lusher J. People with a body mass index 30 under-report their dietary intake: a systematic review. J Health Psychol 2019 Dec;24(14):2042-2059. [doi: 10.1177/1359105317714318] [Medline: 28810493]
17. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
18. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. London, UK: CRC Press; 1996:189-194.

XSL•FO
RenderX

19. Kirkpatrick SI, Subar AF, Douglass D, Zimmerman TP, Thompson FE, Kahle LL, et al. Performance of the automated self-administered 24-hour recall relative to a measure of true intakes and to an interviewer-administered 24-h recall. Am J Clin Nutr 2014 Jul;100(1):233-240 [FREE Full text] [doi: 10.3945/ajcn.114.083238] [Medline: 24787491]
20. Park Y, Dodd KW, Kipnis V, Thompson FE, Potischman N, Schoeller DA, et al. Comparison of self-reported dietary intakes from the automated self-administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am J Clin Nutr 2018 Jan 01;107(1):80-93 [FREE Full text] [doi: 10.1093/ajcn/nqx002] [Medline: 29381789]
21. Canadian Nutrient File (CNF) - Search by food. Health Canada, Government of Canada. URL: https://food-nutrition.canada.ca/cnf-fce/ [accessed 2022-06-21]
22. FoodData Central. U.S. Department of Agriculture. URL: https://fdc.nal.usda.gov/ [accessed 2022-06-21]
23. Korean Food Composition Database. National Institute of Agricultural Sciences. URL: http://koreanfood.rda.go.kr/eng/fctFoodSrchEng/engMain [accessed 2022-06-21]
24. Food Nutrient Finder. Centre for Food Safety. URL: https://www.cfs.gov.hk/english/nutrient/presearch3.php [accessed 2022-06-21]
25. Longvah T, Ananthan R, Bhaskar K, Venkaiah K. Indian Food Composition Tables. Hyderabad, India: National Institute of Nutrition; 2017.
26. Australian Food Composition Database - Release 2.0. Food Standards Australia & New Zealand. URL: https://www.foodstandards.gov.au/science/monitoringnutrients/afcd/Pages/foodsearch.aspx [accessed 2022-06-21]
27. Ji Y, Plourde H, Bouzo V, Kilgour RD, Cohen TR. Validity and usability of a smartphone image-based dietary assessment app compared to 3-day food diaries in assessing dietary intake among Canadian adults: randomized controlled trial. JMIR Mhealth Uhealth 2020 Sep 09;8(9):e16953 [FREE Full text] [doi: 10.2196/16953] [Medline: 32902389]
28. Bouzo V, Plourde H, Beckenstein H, Cohen TR. Evaluation of the diet tracking smartphone application Keenoa™: a qualitative analysis. Can J Diet Pract Res 2022 Mar 01;83(1):25-29 [FREE Full text] [doi: 10.3148/cjdpr-2021-022] [Medline: 34582258]
29. Serra-Majem L, Frost Andersen L, Henríque-Sánchez P, Doreste-Alonso J, Sánchez-Villegas A, Ortiz-Andrelluchi A, et al. Evaluating the quality of dietary intake validation studies. Br J Nutr 2009 Dec;102 Suppl 1:S3-S9. [doi: 10.1017/S0007114509993114] [Medline: 20100366]
30. Brown CC, Kipnis V, Freedman LS, Hartman AM, Schatzkin A, Wacholder S. Energy adjustment methods for nutritional epidemiology: the effect of categorization. Am J Epidemiol 1994 Feb 01;139(3):323-338. [doi: 10.1093/oxfordjournals.aje.a117000] [Medline: 8116608]
31. Trafimow D. The attenuation of correlation coefficients: a statistical literacy issue. Teach Stat 2016;38(1):25-28. [doi: 10.1111/test.12087]
32. Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. Nutr J 2015 Apr 22;14:40 [FREE Full text] [doi: 10.1186/s12937-015-0027-y] [Medline: 25897837]
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
34. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, et al. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. Eur J Clin Nutr 1991 Dec;45(12):569-581. [Medline: 1810719]
35. Johansson L, Solvoll K, Bjørneboe GE, Drevon CA. Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. Am J Clin Nutr 1998 Aug;68(2):266-274. [doi: 10.1093/ajcn/68.2.266] [Medline: 9701182]
36. Mifflin MD, St Jeor ST, Hill LA, Scott BJ, Daugherty SA, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. Am J Clin Nutr 1990 Feb;51(2):241-247. [doi: 10.1093/ajcn/51.2.241] [Medline: 2305711]
37. Basiotis PP, Welsh SO, Cronin FJ, Kelsay JL, Mertz W. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. J Nutr 1987 Sep;117(9):1638-1641. [doi: 10.1093/jn/117.9.1638] [Medline: 3655942]
38. Tarasuk V, Beaton GH. Day-to-day variation in energy and nutrient intake: evidence of individuality in eating behaviour? Appetite 1992 Feb;18(1):43-54. [doi: 10.1016/0195-6663(92)90209-o] [Medline: 1562201]
39. Sebastian RS, Enns CW, Steinfeldt LC, Goldman JD, Moshfegh AJ. Discontinuation of data processing step:salt adjustment on designated foods likely to be home prepared. U.S. Department of Agriculture. 2012. URL: https://naldc.nal.usda.gov/catalog/55318 [accessed 2022-06-16]
40. Subar AF, Potischman N, Dodd KW, Thompson FE, Baer DJ, Schoeller DA, et al. Performance and feasibility of recalls completed using the automated self-administered 24-hour dietary assessment tool in relation to other self-report tools and biomarkers in the interactive diet and activity tracking in AARP (IDATA) study. J Acad Nutr Diet 2020 Nov;120(11):1805-1820 [FREE Full text] [doi: 10.1016/j.jand.2020.06.015] [Medline: 32819883]
41. Chen J, Berkman W, Bardouh M, Ng CY, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. Nutrition 2019 Jan;57:208-216. [doi: 10.1016/j.nut.2018.05.003] [Medline: 30184514]

XSL•FO
RenderX

42. Van Asbroeck S, Matthys C. Use of different food image recognition platforms in dietary assessment: comparison study. JMIR Form Res 2020 Dec 07;4(12):e15602 [FREE Full text] [doi: 10.2196/15602] [Medline: 33284118]
43. Recio-Rodriguez JI, Rodriguez-Martin C, Gonzalez-Sanchez J, Rodriguez-Sanchez E, Martin-Borras C, Martínez-Vizcaino V, EVIDENT Investigators. EVIDENT smartphone app, a new method for the dietary record: comparison with a food frequency questionnaire. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e11463 [FREE Full text] [doi: 10.2196/11463] [Medline: 30735141]
44. Lee JE, Song S, Ahn JS, Kim Y, Lee JE. Use of a mobile application for self-monitoring dietary intake: feasibility test and an intervention study. Nutrients 2017 Jul 13;9(7):748 [FREE Full text] [doi: 10.3390/nu9070748] [Medline: 28703758]
45. Mescoloto SB, Caivano S, Domene SM. Evaluation of a mobile application for estimation of food intake. Rev Nutr 2017 Feb;30(1):91-98. [doi: 10.1590/1678-98652017000100009]
46. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2013 Feb 14;109(3):539-546. [doi: 10.1017/S0007114512001353] [Medline: 22717334]
47. Rangan AM, O'Connor S, Giannelli V, Yap ML, Tang LM, Roy R, et al. Electronic dietary intake assessment (e-DIA): comparison of a mobile phone digital entry app for dietary data collection with 24-hour dietary recalls. JMIR Mhealth Uhealth 2015 Oct 27;3(4):e98 [FREE Full text] [doi: 10.2196/mhealth.4613] [Medline: 26508282]
48. Wellard-Cole L, Chen J, Davies A, Wong A, Huynh S, Rangan A, et al. Relative validity of the eat and track (EaT) smartphone app for collection of dietary intake data in 18-to-30-year olds. Nutrients 2019 Mar 14;11(3):621 [FREE Full text] [doi: 10.3390/nu11030621] [Medline: 30875772]
49. Ambrosini GL, Hurworth M, Giglia R, Trapp G, Strauss P. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. Nutr J 2018 Jan 09;17(1):5 [FREE Full text] [doi: 10.1186/s12937-018-0315-4] [Medline: 29316930]
50. Pendergast FJ, Ridgers ND, Worsley A, McNaughton SA. Evaluation of a smartphone food diary application using objectively measured energy expenditure. Int J Behav Nutr Phys Act 2017 Mar 14;14(1):30 [FREE Full text] [doi: 10.1186/s12966-017-0488-9] [Medline: 28288657]
51. Boushey CJ, Spoden M, Delp EJ, Zhu F, Bosch M, Ahmad Z, et al. Reported energy intake accuracy compared to doubly labeled water and usability of the mobile food record among community dwelling adults. Nutrients 2017 Mar 22;9(3):312 [FREE Full text] [doi: 10.3390/nu9030312] [Medline: 28327502]
52. Zhang L, Misir A, Boshuizen H, Ocké M. A systematic review and meta-analysis of validation studies performed on dietary record apps. Adv Nutr 2021 Dec 01;12(6):2321-2332 [FREE Full text] [doi: 10.1093/advances/nmab058] [Medline: 34019624]
53. Chui TK, Tan J, Li Y, Raynor HA. Validating an automated image identification process of a passive image-assisted dietary assessment method: proof of concept. Public Health Nutr 2020 Oct;23(15):2700-2710. [doi: 10.1017/S1368980020000816] [Medline: 32517834]
54. Sun M, Burke LE, Mao ZH, Chen Y, Chen HC, Bai Y, et al. eButton: a wearable computer for health monitoring and personal assistance. Proc Des Autom Conf 2014;2014:1-6 [FREE Full text] [doi: 10.1145/2593069.2596678] [Medline: 25340176]
55. Foster E, Lee C, Imamura F, Hollidge SE, Westgate KL, Venables MC, et al. Validity and reliability of an online self-report 24-h dietary recall method (Intake24): a doubly labelled water study and repeated-measures analysis. J Nutr Sci 2019 Aug 30;8:e29 [FREE Full text] [doi: 10.1017/jns.2019.20] [Medline: 31501691]
56. Wark PA, Hardie LJ, Frost GS, Alwan NA, Carter M, Elliott P, et al. Validity of an online 24-h recall tool (myfood24) for dietary assessment in population studies: comparison with biomarkers and standard interviews. BMC Med 2018 Aug 09;16(1):136 [FREE Full text] [doi: 10.1186/s12916-018-1113-8] [Medline: 30089491]
57. Lafrenière J, Lamarche B, Laramée C, Robitaille J, Lemieux S. Validation of a newly automated Web-based 24-hour dietary recall using fully controlled feeding studies. BMC Nutr 2017 Apr 5;3:34 [FREE Full text] [doi: 10.1186/s40795-017-0153-3] [Medline: 32153814]
58. Greenwood DC, Hardie LJ, Frost GS, Alwan NA, Bradbury KE, Carter M, et al. Validation of the Oxford WebQ online 24-hour dietary questionnaire using biomarkers. Am J Epidemiol 2019 Oct 01;188(10):1858-1867 [FREE Full text] [doi: 10.1093/aje/kwz165] [Medline: 31318012]
59. Bailey RL. Overview of dietary assessment methods for measuring intakes of foods, beverages, and dietary supplements in research studies. Curr Opin Biotechnol 2021 Aug;70:91-96 [FREE Full text] [doi: 10.1016/j.copbio.2021.02.007] [Medline: 33714006]

Abbreviations

ASA24: automated self-administered 24-hour recall
BETTER: Behaviors, Therapies, Technologies, and Hypoglycemic Risk in Type 1 Diabetes
CNF: Canadian nutrient file
LOA: limit of agreement

XSL•FO
RenderX

SUS: System Usability Scale

Edited by G Eysenbach; submitted 22.06.22; peer-reviewed by S Lemieux, G Ash; comments to author 29.08.22; revised version received 09.09.22; accepted 23.10.22; published 21.11.22.

Please cite as:
Moyen A, Rappaport AI, Fleurent-Grégoire C, Tessier AJ, Brazeau AS, Chevalier S
Relative Validation of an Artificial Intelligence–Enhanced, Image-Assisted Mobile App for Dietary Assessment in Adults: Randomized Crossover Study
J Med Internet Res 2022;24(11):e40449
URL: https://www.jmir.org/2022/11/e40449
doi:10.2196/40449
PMID:36409539

©Audrey Moyen, Aviva Ilysse Rappaport, Chloé Fleurent-Grégoire, Anne-Julie Tessier, Anne-Sophie Brazeau, Stéphanie Chevalier. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Psychometric Properties of the Chinese Warwick-Edinburgh Mental Well-being Scale in Medical Staff: Cross-sectional Study

Aishu Dong[1*], MD; Jing Huang[2*], BA; Shudan Lin[2], MD; Jianing Zhu[2], BA; Haitao Zhou[1], BA; Qianqian Jin[3], BA; Wei Zhao[1], BA; Lianlian Zhu[2], MD; Wenjian Guo[4], PhD

[1]Cardiology Department, Second Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[2]Wenzhou Medical University, Wenzhou, China

[3]Emergency Department, Second Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

[4]Oncology Department, Second Affiliated Hospital of Wenzhou Medical University, Whenzhou, China

[*]these authors contributed equally

Corresponding Author:
Wenjian Guo, PhD
Oncology Department
Second Affiliated Hospital of Wenzhou Medical University
College West Road 109, 0577
Whenzhou, 325000
China
Phone: 86 13957729220
Email: guowenjian@wmu.edu.cn

Abstract

Background: Worldwide, mental well-being is a critical issue for public health, especially among medical staff; it affects professionalism, efficiency, quality of care delivery, and overall quality of life. Nevertheless, assessing mental well-being is a complex problem.

Objective: This study aimed to evaluate the psychometric properties of the Chinese-language version of the 14-item Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in medical staff recruited mainly from 6 hospitals in China and provide a reliable measurement of positive mental well-being.

Methods: A cross-sectional online survey was conducted of medical staff from 15 provinces in China from May 15 to July 15, 2020. Confirmatory factor analysis (CFA) was conducted to test the structure of the Chinese WEMWBS. The Spearman correlations of the Chinese WEMWBS with the 5-item World Health Organization Well-Being Index (WHO-5) were used to evaluate convergent validity. The Cronbach α and split-half reliability (λ) represented internal consistency. A graded response model was adopted for an item response theory (IRT) analysis. We report discrimination, difficulty, item characteristic curves (ICCs), and item information curves (IICs). ICCs and IICs were used to estimate reliability and validity based on the IRT analysis.

Results: A total of 572 participants from 15 provinces in China finished the Chinese WEMWBS. The CFA showed that the 1D model was satisfactory and internal consistency reliability was excellent, with α=.965 and λ=0.947, while the item-scale correlation coefficients ranged from r=0.727 to r=0.900. The correlation coefficient between the Chinese WEMWBS and the WHO-5 was significant, at r=0.746. The average variance extraction value was 0.656, and the composite reliability value was 0.964, with good aggregation validity. The discrimination of the Chinese WEMWBS items ranged from 2.026 to 5.098. The ICCs illustrated that the orders of the category thresholds for the 14 items were satisfactory.

Conclusions: The Chinese WEMWBS showed good psychometric properties and can measure well-being in medical staff.

(J Med Internet Res 2022;24(11):e38108) doi:10.2196/38108

KEYWORDS

psychometric property; Chinese Warwick-Edinburgh Mental Well-being Scale; classical test theory; well-being; item response theory; medical staff; China

XSL•FO
RenderX

Introduction

Background

Mental well-being is a public health concern worldwide; adequate mental well-being is associated with better health-related quality of life and longer life expectancy [1]. In recent years, the mental well-being of employees in several occupations has gained substantial attention [2-6]. A meta-analysis revealed that numerous health care workers had various psychological problems [7]. It is well known that medical staff experience many work-related stresses (eg, prolonged and irregular working hours, night shifts, high-intensity work, emotional exhaustion, chronicity of care, and moral conflicts), which may negatively influence their mental well-being, causing depression, anxiety, sleeping disorders, and other problems. Impaired mental well-being can affect health care providers' professionalism, quality of care delivery, efficiency, and overall quality of life [8,9].

Moreover, it has been reported that the overall mental health status of Chinese medical staff is unfavorable [10,11]. This finding suggests that the mental well-being of medical staff is critically important to public health [12,13]. For this reason, it is crucial to measure the mental health status of medical staff and identify work-related risk factors to protect their well-being [14].

The Warwick-Edinburgh Mental Well-being Scale (WEMWBS) is a relatively new, short, acceptable scale that has been translated into several languages [15-18]. It has demonstrated excellent reliability, good validity, and internal consistency [19]. Studies of public mental health have confirmed the WEMWBS's ability to offer rigor in psychological evaluations [20]; it focuses on protective and promoting factors that can provide a rational basis for the orientation of policy makers formulating interventions [21].

Previous studies have reported the psychological performance of the Chinese-language version of the WEMWBS in clinical and nonclinical settings in China, but all have had limitations [15,22]. Research by Liu et al [23] appears to be the earliest psychometric analysis of the Chinese WEMWBS; however, 2 issues need addressing. First, their paper was written in Chinese, making it burdensome to read for non–Chinese-speaking investigators and impeding comparisons of China with other countries. Second, the age of the study participants ranged from 60 to 97 years, resulting in information and selection bias. The generalizability of the findings from Dong et al [22] is problematic, because the 191 patients with chronic heart failure in that study came from 1 hospital in a Chinese city. A study by Fung [24] and an earlier study by Dong et al [15] were limited because all respondents were university students recruited from either a single university or a single hospital nursing internship program in a Chinese city; this could have caused pervasive information and selection bias in these studies' assessment of the psychometric properties of the WEMWBS. A study by Waqas et al [25] explored the reliability and validity of the WEMWBS in Pakistan; Taggart et al [26] investigated the WEMWBS in a targeted sample of minority ethnic groups living in the UK who self-identified as Chinese or Pakistani by background. Additionally, no previous investigation has combined a graded response model (GRM), item response theory (IRT), and classical test theory (CTT) to evaluate the psychometric properties of the WEMWBS. It is necessary to find a comprehensive method and a better representative sample that covers participants from southern and northern areas to assess the performance of the Chinese WEMWBS.

Objective of the Study

We administered the Chinese WEMWBS to medical staff to evaluate their psychological characteristics and explore and popularize this questionnaire on mental well-being, which is suitable for Chinese national conditions. We aim to provide theoretical support for improving the mental well-being of medical staff.

Methods

Data Collection

From May 15 to July 15, 2020, purposeful sampling was conducted to recruit 572 medical staff online, mainly from 6 hospitals in mainland China (the First Affiliated Hospital of Wenzhou Medical University, the Second Affiliated Hospital of Wenzhou Medical University, the Second Hospital of Dalian Medical University, the Second Affiliated Hospital of Zhongguo Medical University, Lishui People's Hospital, and Chenzhou Third People's Hospital).

Ethics Approval

All participants provided informed consent before participation, and the Medical Ethics Committee of the Second Affiliated Hospital of Wenzhou Medical University approved the study (LCKY2019-288).

Instruments

Data were collected via a self-administered online questionnaire. The first section collected sociodemographic characteristics, including age, marital status, gender, body weight (in kilograms), height (in meters), professional status, and education level. The second section examined lifestyle habits, including working hours, night shifts per week, smoking history, drinking history, consumption of vegetables and fruit, physical exercise, and self-reported personality. The third section examined mental well-being using the WEMWBS and self-perceived quality of life (QoL). The WEMWBS is a 14-item sequential scale that measures 3 aspects of mental well-being: positive psychological function, emotion, and interpersonal relationship satisfaction. All items were scored on a 5-point Likert scale, including 1 (never), 2 (occasionally), 3 (yes), 4 (often), and 5 (always). The total score ranged from 14 to 70, with higher scores representing stronger subjective well-being. The third section of the questionnaire used the 36-Item Short Form Health Survey, Version 2 (SF-36 v2) to assess self-perceived QoL. The SF-36 v2 is a 36-item structured scale that comprehensively summarizes respondents' QoL across 8 dimensions: physical functioning (10 items), role-physical (4 items), bodily pain (2 items), general health (5 items), vitality (4 items), social functioning (2 items), role-emotional (3 items), and mental health (5 items). The physical component summary and the

XSL•FO
RenderX

mental component summary are 2 subscales of the 8 dimensions. In addition to the 8 dimensions listed above, the SF-36 v2 includes another health condition, reported health transition, which measures overall changes in health status over the past year.

Statistical Analysis

We used EpiData (version 3.1; EpiData Association) for double entry and data management. Data collection and analysis were carried out using SPSS (version 27.0; IBM Corp) and R (version 4.1.1; R Foundation for Statistical Computing). Means and SDs were calculated for continuous data and frequencies and percentages for categorical data.

Dimensionality Test

Principal component analysis of the Chinese WEMWBS was used to independently identify a 1D hypothesis; this analysis indicates good quality (ie, statistical power) of the 1D structure of the model when the first eigenvalue is more than 50% of the total variation.

Ceiling Effect and Floor Effect

A ceiling or floor effect is present when subjects receive the scale's highest or lowest score. Measurement scales with ceiling or floor effects may have questionable validity, reliability, and reactivity. The significance level should be 20%.

Item Analysis

Item analysis determines effectiveness and the ability to discriminate the entire scale. The process used is to sum the scores of the items for each participant, divide them into high-score and low-score groups (with 27% and 73% quantiles as the boundaries), and finally use a 2-tailed *t* test to identify differences between the groups. If there is a difference, the scale item is appropriately designed; otherwise, it indicates that the item has a questionable ability to discriminate between respondents, meaning that the item should be deleted or rearranged.

Reliability Analysis: Internal Consistency of the Scale

We used the Cronbach α and split-half reliability (λ) to represent internal consistency reliability. The former indicates the homogeneity of each item in the scale; we considered α=.7 as the threshold above which the scale showed desired reliability. The latter measures consistency between the 2 halves of these items, divided according to the precedence and the odd-even sequence of the serial number. Generally, a correlation coefficient of $r \geq 0.70$ is considered acceptable.

Test-Retest Reliability

The test-retest reliability of the WEMWBS scale was estimated within a 2-week interval by comparing 2 sets of scores using the intraclass correlation coefficient.

Construct Validity

Confirmatory factor analysis (CFA) of item responses was implemented using the weighted least-squares method to test the structural equation modeling of the hypothesized unidimensionality of the WEMWBS. Statistical analysis of correlations was performed using SAS (version 9.4; SAS Institute Inc), assuming no relationship between the residuals. A stepwise strategy was then used to add the matrix elements with the highest dependencies until sufficient fit statistics were achieved.

The predicted levels of the goodness-of-fit index and adjusted goodness-of-fit index based on degrees of freedom correction were >0.9 and >0.8, respectively.

A root mean square error of approximation (RMSEA) below the accepted level of 0.06 [27] indicates only a tiny number of unintended deviations. A chi-square statistic with $P<.05$ indicates a considerable amount of actual covariance between measurements that the model cannot explain [28]. Nevertheless, large sample sizes may exaggerate this and are therefore unsuitable [29].

Compatible Validity

This parameter refers to the extent to which the scores of the new scale are relevant to the scores of another scale with the same content and known validity. If the compatibility coefficient is high, the 2 scales measure the same content, and the new scale is equally effective. Based on the range of these 2 scales, we hypothesized a strong correlation between the WEMWBS and the 5-item World Health Organization Well-Being Index (WHO-5) scale for capturing mental well-being, with a coefficient above $r=0.7$.

Convergence Validity

Convergence validity refers to the similarity of measurement results when different algorithmic methods are grouped to determine the same feature. The evaluation indices usually include composite reliability (CR), factor loading, and average variance extracted (AVE), where AVE greater than 0.5 and CR greater than 0.7 indicate that the aggregation validity is acceptable.

IRT Analysis

IRT, also known as latent trait theory, is a modern psychometric theory proposed to compensate for the limitations of CTT. According to an exploratory factor analysis of CTT, the Chinese WEMWBS is a 1D scale. Therefore, in this study, the responses of the 572 participants to the WEMWBS on a 5-point Likert-type scale were interpreted with the Samejima GRM [30]. These parameters, including a discrimination parameter (referred to as *a*), a difficulty parameter (referred to as *b*), item characteristic curves (ICCs), and item information curves (IICs), were administered to implement filtering entry. The discrimination parameter evaluates the strength of the relationship between each item and the scale; the difficulty parameter identifies an item in the potential continuum of the structure that best distinguishes each item. Each item has 5 levels; we used level 1 as a reference and set the remaining 4 levels as difficulty levels. The difficulty level parameter was calculated between 1 and 2, 2 and 3, 3 and 4, and 4 and 5, denoted as thresholds: ≥2, ≥3, ≥4, and 5.

When the discrimination parameter is <0.4 or >3 and the difficulty parameter range exceeds –3 to 3, the item should be considered for deletion. The model simulates ICCs for each option for the 14 items. The first and fifth ICCs change

XSL•FO
RenderX

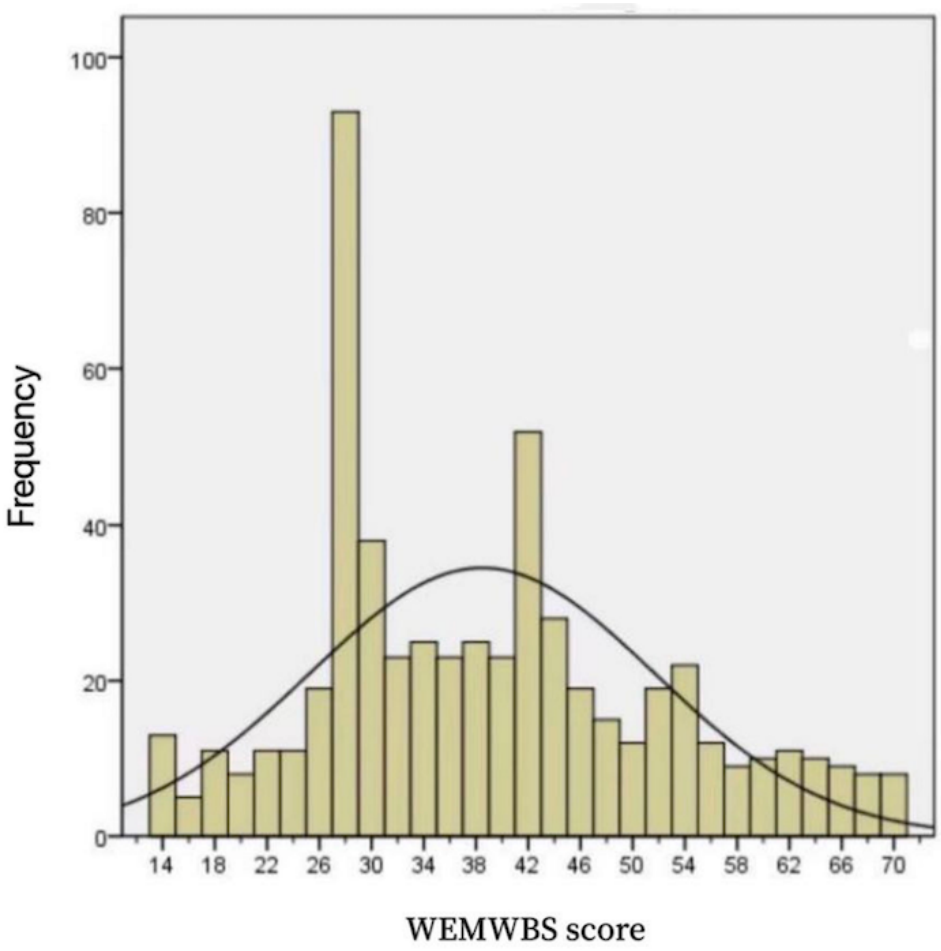
unvaryingly, and the second, third, and fourth ICCs are typically distributed, which can be considered ideal. The more ideal the ICC distribution, the more considerable the corresponding project information. Moreover, a larger item information function results in greater accuracy. Item screening was then carried out. When an item did not meet the requirement for 3 or more parameters, it was considered for deletion based on professional knowledge and expert opinion. These calculations were performed using Stata/MP (version 14.0 for Mac; StataCorp LP).

Results

Descriptive Statistics of the Scale

The total sample of 572 medical staff had a mean score for the Chinese WEMWBS of 38.47 (95% CI 37.45-39.61; SD 13.23; skewness 0.449; kurtosis –0.486) and a median score of 37, indicating a latent skewed trait distribution (Figure 1). An independent-sample *t* test showed no difference between the total WEMWBS score and gender (t_1=–1.477; *P*=.14). A Pearson correlation analysis did not indicate any significant relationship between the score for mental well-being and age; therefore, further validation analyses did not include participant age.

Figure 1. Histogram showing the scores of medical staff (N=572) on the Chinese-language version of the WEMWBS. The mean score was 38.47 (SD 13.227). WEMWBS: Warwick-Edinburgh Mental Well-being Scale.



Item Analysis

As shown in Table 1, the values for specific items were significantly different in the high-score and low-score groups (*P*<.001), meaning that all 14 items could differentiate the 2 groups well, and that none should be discarded. The correlation coefficient between each item and the total score of the instrument ranged from *r*=0.727 to *r*=0.900. As seen in Table 2, none of the items reached a rate of 20%, suggesting that there were no ceiling or floor effects.

XSL•FO
RenderX

Table 1. Item analysis (discrimination analysis) and item-scale correlation of the Chinese version of the Warwick-Edinburgh Mental Well-being Scale.

Item	Low-score group (n=171), mean (SD) score	High-score group (n=164), mean (SD) score	*t* (decision value)	*P* value	Item-scale correlation, *r*
1	1.78 (0.68)	3.91 (0.98)	–23.125	<.001	0.778
2	1.56 (0.51)	3.66 (0.99)	–24.248	<.001	0.823
3	1.81 (0.53)	4.24 (0.81)	–32.428	<.001	0.831
4	1.77 (0.65)	4.05 (0.94)	–25.866	<.001	0.727
5	1.88 (0.56)	4.43 (0.67)	–37.806	<.001	0.859
6	1.75 (0.46)	3.89 (0.87)	–27.92	<.001	0.862
7	1.75 (0.43)	3.84 (0.90)	–26.892	<.001	0.871
8	1.76 (0.45)	4.16 (0.78)	–34.36	<.001	0.900
9	1.70 (0.48)	3.73 (0.96)	–24.182	<.001	0.842
10	1.82 (0.53)	4.12 (0.79)	–31.044	<.001	0.870
11	1.72 (0.48)	3.70 (1.00)	–22.947	<.001	0.819
12	1.69 (0.48)	3.77 (0.89)	–26.601	<.001	0.817
13	1.73 (0.50)	3.96 (0.91)	–27.705	<.001	0.789
14	1.77 (0.46)	4.04 (0.81)	–31.21	<.001	0.879

Table 2. Floor effect, ceiling effect, and item-scale correlation of Chinese-language version of the Warwick-Edinburgh Mental Well-being Scale. The floor and ceiling effects were defined as the lowest (1 point) and highest (5 points) scores, respectively (N=572).

Item	Subjects with floor effect, n (%)	Subjects with ceiling effect, n (%)	Item-scale correlation[a], *r*
1	69 (12.1)	67 (11.7)	0.764
2	100 (17.5)	44 (7.7)	0.809
3	45 (7.9)	86 (15)	0.823
4	79 (13.8)	82 (14.3)	0.697
5	34 (5.9)	92 (16.1)	0.837
6	47 (8.2)	49 (8.6)	0.867
7	51 (8.9)	43 (7.5)	0.885
8	45 (7.9)	69 (12.1)	0.904
9	63 (11)	40 (7)	0.835
10	45 (7.9)	64 (11.2)	0.865
11	63 (11)	45 (7.9)	0.832
12	66 (11.5)	44 (7.7)	0.825
13	67 (11.7)	58 (10.1)	0.768
14	47 (8.2)	52 (9.1)	0.875

[a]Correlations were deemed significant at the *P*<.01 (ie, 2 significant figures) level.

Reliability Analysis

Internal consistency reliability was good (Cronbach α=.965). The corrected item-total correlation values of the items were all greater than 0.5, indicating a good correlation between items and reliability (Table 3). Two weeks after completing the questionnaire, 35 subjects completed it again; the test-retest reliability was measured at 0.810, indicating that the scale had good stability. The split-half reliability of the scale was λ=0.947 according to the first half and the second half of the serial number, while the value was λ=0.970 according to the odd-even status of the serial number.

XSL•FO
RenderX

Table 3. Cronbach reliability analysis of Chinese-language version of the Warwick-Edinburgh Mental Well-being Scale.

Item	Average score after deleting each item	Scaled variance after deleting terms	Corrected item-total correlation	Squared multiple correlation	Cronbach α if item deleted
1	35.73	151.766	0.739	0.642	.964
2	35.97	151.654	0.792	0.691	.963
3	35.52	149.833	0.799	0.711	.963
4	35.60	152.174	0.676	0.510	.966
5	35.43	149.269	0.832	0.743	.962
6	35.75	151.830	0.839	0.769	.962
7	35.84	151.722	0.850	0.791	.962
8	35.63	148.971	0.881	0.822	.961
9	35.87	152.497	0.817	0.716	.963
10	35.61	150.137	0.846	0.774	.962
11	35.87	152.946	0.790	0.677	.963
12	35.83	152.408	0.786	0.664	.963
13	35.71	151.672	0.752	0.608	.964
14	35.75	150.636	0.858	0.756	.962

Construct Validity: Exploratory Factor Analysis

A Kaiser-Meyer-Olkin value of 0.963 for the 14 items and a value of 7844.584 for the Bartlett sphericity test (P<.001) demonstrated that the data obtained were suitable for factor analysis. A principal component factor analysis was used with varimax rotation to evaluate construct validity. Table 4 shows factor loadings for the 14 items, which ranged from 0.714 for item 4 to 0.903 for item 8.

Table 4. Validity analysis result of Chinese-language version of the Warwick-Edinburgh Mental Well-being Scale.

Item	Factor loading	Common degree (common factor variance)
1	0.772	0.597
2	0.821	0.675
3	0.828	0.685
4	0.714	0.510
5	0.856	0.733
6	0.868	0.753
7	0.877	0.769
8	0.903	0.816
9	0.848	0.719
10	0.873	0.762
11	0.823	0.678
12	0.818	0.669
13	0.784	0.615
14	0.881	0.776

CFA Results

An analysis of mean average precision (MAP) showed that the WEMWBS had a 1D structure. The minor average squared partial correlation was 0.02221, and the most negligible average fourth-power partial correlation was 0.00100. According to the revised MAP test [31], the number of factors was 1.

We conducted a CFA test of the hypothetical single-factor structure of the Chinese WEMWBS and measured the goodness-of-fit of the single confirmatory factor model. Assuming that there was no correlation between the residuals, the initial model fit poorly. The χ^2/df was 8.437; the comparative fitting index (CFI) was 0.927; the RMSEA was 0.114; for the normed fit index (NFI), delta 1 was 0.918; for the relative fit index (RFI), rho 1 was 0.903; for the incremental fit index (IFI),

XSL•FO
RenderX

delta 2 was 0.927; for and the Tacker-Lewis index (TLI), rho was 2.914.

Compatible Validity

There was a significant positive correlation between the Chinese WEMWBS and the WHO-5, with a correlation coefficient of 0.746 (95% CI 0.722-0.794; *P*<.01).

Combination Reliability and Convergent Validity

A CFA showed that the AVE value was 0.674 (ie, greater than 0.5). The CR value was 0.966 (ie, greater than 0.7), suggesting that the sample had good convergence validity.

IRT Analysis

Table 5 shows the results of the GRM analysis. The discrimination difference indices of the items ranged from 2.026 to 5.098, which demonstrates that the Chinese WEMWBS scores of low-score individuals differed from high-score individuals, corresponding to latent trait sensitivity. The item difficulty of thresholds ≥2, ≥3, ≥4, and 5 ranged from 1.06 to 1.73, 0 to 0.23, 0.56 to 1.06, and 1.12 to 1.66, respectively.

XSL•FO
RenderX

Table 5. Results of the graded response model analysis of the Chinese-language version of the Warwick-Edinburgh Mental Well-being Scale.

Item	Coefficient	95% CI	SE	*z*	*P*>*z*
1					
Discrimination difference	2.526	2.200-2.853	0.167	15.160	<.001
Item difficulty					
≥2	1.394	1.578-1.210	0.094	14.84	<.001
≥3	0.081	0.039-0.201	0.061	1.32	.19
≥4	0.806	0.669-0.943	0.070	11.54	<.001
5	1.455	1.270-1.640	0.094	15.42	<.001
2					
Discrimination difference	3.010	2.621-3.400	0.199	15.140	<.001
Item difficulty					
≥2	1.064	1.212-0.916	0.075	14.09	<.001
≥3	0.232	0.117-0.348	0.059	3.94	<.001
≥4	1.058	0.914-1.203	0.074	14.38	<.001
5	1.656	1.463-1.850	0.099	16.78	<.001
3					
Discrimination difference	3.024	2.641-3.407	0.195	15.490	<.001
Item difficulty					
≥2	1.619	1.814-1.424	0.100	16.25	<.001
≥3	0.119	0.233-0.004	0.058	2.03	.04
≥4	0.559	0.439-0.679	0.061	9.13	<.001
5	1.208	1.053-1.364	0.079	15.22	<.001
4					
Discrimination difference	2.026	1.756-2.297	0.138	14.670	<.001
Item difficulty					
≥2	1.422	1.625-1.219	0.103	13.74	<.001
≥3	0.115	0.245-0.016	0.067	1.72	.09
≥4	0.643	0.501-0.785	0.073	8.87	<.001
5	1.405	1.207-1.602	0.101	13.19	<.001
5					
Discrimination difference	3.495	3.051-3.939	0.227	15.420	<.001
Item difficulty					
≥2	1.726	1.928-1.525	0.103	16.79	<.001
≥3	0.220	0.331-0.109	0.057	3.88	<.001
≥4	0.469	0.356-0.583	0.058	8.10	<.001
5	1.117	0.974-1.261	0.073	15.26	<.001
6					
Discrimination difference	4.110	3.573-4.646	0.274	15.020	<.001
Item difficulty					
≥2	1.480	1.650-1.310	0.087	17.06	<.001
≥3	0.001	0.108-0.106	0.055	0.01	.99
≥4	0.873	0.748-0.997	0.063	13.79	<.001
5	1.459	1.296-1.623	0.084	17.45	<.001

XSL•FO
RenderX

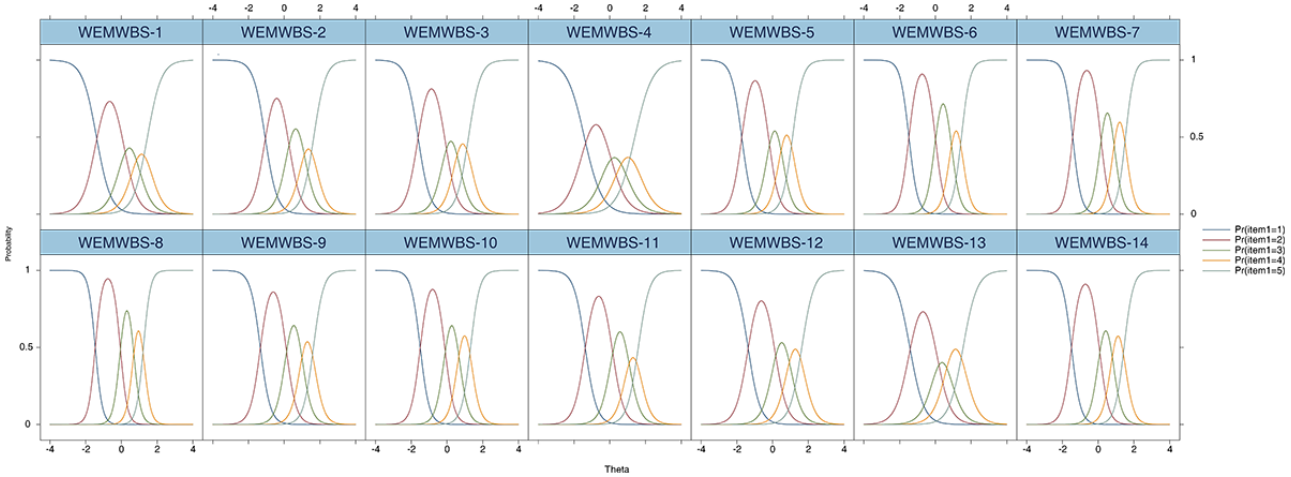
Item	Coefficient	95% CI	SE	*z*	*P*>*z*
7					
Discrimination difference	4.258	3.680-4.835	0.295	14.450	<.001
Item difficulty					
≥2	1.422	1.586-1.258	0.084	17.02	<.001
≥3	0.149	0.042-0.255	0.054	2.74	.006
≥4	0.887	0.762-1.011	0.063	13.99	<.001
5	1.532	1.364-1.701	0.086	17.82	<.001
8					
Discrimination difference	5.098	4.400-5.796	0.356	14.310	<.001
Item difficulty					
≥2	1.459	1.622-1.297	0.083	17.58	<.001
≥3	0.060	0.164-0.043	0.053	1.14	.25
≥4	0.682	0.571-0.793	0.057	12.00	<.001
5	1.235	1.094-1.375	0.072	17.25	<.001
9					
Discrimination difference	3.571	3.111-4.032	0.235	15.200	<.001
Item difficulty					
≥2	1.334	1.496-1.173	0.083	16.17	<.001
≥3	0.113	0.004-0.223	0.056	2.02	.04
≥4	0.967	0.833-1.100	0.068	14.19	<.001
5	1.638	1.455-1.822	0.094	17.47	<.001
10					
Discrimination difference	3.942	3.439-4.446	0.257	15.360	<.001
Item difficulty					
≥2	1.498	1.672-1.323	0.089	16.18	<.001
≥3	0.117	0.224-0.009	0.055	2.12	.03
≥4	0.655	0.539-0.771	0.059	11.05	<.001
5	1.320	1.167-1.473	0.078	16.91	<.001
11					
Discrimination difference	3.171	2.762-3.580	0.209	15.200	<.001
Item difficulty					
≥2	1.374	1.544-1.204	0.087	15.85	<.001
≥3	0.134	0.021-0.247	0.058	2.33	.02
≥4	1.011	0.872-1.151	0.071	14.17	<.001
5	1.597	1.411-1.783	0.095	16.79	<.001
12					
Discrimination difference	2.962	2.583-3.341	0.193	15.320	<.001
Item difficulty					
≥2	1.373	1.545-1.200	0.088	15.59	<.001
≥3	0.118	0.003-0.234	0.059	2.01	.04
≥4	0.917	0.781-1.053	0.069	13.20	<.001
5	1.639	1.445-1.833	0.099	16.56	<.001
13					

XSL•FO
RenderX

Item	Coefficient	95% CI	SE	*z*	*P*>*z*
Discrimination difference	2.539	2.215-2.864	0.166	15.330	<.001
Item difficulty					
≥2	1.427	1.613-1.241	0.095	15.02	<.001
≥3	0.042	0.078-0.162	0.061	0.69	.49
≥4	0.713	0.579-0.846	0.068	10.48	<.001
5	1.554	1.361-1.747	0.099	15.77	<.001
14					
Discrimination difference	3.925	3.411-4.440	0.262	14.960	<.001
Item difficulty					
≥2	1.493	1.666-1.320	0.088	16.92	<.001
≥3	0.066	0.041-0.174	0.055	1.21	.23
≥4	0.785	0.663-0.907	0.062	12.64	<.001
5	1.451	1.287-1.614	0.083	17.38	<.001

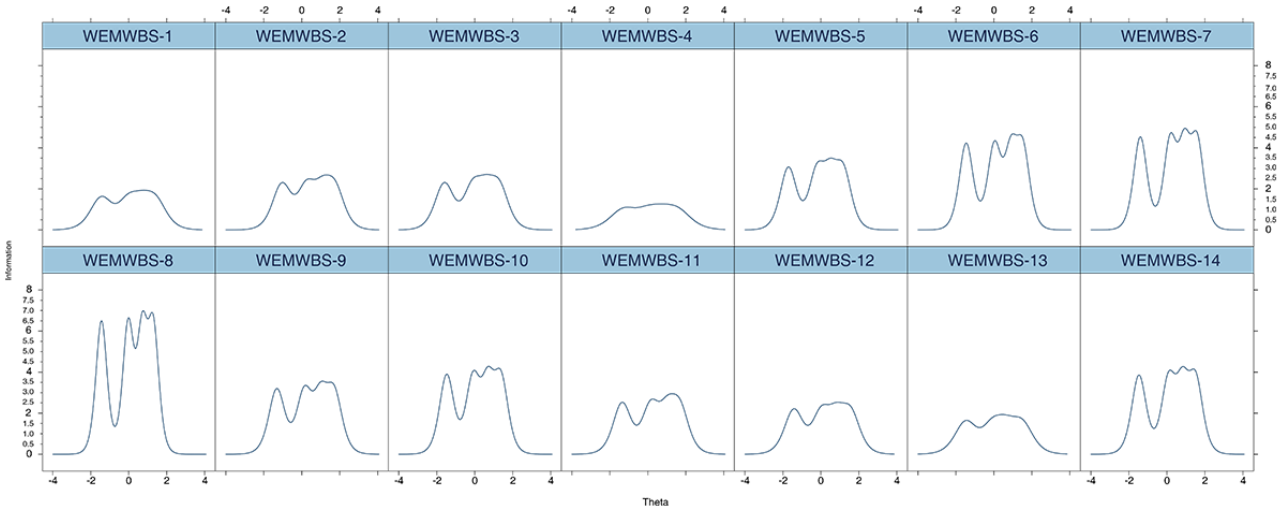
The ICCs and IICs for the Chinese WEMWBS are shown in Figures 2 and 3, respectively. The ICCs demonstrated that the sequence of the categories' thresholds for the 14 items was as predicted, meaning that all regimentations were sufficient in including respondents; this finding, in turn, suggests that all categories were adequate based on placing a participant on the scale. The IICs displayed multimodal distribution. The shape of item 8 was the most precipitous and provided more knowledge than the other 13 items. The shape of item 4 was the flattest, indicating that the item provided the least information.

Figure 2. Item-category characteristic curves for the Chinese-language version of the WEMWBS. The numbers indicate each item on the scale. WEMWBS: Warwick-Edinburgh Mental Well-being Scale.



XSL•FO
RenderX

Figure 3. The item information curves for items of the Chinese-language version of the WEMWBS. The numbers indicate each item on the scale. WEMWBS: Warwick-Edinburgh Mental Well-being Scale.



Discussion

Principal Findings

This is the first study to combine CTT and a GRM incorporating IRT to evaluate psychometric properties of the Chinese-language version of the WEMWBS in a sample of medical staff. Our results confirm the initial hypothesis that the WEMWBS is 1D. Since its establishment in 2006, the WEMWBS has been used in trials of patients and the general population with commendable results according to CTT and the Rasch model [32,33]. Given the broad and complicated spectrum of psychometric processes other than CTT, each with new evaluations and fixed statistical analyses in diverse models [34], we adopted the GRM to evaluate the contribution of the 14 items and their responses to the assessment of subjective well-being (SWB).

Comparisons With Previous Studies

The mean score for the Chinese version of the WEMWBS used in this study was 38.47 (SD 13.23), which is lower than WEMWBS scores in medical staff surveys in other countries (eg, the United Kingdom [35], Pakistan [25,36], and Northern Ireland [37]). This discrepancy may be due to the data having been collected during the outbreak of COVID-19, meaning that the SWB of the medical staff would have been impacted to a certain extent [38]. Moreover, with the aging population of China, medical staff are under a great deal of pressure and need to master multidisciplinary knowledge and skills even as their work intensity increases [39].

The original 1D structure of the WEMWBS, as confirmed by previous studies in other countries [24,27,29], was not fully supported by earlier research in mainland China. This outcome was expected; some studies [28,40] identified a 2D structure that differed from the original assumption.

Researchers have pointed to differences between Eastern and Western cultures to explain this: the original meaning of the individual items might be changed in translated versions, and this alteration could affect the perceived intentions of the target population [22]. Furthermore, previous studies [33] adopted the Likert ordinal interval for a comprehensive rating, in which the 14 individual item scores were added to produce a total score. Bartram [35] found that using only a CFA may lead to misunderstanding, because the total score has a serial order, and the intervals between each score are not necessarily equal. The unidimensional structure was not without problems in this study. First, the model fitting effect was insufficient, because the χ^2/df was greater than 5, and the RMSEA was greater than 0.08. Only the NFI, TLI, and CFI values supported the unidimensionality of the model. However, the AVE was greater than 0.5, and the CR was greater than 0.8, suggesting a relevant result. Second, the 1D model's factor loadings for the 14 items were similar to the 2-factor model. Third, considering that the number of factors according to the revised MAP test was 1 [31], we adopted the 1D structure. An exemplary configuration of the Chinese WEMWBS would be favorable for facilitating IRT analyses in the future. Administering the Chinese WEMWBS based on IRT could strengthen its sensitivity and precision, guaranteeing that the items reflect the participants' SWB levels.

The proportion of participants selecting the options "sometimes" and "often" was high in this study, suggesting that most respondents had relatively good SWB. To test the accuracy of the results, we examined the 14 items for floor and ceiling effects; we did not find extreme ceiling or floor effects, indicating that the process was reliable. There have been no reports on the distribution of responses to the WEMWBS in mainland China. In addition, the Chinese version of the WEMWBS displayed outstanding reliability, with a Cronbach α of .96, more significant than other studies for Chinese and other language versions [18,19,28,29,41].

The GRM was the best-match IRT model in this study. No previous studies have used the GRM to evaluate the psychometric properties of the WEMWBS. Our study reinforces the use of IRT models and supports existing studies on the psychometric evaluation of the WEMWBS with IRT methods.

The GRM analysis demonstrated that the global performance of WEMWBS items was satisfactory. The ICCs showed that the feedback categories of all the items were ordered and that

XSL•FO
RenderX

all categories were presumably at the same point on the continuum [42].

Prospects for Application of the Chinese WEMWBS

Mental health assessment has drawn increasing attention from the Chinese government. In 2017, the Chinese government released the first guidelines to improve mental health in schools, workplaces, and hospitals. The WEMWBS has proven to be a convenient and valuable psychometric tool for academics, medical professionals, and other prominent stakeholders to measure the SWB of medical staff [43,44]. The Chinese WEMWBS has good reliability and validity with comprehensive and understandable content [15,24,26,45].

Limitations

There are some limitations to this study. First, our investigation concentrated on hospitals in Zhejiang and Hunan provinces, and most participants were nurses, suggesting some selection bias. Follow-up research needs a larger sample size that includes therapists, physicians, and surgeons to assess the psychometric properties of the Chinese WEMWBS. Second, the sample size was only 572, which is less than 1000; this may have caused ambiguity in evaluating the IRT model. A larger sample size is needed in future research to confirm our findings. Third, we did not discriminate between medical staff with anxiety or depression when calculating the psychometric properties of the Chinese WEMWBS, which may have caused difficulty in demonstrating the scale's validity. The performance of the Chinese WEMWBS should be further assessed in distinct staff groups.

Conclusion

Detailed provisions were made for the Chinese version of the WEMWBS in this study, and its psychometric properties were evaluated in a group of medical staff. We found that the Chinese WEMWBS has good reliability and validity and that it could be used as a reliable tool to evaluate the SWB of medical staff. It is critical to adopt measures to enable decision-making departments of hospitals to reduce work pressure, improve the SWB of clinical medical staff, improve patient satisfaction, and promote the development of the medical industry in a favorable direction.

Acknowledgments

This work was supported by the Wenzhou Science and Technology Bureau, Wenzhou, China (Y2019038), and the Education Department of Zhejiang Province, Zhejiang, China (Y202147054), which had no role in the study design; collection, analysis, or interpretation of the data; writing the manuscript; or the decision to submit the paper for publication. The table of contents image was obtained from the Freepik image bank.

Authors' Contributions

All authors provided scientific input and edited and reviewed the manuscript content. All authors provided their final approval and agreed to be accountable for all aspects of the work, ensuring integrity and accuracy. WG was responsible for the manuscript; AD and JH wrote the manuscript; SL completed the data analysis; JZ tabulated the data; other authors collected case materials. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Diener E, Oishi S, Tay L. Advances in subjective well-being research. Nat Hum Behav 2018 Apr 12;2(4):253-260. [doi: 10.1038/s41562-018-0307-6] [Medline: 30936533]
2. Melnyk BM, Kelly SA, Stephens J, Dhakal K, McGovern C, Tucker S, et al. interventions to improve mental health, well-being, physical health, and lifestyle behaviors in physicians and nurses: a systematic review. Am J Health Promot 2020 Nov 27;34(8):929-941 [FREE Full text] [doi: 10.1177/0890117120920451] [Medline: 32338522]
3. O'Reilly M. Social media and adolescent mental health: the good, the bad and the ugly. J Ment Health 2020 Apr 28;29(2):200-206. [doi: 10.1080/09638237.2020.1714007] [Medline: 31989847]
4. Sim F, Mackie P. Well-being of the migrant workforce: who cares? Public Health 2018 Jul;160:A1-A2. [doi: 10.1016/j.puhe.2018.06.005] [Medline: 29914732]
5. Søvold LE, Naslund JA, Kousoulis AA, Saxena S, Qoronfleh MW, Grobler C, et al. Prioritizing the mental health and well-being of healthcare workers: an urgent global public health priority. Front Public Health 2021 May 7;9:679397 [FREE Full text] [doi: 10.3389/fpubh.2021.679397] [Medline: 34026720]
6. Wainberg ML, Scorza P, Shultz JM, Helpman L, Mootz JJ, Johnson KA, et al. Challenges and opportunities in global mental health: a research-to-practice perspective. Curr Psychiatry Rep 2017 May;19(5):28 [FREE Full text] [doi: 10.1007/s11920-017-0780-z] [Medline: 28425023]
7. Salazar de Pablo G, Vaquerizo-Serrano J, Catalan A, Arango C, Moreno C, Ferre F, et al. Impact of coronavirus syndromes on physical and mental health of health care workers: systematic review and meta-analysis. J Affect Disord 2020 Oct 01;275:48-57 [FREE Full text] [doi: 10.1016/j.jad.2020.06.022] [Medline: 32658823]

XSL•FO
RenderX

8. Berlanda S, de Cordova F, Fraizzoli M, Pedrazza M. Risk and protective factors of well-being among healthcare staff. A thematic analysis. Int J Environ Res Public Health 2020 Sep 12;17(18):6651 [FREE Full text] [doi: 10.3390/ijerph17186651] [Medline: 32932588]
9. Ruiz-Fernández MD, Pérez-García E, Ortega-Galán ÁM. Quality of life in nursing professionals: burnout, fatigue, and compassion satisfaction. Int J Environ Res Public Health 2020 Feb 15;17(4):1253 [FREE Full text] [doi: 10.3390/ijerph17041253] [Medline: 32075252]
10. Zhou C, Shi L, Gao L, Liu W, Chen Z, Tong X, et al. Determinate factors of mental health status in Chinese medical staff: A cross-sectional study. Medicine (Baltimore) 2018 Mar;97(10):e0113 [FREE Full text] [doi: 10.1097/MD.0000000000010113] [Medline: 29517690]
11. Shi Y, Xue H, Ma Y, Wang L, Gao T, Shi L, et al. Prevalence of occupational exposure and its influence on job satisfaction among Chinese healthcare workers: a large-sample, cross-sectional study. BMJ Open 2020 Apr 16;10(4):e031953 [FREE Full text] [doi: 10.1136/bmjopen-2019-031953] [Medline: 32303512]
12. Ledikwe JH, Kleinman NJ, Mpho M, Mothibedi H, Mawandia S, Semo B, et al. Associations between healthcare worker participation in workplace wellness activities and job satisfaction, occupational stress and burnout: a cross-sectional study in Botswana. BMJ Open 2018 Mar 16;8(3):e018492 [FREE Full text] [doi: 10.1136/bmjopen-2017-018492] [Medline: 29549200]
13. Yakes EA, Dean S, Labadie RF, Byrne D, Estrada C, Thompson R, et al. Factors associated with physician empowerment and well-being at an academic medical center. J Occup Environ Med 2020 Jul;62(7):478-483. [doi: 10.1097/JOM.0000000000001874] [Medline: 32730023]
14. Zhou H, Jiang F, Rakofsky J, Hu L, Liu T, Wu S, et al. Job satisfaction and associated factors among psychiatric nurses in tertiary psychiatric hospitals: Results from a nationwide cross-sectional study. J Adv Nurs 2019 Dec;75(12):3619-3630. [doi: 10.1111/jan.14202] [Medline: 31566793]
15. Dong A, Chen X, Zhu L, Shi L, Cai Y, Shi B, et al. Translation and validation of a Chinese version of the Warwick-Edinburgh Mental Well-being Scale with undergraduate nursing trainees. J Psychiatr Ment Health Nurs 2016 Nov 18;23(9-10):554-560. [doi: 10.1111/jpm.12344] [Medline: 27860080]
16. Konaszewski K, Niesiobędzka M, Surzykiewicz J. Factor structure and psychometric properties of a Polish adaptation of the Warwick-Edinburgh Mental Wellbeing Scale. Health Qual Life Outcomes 2021 Mar 02;19(1):70 [FREE Full text] [doi: 10.1186/s12955-021-01716-w] [Medline: 33653357]
17. Lang G, Bachinger A. Validation of the German Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a community-based sample of adults in Austria: a bi-factor modelling approach. J Public Health 2016 Dec 3;25(2):135-146. [doi: 10.1007/s10389-016-0778-8]
18. Trousselard M, Steiler D, Dutheil F, Claverie D, Canini F, Fenouillet F, et al. Validation of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in French psychiatric and general populations. Psychiatry Res 2016 Nov 30;245:282-290. [doi: 10.1016/j.psychres.2016.08.050] [Medline: 27565700]
19. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
20. Liddell C, Guiney C. Living in a cold and damp home: frameworks for understanding impacts on mental well-being. Public Health 2015 Mar;129(3):191-199. [doi: 10.1016/j.puhe.2014.11.007] [Medline: 25726123]
21. Tamminen N, Solin P, Stengård E, Kannas L, Kettunen T. Mental health promotion competencies in the health sector in Finland: a qualitative study of the views of professionals. Scand J Public Health 2019 Mar 12;47(2):115-120. [doi: 10.1177/1403494817711360] [Medline: 28699393]
22. Dong A, Zhang X, Zhou H, Chen S, Zhao W, Wu M, et al. Applicability and cross-cultural validation of the Chinese version of the Warwick-Edinburgh mental well-being scale in patients with chronic heart failure. Health Qual Life Outcomes 2019 Mar 29;17(1):55 [FREE Full text] [doi: 10.1186/s12955-019-1120-2] [Medline: 30922342]
23. Liu Y, Guo L, Liu K. Validity and reliability of Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in older people. Chin Ment Health J 2016;30(3):174-178. [doi: 10.3969/j.issn.1000-6729.2016.03.004]
24. Fung S. Psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) with Chinese university students. Health Qual Life Outcomes 2019 Mar 14;17(1):46 [FREE Full text] [doi: 10.1186/s12955-019-1113-1] [Medline: 30871563]
25. Waqas A, Ahmad W, Haddad M, Taggart FM, Muhammad Z, Bukhari MH, et al. Measuring the well-being of health care professionals in the Punjab: a psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale in a Pakistani population. PeerJ 2015;3:e1264 [FREE Full text] [doi: 10.7717/peerj.1264] [Medline: 26557420]
26. Taggart F, Friede T, Weich S, Clarke A, Johnson M, Stewart-Brown S. Cross cultural evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) --a mixed methods study. Health Qual Life Outcomes 2013 Feb 27;11:27 [FREE Full text] [doi: 10.1186/1477-7525-11-27] [Medline: 23445544]
27. Bass M, Dawkin M, Muncer S, Vigurs S, Bostock J. Validation of Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in a population of people using Secondary Care Mental Health Services. J Ment Health 2016 Aug 28;25(4):323-329. [doi: 10.3109/09638237.2015.1124401] [Medline: 26821730]

XSL•FO
RenderX

28. Clarke A, Friede T, Putz R, Ashdown J, Martin S, Blake A, et al. Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment. BMC Public Health 2011 Jun 21;11:487 [FREE Full text] [doi: 10.1186/1471-2458-11-487] [Medline: 21693055]
29. Smith ORF, Alves DE, Knapstad M, Haug E, Aarø LE. Measuring mental well-being in Norway: validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS). BMC Psychiatry 2017 May 12;17(1):182 [FREE Full text] [doi: 10.1186/s12888-017-1343-x] [Medline: 28499368]
30. Yamada A, Kasahara K, Ogawa Y, Samejima K, Eriguchi M, Yano H, et al. Peritonitis due to Moraxella osloensis: A case report and literature review. J Infect Chemother 2019 Dec;25(12):1050-1052. [doi: 10.1016/j.jiac.2019.05.018] [Medline: 31196771]
31. Velicer WF, Eaton CA, Fava JL. Construct explication through factor or component analysis: a review and evaluation of alternative procedures for determining the number of factors or components. In: Goffin RD, Helmes E, editors. Problem and Solutions in Human Assessment. Boston, MA: Springer; 2000.
32. Khumalo IP, Temane QM, Wissing MP. Further validation of the General Psychological Well-being Scale among a Setswana-speaking group. In: Wissing M, editor. Well-Being Research in South Africa. Dordrecht, Netherlands: Springer; 2013:199-224.
33. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes 2009 Feb 19;7:15 [FREE Full text] [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
34. Bilder RM, Reise SP. Neuropsychological tests of the future: How do we get there from here? Clin Neuropsychol 2019 Feb 13;33(2):220-245 [FREE Full text] [doi: 10.1080/13854046.2018.1521993] [Medline: 30422045]
35. Bartram DJ, Sinclair JM, Baldwin DS. Further validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in the UK veterinary profession: Rasch analysis. Qual Life Res 2013 Mar 2;22(2):379-391. [doi: 10.1007/s11136-012-0144-4] [Medline: 22383106]
36. Ahmad W, Taggart F, Shafique MS, Muzafar Y, Abidi S, Ghani N, et al. Diet, exercise and mental-wellbeing of healthcare professionals (doctors, dentists and nurses) in Pakistan. PeerJ 2015;3:e1250 [FREE Full text] [doi: 10.7717/peerj.1250] [Medline: 26401460]
37. Murray MA, Cardwell C, Donnelly M. GPs' mental wellbeing and psychological resources: a cross-sectional survey. Br J Gen Pract 2017 Jul 17;67(661):e547-e554. [doi: 10.3399/bjgp17x691709]
38. Hu Z, Lin X, Chiwanda Kaminga A, Xu H. Impact of the COVID-19 epidemic on lifestyle behaviors and their association with subjective well-being among the general population in mainland China: cross-sectional study. J Med Internet Res 2020 Aug 25;22(8):e21176 [FREE Full text] [doi: 10.2196/21176] [Medline: 32759103]
39. Du J, Mayer G, Hummel S, Oetjen N, Gronewold N, Zafar A, et al. Mental health burden in different professions during the final stage of the COVID-19 lockdown in China: cross-sectional survey study. J Med Internet Res 2020 Dec 02;22(12):e24240 [FREE Full text] [doi: 10.2196/24240] [Medline: 33197231]
40. López MA, Gabilondo A, Codony M, García-Forero C, Vilagut G, Castellví P, et al. Adaptation into Spanish of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) and preliminary validation in a student sample. Qual Life Res 2013 Jun 27;22(5):1099-1104. [doi: 10.1007/s11136-012-0238-z] [Medline: 22836376]
41. Mavali S, Mahmoodi H, Sarbakhsh P, Shaghaghi A. Psychometric properties of the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) in the Iranian older adults. Psychol Res Behav Manag 2020;13:693-700 [FREE Full text] [doi: 10.2147/PRBM.S256323] [Medline: 32884372]
42. Chorpita BF, Daleiden EL, Weisz JR. Identifying and selecting the common elements of evidence based interventions: a distillation and matching model. Ment Health Serv Res 2005 Mar;7(1):5-20. [doi: 10.1007/s11020-005-1962-6] [Medline: 15832690]
43. Adamou M, Goddard A, Kyriakidou N, Mooney A, O'Donoghue D, Pattani S, et al. The wellbeing thermometer: a novel framework for measuring wellbeing. Psychology 2020;11(10):1471-1480 [FREE Full text] [doi: 10.4236/psych.2020.1110093]
44. Oates J, Jones J, Drey N. Subjective well-being of mental health nurses in the United Kingdom: results of an online survey. Int J Ment Health Nurs 2017 Aug 23;26(4):391-401. [doi: 10.1111/inm.12263] [Medline: 27878917]
45. Ng SSW, Lo AWY, Leung TKS, Chan FSM, Wong ATY, Lam RWT, et al. Translation and validation of the Chinese version of the short Warwick-Edinburgh Mental Well-being Scale for patients with mental illness in Hong Kong. East Asian Arch Psychiatry 2014 Mar;24(1):3-9 [FREE Full text] [Medline: 24676481]

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fitting index
CR: composite reliability
CTT: classical test theory
GRM: graded response model

XSL•FO
RenderX

ICC: item characteristic curve
IFI: incremental fit index
IIC: item information curve
IRT: item response theory
MAP: mean average precision
NFI: normed fit index
QoL: quality of life
RFI: relative fit index
RMSEA: root mean square error of approximation
SF-36 v2: 36-Item Short Form Health Survey, Version 2
SWB: subjective well-being
TLI: Tacker-Lewis index
WEMWBS: Warwick-Edinburgh Mental Well-being Scale
WHO-5: 5-item World Health Organization Well-Being Index

Edited by R Kukafka; submitted 20.03.22; peer-reviewed by S Stewart-Brown, A Joseph, S Pesälä; comments to author 25.05.22; revised version received 07.07.22; accepted 01.11.22; published 30.11.22.

Please cite as:
Dong A, Huang J, Lin S, Zhu J, Zhou H, Jin Q, Zhao W, Zhu L, Guo W
Psychometric Properties of the Chinese Warwick-Edinburgh Mental Well-being Scale in Medical Staff: Cross-sectional Study
J Med Internet Res 2022;24(11):e38108
URL: https://www.jmir.org/2022/11/e38108
doi:10.2196/38108
PMID:36449336

©Aishu Dong, Jing Huang, Shudan Lin, Jianing Zhu, Haitao Zhou, Qianqian Jin, Wei Zhao, Lianlian Zhu, Wenjian Guo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Text Topics and Treatment Response in Internet-Delivered Cognitive Behavioral Therapy for Generalized Anxiety Disorder: Text Mining Study

Sanna Mylläri[1], MA; Suoma Eeva Saarni[2], PhD; Ville Ritola[2], MA; Grigori Joffe[2], MD, PhD; Jan-Henry Stenberg[2], PhD; Ole André Solbakken[3], PhD; Nikolai Olavi Czajkowski[3,4], PhD; Tom Rosenström[1], PhD

[1]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[2]Department of Psychiatry, Brain Center, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

[3]Department of Psychology, University of Oslo, Oslo, Norway

[4]Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway

Corresponding Author:
Sanna Mylläri, MA
Department of Psychology and Logopedics
Faculty of Medicine
University of Helsinki
PO Box 21, Haartmaninkatu 3 E, H218
Helsinki, 00014
Finland
Phone: 358 407678891
Email: sanna.myllari@helsinki.fi

Abstract

Background: Text mining methods such as topic modeling can offer valuable information on how and to whom internet-delivered cognitive behavioral therapies (iCBT) work. Although iCBT treatments provide convenient data for topic modeling, it has rarely been used in this context.

Objective: Our aims were to apply topic modeling to written assignment texts from iCBT for generalized anxiety disorder and explore the resulting topics' associations with treatment response. As predetermining the number of topics presents a considerable challenge in topic modeling, we also aimed to explore a novel method for topic number selection.

Methods: We defined 2 latent Dirichlet allocation (LDA) topic models using a novel data-driven and a more commonly used interpretability-based topic number selection approaches. We used multilevel models to associate the topics with continuous-valued treatment response, defined as the rate of per-session change in GAD-7 sum scores throughout the treatment.

Results: Our analyses included 1686 patients. We observed 2 topics that were associated with better than average treatment response: "well-being of family, pets, and loved ones" from the data-driven LDA model (B=–0.10 SD/session/Δtopic; 95% CI –016 to –0.03) and "children, family issues" from the interpretability-based model (B=–0.18 SD/session/Δtopic; 95% CI –0.31 to –0.05). Two topics were associated with worse treatment response: "monitoring of thoughts and worries" from the data-driven model (B=0.06 SD/session/Δtopic; 95% CI 0.01 to 0.11) and "internet therapy" from the interpretability-based model (B=0.27 SD/session/Δtopic; 95% CI 0.07 to 0.46).

Conclusions: The 2 LDA models were different in terms of their interpretability and broadness of topics but both contained topics that were associated with treatment response in an interpretable manner. Our work demonstrates that topic modeling is well suited for iCBT research and has potential to expose clinically relevant information in vast text data.

(J Med Internet Res 2022;24(11):e38911) doi:10.2196/38911

KEYWORDS

iCBT; CBT; psychotherapy; internet therapy; anxiety; topic modeling; natural language processing

XSL•FO
RenderX

Introduction

Internet-delivered cognitive behavioral therapy (iCBT) is an effective treatment for generalized anxiety disorder (GAD) [1-4]. Additionally, iCBT programs typically store data automatically, which is convenient in terms of computerized text analysis methods, or text mining. Such methods can vastly extend the scale of traditional human-based content analysis [5]. Together with increasing data availability, text mining provides opportunities for treatment personalization and may reveal mechanisms of or obstacles to behavior change. For example, a previous study analyzed texts written during an iCBT for GAD, demonstrating a covariation between negative emotion words and symptom change over the course of treatment [6].

Many of the previous studies that have used computerized methods to analyze psychotherapy texts have relied on predetermined word categories in text classification [7-9]. Word categorization tools developed for the purposes of psychological research are theory-driven and easy to interpret [10]. Nonetheless, approaches that are more data-driven might reveal textual aspects not considered in theory-driven categorizations. An example of a data-driven approach applicable to iCBT research is topic modeling.

Topic models, such as the latent Dirichlet allocation (LDA), are unsupervised machine learning models that reduce data dimensionality by expressing a text as a mixture of latent topics [11]. LDA has been successful in detecting meaningful topics that occur in face-to-face psychotherapy transcripts [12-14]. A recent study that applied LDA to psychotherapy session transcripts found a covariation between topics with descriptions of positive experiences and a symptom decrease whereas topics that reflected discussion about treatment were associated with a symptom increase [13]. The latter also predicted alliance rupture in therapy sessions. Thus, LDA may reveal information on how to tailor interventions and improve treatment outcomes.

Previous psychotherapy topic modeling studies have used text data from whole therapy sessions with relatively free-flowing speech. This type of data is rich and has potential to reveal a wide spectrum of contents in language use during the psychotherapy process. We argue, however, that the more structured iCBT data has some benefits. First, data with a spectrum of contents that is too wide may not be ideal in terms of exploratory statistical analysis due to a phenomenon known as "the statistical curse of dimensionality: If the data have a dimension *d*, then we need a sample size *n* that grows exponentially with *d*" [15]. Simply put, rare word combinations need very large data sets to occur frequently enough for statistical estimation, and with an increasing number of utterances, most combinations get rare. Naturalistic iCBT data accrue rapidly and pertain to a comparatively narrow language space as iCBTs generally consist of standardized assignments. Second, focusing on assignment-specific data may increase the currently lacking and sought-after understanding of the meaning of specific therapeutic components [16,17]. Compared with more traditional component studies, topic modeling assignment texts could reveal benefits or harms specific to some individuals or contents that are missed in group-level comparisons in randomized controlled trials. From a topic modeling perspective, focusing on texts derived from one predetermined therapy assignment as opposed to many helps to avoid the topic model picking up task-related variation in the language use. As the meaning of language use could be context-specific, use of one assignment should also serve the aim of finding interpretable topic-outcome associations with practical implications. Despite these beneficial aspects, iCBT data have not yet been widely used for topic modeling (for an exception, see the study by Hoogendoorn et al [18]).

Regardless of the specific application context, estimating an LDA topic model requires an analyst to specify the number of latent topics in the model. This poses a challenge when using large naturalistic data sets such as iCBT texts where it is rarely possible to predetermine the distinct semantic contents in the data. Selection of the number of topics needs to be performed with care, as too few and too many topics can both affect the reliability of LDA model estimation [19]. Previous psychotherapy studies, for example a study by Atzil-Slonim et al [13], have used heuristic methods to select topic number. This may lead to suboptimal models containing idiosyncratic topics, which in turn can reduce the comparability and performance of topic models in psychotherapy research. However, the optimal strategy for topic number selection remains an unresolved challenge in topic modeling literature. New, promising, fully data-driven methods for topic-number selection are emerging, and here we examined their potential in iCBT topic modeling [19].

In this paper, we applied topic modeling to a large, naturalistic set of text data from iCBT for GAD that is offered as a part of public health care in Finland [2]. As a central element in GAD is worry, we focused on worry diary task sheets that contain patients' descriptions of their worrisome thoughts. The worry diary was introduced at the early stages of treatment and carried on throughout the treatment, thus offering a good representation of the patient population's writing behaviors. As a preliminary analysis, we examined whether worry diary writing activity was associated with treatment response. Our aims were to (1) explore topic modeling in iCBT data, focusing specifically on defining the optimal number of topics, and (2) investigate associations between found topics and treatment response. We expected to find meaningful topics associated with treatment response in an interpretable way. Our findings should be useful when designing optimal psychotherapy programs and instructions for worry-diary tasks and potentially when predicting who will benefit from these tasks.

Methods

Data

Participants

The data were obtained as part of routine care from the therapist-assisted iCBT for GAD, manufactured and delivered by the HUS Helsinki University Hospital (HUS-iCBT). HUS-iCBT for GAD is a standardized treatment consisting of 12 weekly sessions and a follow-up session 3 months after treatment completion. The treatment is part of the Finnish public

XSL•FO
RenderX

specialized mental health care and targets adult patients and minors aged 16 years and older with mild to moderately severe symptoms. The exclusion criteria are suicidality, acute psychosis, serious personality disorder, and neurological or neuropsychiatric disorders that affect cognitive functioning. For a more detailed description of the treatment, see Ritola et al [2].

The original data set consisted of 2218 patients who had entered and completed or dropped out of the treatment between January 2015 and September 2019. As we were interested in the actual observed per-session symptom decline, we used all available data efficiently in multilevel models (see the section Treatment Response Models). Our goal was to model the symptom change of both completers and dropouts in a naturalistic manner, and therefore we did not impute any missing data. As the patient was required to complete the symptom questionnaires to proceed within each session, complete symptom data were available from all the sessions that each patient completed.

Text Data

Our text data were drawn from a worry diary task sheet, used as a part of 3 different between-session assignments throughout the treatment. The assignments were (1) simple worry diary, where the patients write observations about their worries and related behaviors, (2) worry postponement by writing in the worry diary within a certain time frame during the day, and (3) practicing problem-solving skills. Patients were not required to complete the between-session assignments to proceed in the treatment, and they were free to use the task sheet as often as they wanted. For a more detailed description of the worry diary, see Multimedia Appendix 1.

Outcome Measure

At the beginning of each session, the patient's anxiety symptoms were assessed using the Generalized Anxiety Disorder 7-item scale (GAD-7) [20]. The GAD-7 sum score is a suitable measure for symptom severity with good temporal measurement invariance [21,22]. We defined continuous-valued treatment response as the rate of per-session change in the GAD-7 sum scores throughout the treatment. The exclusion criterion for the study was a baseline GAD-7 score of less than 8, which is a recommended cutoff point for GAD screening [23].

Ethics Approval

This study is a part of a research project that has received permission from the ethical board of HUS Helsinki University Hospital to use the data (approval number HUS/1861/2020).

Topic Modeling

Text Preprocessing

Our text corpus consisted of the worry diary task sheet entries (Multimedia Appendix 1). We preprocessed the texts by tokenizing and stemming the words as well as removing punctuation and common stop words (common words with little meaning such as *and* or *it*). The preprocessing was performed using the R package Corpus [24]. The original data contained a limited number of entries written in English or Swedish, which were removed. For an illustrative example of data preprocessing, see Multimedia Appendix 1.

Latent Dirichlet Allocation

We used LDA for the topic modeling of worry diary entries [11]. LDA is a widely used probabilistic model that represents each text document as a mixture of latent topics, whereas each latent topic is defined by a distribution over the words in the corpus [25]. In our data, the corpus is the whole data set of worry diary entries, whereas each entry written by a patient is a document. Each worry diary entry i is given an estimate θ_{ik} that represents the probability of topic k occurring in that entry. This numeric representation of latent topics in diary entries can then be used to associate writing about each topic with the treatment outcome. We used the R package textmineR to compute the LDA models [26]. For technical details and an illustrative example of the LDA model, see Multimedia Appendix 1.

Selection of Number of Topics

As noted, the selection of the number of topics k is important because it affects a reliable estimation of the posterior LDA distribution and thus the generalizability of observed associations with treatment response. We aimed to solve the problem by using a Bayesian approach, where the data dictate the desirable parameter value, as suggested by the original work by Chen and Doss [19]. For a more detailed description of the topic number selection procedure, see Multimedia Appendix 1. Essentially, the procedure controls for overfitting to data according to Bayesian model selection principles.

The amount of available data can affect how a complex model—how large a value for k—is found using a data-driven approach. Therefore, and to better understand the effects of the choice for k, we formed an additional LDA model using a heuristic selection process that emphasizes the interpretability of the resulting topics. That is, we aimed for topics that were semantically coherent, distinguishable from each other, and easy to identify from the texts. In this approach, we ran LDA models with k starting from 10, increasing it by intervals of 5 until the interpretability no longer continued to improve when more topics were added.

Correlates of Treatment Response

Data Set for Modeling

The patients who had zero worry diary entries were not included in the LDA modeling corpus. To model the full range of writing activity, we included these patients in the multilevel modeling data set. For the interpretability of the treatment response effect sizes, we standardized the GAD-7 sum scores according to the baseline GAD-7 measurements. To facilitate the interpretation of the model intercepts as anxiety at the beginning of the treatment, we set the running number of therapy sessions to start from 0.

To model writing activity, we defined 4 variables for the number of worry diary entries. The first one was the total number of entries throughout the treatment, labeled as total entries. We then divided the number of entries according to the different worry diary task assignments and created 3 additional variables for the number of entries: entries 1 (worry diary), entries 2 (worry postponement), and entries 3 (problem solving).

XSL•FO
RenderX

To use the topics from the LDA models as correlates of treatment response, we assessed the average occurrence of a topic for each patient by calculating the mean value of the LDA model's topic probability parameter θ over the patient's worry diary entries. For those patients with no entries, the occurrence of each topic was set at 0.

Treatment Response Models

We defined 2 baseline treatment response models by including the session number as a fixed-effect covariate and the within-patient time-average level of anxiety as a random intercept [27]. Model 0 included random intercept only, whereas model 1 also included a within-patient random slope. Both models were adjusted for age and sex. We compared model 0 and model 1 using a likelihood ratio test, and the model with better fit was selected as the base model for additional correlates. For assessment of the treatment-response moderator effect, all following models included a correlate-by-session interaction.

We estimated the association of the worry diary writing activity with treatment response using 2 separate models. Model 2 included the total entries as a fixed-effect correlate. In model 3, we used the 3 other entry variables as fixed-effect correlates to estimate the effects of the different worry diary task assignments.

We then estimated the association of the topics from the 2 LDA models with treatment response. Each topic was separately added as a fixed-effect correlate to the base model (model 4). We then adjusted the models for the 3 entry variables to separate the independent association of a topic with treatment response from its association with writing activity under different assignments (model 5). Finally, the topics with a significant treatment-response effect were additionally adjusted with other significant topics within the same LDA model to account for potential confounding effects of topics with each other (model 6). All analyses were conducted using R (version 3.6.3, R Foundation for Statistical Computing) [28]. R code equations for the models are presented in Multimedia Appendix 1.

Results

Patient Characteristics

Table 1 presents the descriptive characteristics of our data. After data preprocessing, the final LDA modeling corpus consisted of 11,897 worry diary entries. In the multilevel modeling sample, the number of diary entries per patient varied between 0 and 97. Those patients with 0 diary entries were on average younger, had finished fewer sessions, were much less likely to complete all treatment sessions, and were more likely men.

Table 1. Baseline characteristics after data preprocessing.

	Multilevel modeling data set (n=1686[a])	LDA[b] modeling corpus (n=1448)	Patients with 0 entries (n=239)
Gender, n (%)			
Female	1322 (78)	1155 (80)	165 (69)
Male	364 (22)	239 (20)	74 (31)
Age (years), mean (SD)	33.2 (12.0)	33.5 (12.0)	31.7 (11.7)
Number of finished sessions, mean (SD)	7.8 (4.4)	8.6 (4.0)	3.2 (3.2)
Finished all 12 treatment sessions, n (%)	730 (43)	712 (49)	16 (0.1)
Number of diary entries, mean (SD)	7.1 (8.7)	8.2 (8.9)	0
Entries 1 (worry diary)	4.7 (5.3)	5.5 (5.4)	0
Entries 2 (worry postponement)	1.6 (4.2)	1.9 (4.5)	0
Entries 3 (problem solving)	0.7 (1.6)	0.9 (1.7)	0
GAD-7[c] at beginning of treatment, mean (SD)	13.1 (3.6)	13.1 (3.6)	13.2 (3.4)

[a]Patients with missing background data (n=11) or GAD-7 score < 8 (n=521) were excluded.

[b]LDA: Latent Dirichlet allocation.

[c]GAD-7: Generalized Anxiety Disorder–7 item.

LDA Models

We selected 7 as the optimal number of topics for the data-driven model using the Bayesian approach for topic number selection (Multimedia Appendix 1, Figure S3 and Table S1). For the interpretability-based model, we selected 25 as the optimal number of topics. For descriptions of the full LDA models, see Multimedia Appendix 1.

Continuous-Valued Treatment Response Model

There was heterogeneity in symptom trajectories between patients as indicated by the better fit of model 1 with a random slope compared with model 0 (random intercept only; χ^2_2=760.17; P<.001). Thus, model 1 was selected as a base model for additional correlates. In model 1, there was a significant association between session number and anxiety symptoms (B=–0.14; 95% CI –0.15 to –0.13). That is, the GAD-7 score declined on average 0.14 standard deviations in each treatment session. The variability between patient symptom

XSL•FO
RenderX

trajectories was large, as indicated by the random slope standard deviation of 0.1.

Writing Activity as a Correlate of Continuous-Valued Treatment Response

In model 2, the total number of worry diary entries had a significant treatment-response moderator effect (B=0.001; 95% CI 0.000 to 0.002; for session-by-entry interaction). The effect size was modest: an increase of 1 entry in the total number of entries was associated with a 0.001 standard deviation slower than average decline in anxiety. In model 3, only the number of entries written during later phases of treatment remained significant treatment-response moderators (task assignments worry postponement and problem solving; Table 2). A larger number of entries written during the first task assignment was associated with on average more severe baseline anxiety, as reflected by the baseline effect of entries 1 on anxiety (Table 2).

Table 2. A multilevel regression model associating continuous-valued treatment response with writing activity during different worry diary task assignments (n=1686). Number of observations (GAD-7 measurements)=13,205.

Effect	Estimate[a]	95% CI	*P* value
Fixed effects			
Intercept	−0.087	−0.260 to 0.087	.33
Session number	−0.147	−0.158 to −0.135	<.001
Treatment moderation effect, entries 1 (change/session)[b]	−0.001	−0.002 to 0.001	.23
Treatment moderation effect, entries 2 (change/session)	0.002	0.000 to 0.003	.03
Treatment moderation effect, entries 3 (change/session)	0.005	0.001 to 0.010	.02
Baseline effect, entries 1	0.014	0.004 to 0.023	.006
Baseline effect, entries 2	−0.003	−0.017 to 0.010	.61
Baseline effect, entries 3	−0.031	−0.066 to 0.004	.08
Age	−0.009	−0.013 to −0.005	<.001
Sex	0.086	−0.029 to 0.201	.14
Random effects			
Residual variance	0.64	—[c]	—
Between-patients intercept standard deviation	0.90	—	—
Between-patients slope standard deviation	0.10	—	—
Intercept-slope correlation	−0.15	—	—

[a]Estimate: regression coefficient.

[b]The interaction between session number and each entry variable was interpreted as the treatment moderation effect.

[c]Not applicable.

Latent Topics as Correlates of Treatment Response

Both of the LDA models contained 2 topics that moderated treatment response. For descriptions of those topics, see Table 3.

XSL•FO
RenderX

Table 3. Topics in the latent Dirichlet allocation models that moderated treatment response.

Model and topic	Top 10 words[a]	Interpretation of content[b]	Example[c]
7 topic model			
1	child, car, how, son, father, dog, husband's, home, mother, son's	Well-being of family, pets, and loved ones	"What if my father gets in a car accident?"
4	self, thing, thoughts, life, mind, things, try, feeling, own, only	Monitoring of thoughts and worries	"Once again I am thinking about all the things that are wrong in my life."
25 topic model			
21	write, internet therapy, therapy, worry, worry diary, write/book/letter, message, assignment, this, part	Internet therapy	"I'm afraid that the internet therapy does not work for me."
24	child, father, mother, how, mother's, husband's, son, daughter, children's, child	Children, family issues	"Got into an argument with my husband about taking our daughter to daycare."

[a]Words are translated from the Finnish language and appear on descending order based on their word-topic probability in the latent Dirichlet allocation models.

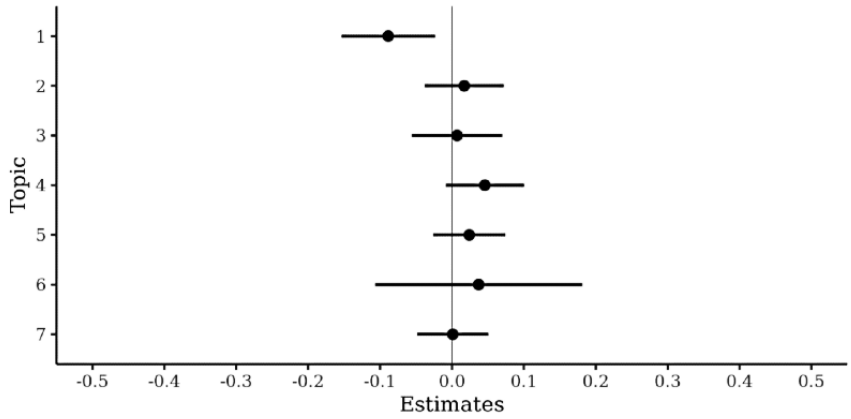
[b]Interpretation of content is based on a qualitative inspection of diary entries with a strong representation of each topic.

[c]Examples are generated by the first author and based on typical diary entries representing each topic.

Data-Driven Model

Topic 1, which was interpreted as worries about the well-being of family and loved ones, was associated with a faster than average per-session decrease in anxiety (B=–0.10 SD/session/Δθ; 95% CI –016 to –0.03). That is, on average, a hypothetical patient who only wrote about topic 1 (mean topic probability θ for topic 1=1) recovered 0.1 GAD-7 standard deviations faster per session as compared with a patient who never wrote about the topic (mean topic probability for topic 1=0). The observed range of mean topic probability for this topic was from 0.0007 to 0.90. Topic 4 (monitoring of thoughts and worries) was associated with a slower than average per-session decline in anxiety (B=0.06 SD/session/Δθ; 95% CI 0.01 to 0.11). After adjusting for the number of entries during the different task assignments, only topic 1 remained a significant moderator of treatment response (Figure 1). Topic 1 also remained a significant moderator when topics 1 and 4 were adjusted with each other.

Figure 1. Topics from the data-driven latent Dirichlet allocation model as moderators of treatment response, adjusted for the writing activity during different worry diary task assignments.

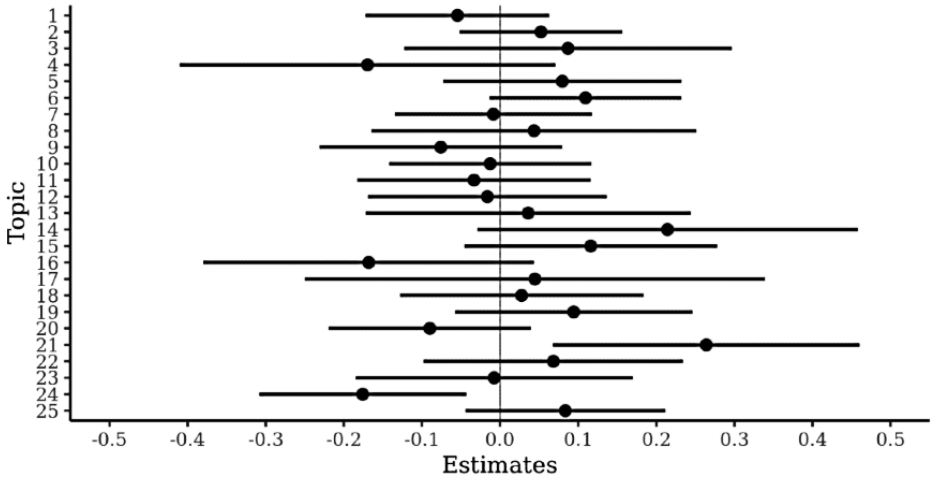


Interpretability-Based Model

Topic 21 (internet therapy) was associated with a slower than average per-session decrease in anxiety (B=0.27 SD/session/Δθ; 95% CI 0.07 to 0.46), whereas topic 24 (children, family issues) was associated with faster than average decrease in anxiety (B=–0.18 SD/session/Δθ; 95% CI –0.31 to –0.05). Both topics remained significant moderators of the treatment response after adjusting for the number of entries during the different task assignments (Figure 2). Finally, both topics remained significant moderators of the treatment response when their treatment effects were adjusted with each other.

XSL•FO
RenderX

Figure 2. Topics from the interpretability-based latent Dirichlet allocation model as moderators of treatment response, adjusted for the writing activity during different worry diary task assignments.



Discussion

Principal Results

In this study, we used topic modeling to analyze text data from worry diary entries written during iCBT treatment for GAD. Higher worry diary writing activity toward the end of the treatment was weakly associated with worse treatment response, defined as a slower per-session symptom change. Our topic models successfully extracted meaningful topics from iCBT texts, some of which were associated with treatment response in an interpretable manner. This is in line with previous psychotherapy topic modeling research [13,18]. Our results extend the previous work by demonstrating topic modeling to be suitable for iCBT task-specific data.

Topics and Their Associations With Treatment Response

Both LDA models contained a topic that reflected worrying about other people and was associated with a faster than average symptom decrease. For the data-driven model, this topic was labeled "well-being of family, pets, and loved ones," based on the contents of entries that were representative of that topic. For the interpretability-based model, this topic was labeled "children, family issues" because it focused more narrowly on worries pertaining to close family. Ruminative self-focus has been associated with depression, anxiety, and negative emotionality [29-31]. Thus, worrying about others rather than merely ruminating about one's self-related problems could reflect a healthy attention to the surrounding world. However, worries considering family members or other important characters may also indicate the presence of important relationships in a patient's life, whereas patients suffering from social isolation would be less likely to write about these topics. Thus, our finding is also in line with research that associates social support with better treatment success, whereas loneliness and a lack of social support are associated with worse outcomes [32,33].

The data-driven model contained another topic that moderated treatment response, labeled "monitoring of thoughts and worries," which was associated with worse treatment response. The entries that represented this topic were typically descriptions of a patient's recurrent focus on worries, appearing as representations of ruminative self-focus [30]. Thus, our finding is in line with the results of a recent meta-analysis that reported the severity of posttreatment anxiety to be associated with higher levels of repetitive negative thinking in the form of rumination or persistent worry [34].

After controlling for the overall writing activity, however, the aforementioned moderation effect related to thought monitoring was no longer significant. A greater total number of worry diary entries was associated with a slower than average symptom decrease, which was explained by the writing activity in the later phases of treatment. This indicates that the occurrence of the "monitoring of thoughts and worries" topic may be associated with high levels of writing activity that continue into the late phases of treatment. In other words, late-phase highly active writers seem to include a group of patients who are not optimally benefitting from the treatment but exhibit persistent worry monitoring behavior. This could partially explain the counterintuitive association between higher writing activity and worse treatment response.

Besides potentially indicating persistent anxiety, the association of high writing activity with the "thought monitoring" topic could also have to do with task-related issues. For example, some patients might have difficulties in adhering to the worry postponement task, which could result in many ruminative entries when the number of entries should be limited. It is also possible that for some patients, the use of a worry diary in the beginning of the treatment could lead to an increased attention to worries and thus lead to a vicious circle of rumination. In any case, our results suggest that what the patients write might be more meaningful than how much they write, supporting the view that the quality of psychotherapy homework completion is meaningful when assessing homework-outcome relations [35]. Furthermore, our findings demonstrate that using topic modeling alongside other correlates (such as writing activity) can offer a broader understanding of treatment effect moderators as compared with using either of them separately.

The interpretation-based model also contained another topic that moderated treatment response labeled "internet therapy." This was associated with a worse treatment response. This topic

XSL•FO
RenderX

was often associated with complaints about the treatment or worries regarding the helpfulness of the treatment. Our finding is in line with Atzil-Slonim et al [13], who reported that treatment-related topics were associated with alliance ruptures and worse treatment outcomes. As these complaints can be easy to identify from iCBT texts, our finding may have applicable value in recognizing patients who are not on track in terms of recovery and may need additional support [36].

When defining the topic model, we specifically focused on defining an unbiased number of topics by adopting a data-driven selection method with a Bayesian approach [19]. The resulting data-driven model consisted of 7 topics. By comparison, our additional interpretability-based model with number of topics selected using a heuristic approach consisted of 25 topics. Based on our qualitative inspection of the worry diary texts that represented topics from the models, the topics in the interpretability-based model appeared easier to identify from the texts and were more semantically coherent than the topics from the data-driven model. The interpretability-based model also included a more diverse range of topics, offering a broader perspective on the contents that appeared in the diary texts.

However, our interpretability-based model also contained some idiosyncratic topics that strongly reflected the writings of one or few patients (Multimedia Appendix 1, Tables S4 and S5). As topic models are descriptive in nature, the idiosyncratic topics do not pose a problem per se. It has been argued that allowing some idiosyncratic topics in a topic model can be useful to separate meaningful or representative topics from "noise" in the data [13]. However, if the topic model is thought to represent a broader patient population, the idiosyncratic topics can be misleading. Furthermore, the idiosyncrasies need to be taken into account when associating the topics with treatment outcomes. In conclusion, neither approach for topic number selection was unambiguously better in our data; rather, both had benefits and drawbacks that should be considered. However, we observed some robust correlates of treatment response across the 2 very different representations.

Strengths and Limitations

The strengths of our study include the nature of our data, which were derived from a naturalistic and nationwide setting of iCBT offered as a part of national public health care. Thus, our data likely constitute a representative sample from the target patient population. Our data set was also fairly large in terms of individual patients, improving the generalizability of our results when compared with previous studies using topic modeling to predict outcomes in samples of under 100 patients [13,18]. Our text corpus used in topic modeling consisted of entries on a task sheet targeted for writing about worries. This type of data offers precise information compared with, for example, data from whole psychotherapy sessions.

Despite our large sample size in comparison with previous research, it is nonetheless a rather moderately sized data set for machine learning. In terms of topic modeling, the amount of text produced per patient in our study was small compared with the previous study by Atzil-Slonim et al [13] that used whole therapy transcripts. Furthermore, it must be noted that the direction of causality cannot be determined from our models. For example, certain topics' association with symptom change may be due to usefulness of writing about that topic or it may reflect patient functioning.

Future Research

As discussed above, the topic number selection method has effects on the topic model estimation. Future iCBT topic modeling research should be mindful about these effects on the quality and generalizability of topics and their associations with treatment outcomes. Furthermore, our study demonstrated topic modeling to be practical and informative when using worry diary texts from an iCBT for GAD. In the future, topic modeling could be used in research on different disorder-specific or transdiagnostic iCBT programs. Topic modeling could also present a means to examine and compare the relative importance and meaning of different text-based tasks within or across different treatment programs, which could offer valuable information in terms of treatment development.

Conclusions

This study demonstrated that topic modeling is a suitable and practical research method for iCBT data. We found topics from a single recurring worry diary task from an iCBT for GAD that were associated with treatment outcomes. Writing about worries regarding people close to the patient was associated with better treatment response. In contrast, monitoring of worries and worries concerning the treatment were associated with worse treatment response. This type of content information has potential for practical implications, such as in informing clinicians about the meaningful patterns in their patients' writing behaviors. The topics also rendered other research variables, such as patients' writing activity, more interpretable.

Acknowledgments

TR and SM were supported by the Academy of Finland (grants 334057 and 335901 to TR). VR, GJ, and JHS were supported by grants TYH2019104 and TYH2015218 from the Government of Finland. SES was supported by grant HUS/441/2022 from Helsinki University Hospital. NOC was supported by grant 288083 from the Research Council of Norway.

Authors' Contributions

SM, TR, and SES conceptualized and planned the study. SM conducted the analyses, with consultative support from TR. SM wrote the original draft. SM, TR, VR, GJ, JHS, OAS, NOC, and SES provided important content and revisions to the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material describing topic modeling procedure, multilevel model equations and topic modeling results in detail.
[DOCX File , 282 KB - jmir_v24i11e38911_app1.docx]

References

1. Eilert N, Enrique A, Wogan R, Mooney O, Timulak L, Richards D. The effectiveness of Internet-delivered treatment for generalized anxiety disorder: an updated systematic review and meta-analysis. Depress Anxiety 2021 Feb;38(2):196-219 [FREE Full text] [doi: 10.1002/da.23115] [Medline: 33225589]
2. Ritola V, Lipsanen JO, Pihlaja S, Gummerus E, Stenberg J, Saarni S, et al. Internet-delivered cognitive behavioral therapy for generalized anxiety disorder in nationwide routine care: effectiveness study. J Med Internet Res 2022 Mar 24;24(3):e29384 [FREE Full text] [doi: 10.2196/29384] [Medline: 35323119]
3. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
4. Etzelmueller A, Vis C, Karyotaki E, Baumeister H, Titov N, Berking M, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. J Med Internet Res 2020 Aug 31;22(8):e18100 [FREE Full text] [doi: 10.2196/18100] [Medline: 32865497]
5. Imel ZE, Caperton DD, Tanana M, Atkins DC. Technology-enhanced human interaction in psychotherapy. J Couns Psychol 2017 Jul;64(4):385-393 [FREE Full text] [doi: 10.1037/cou0000213] [Medline: 28318277]
6. Dirkse D, Hadjistavropoulos HD, Hesser H, Barak A. Linguistic analysis of communication in therapist-assisted internet-delivered cognitive behavior therapy for generalized anxiety disorder. Cogn Behav Ther 2015 Sep;44(1):21-32. [doi: 10.1080/16506073.2014.952773] [Medline: 25244051]
7. Smink WAC, Fox J, Tjong Kim Sang E, Sools AM, Westerhof GJ, Veldkamp BP. Understanding therapeutic change process research through multilevel modeling and text mining. Front Psychol 2019 May 29;10:1186 [FREE Full text] [doi: 10.3389/fpsyg.2019.01186] [Medline: 31191394]
8. Van der Zanden R, Curie K, Van Londen M, Kramer J, Steen G, Cuijpers P. Web-based depression treatment: associations of clients' word use with adherence and outcome. J Affect Disord 2014 May;160:10-13 [FREE Full text] [doi: 10.1016/j.jad.2014.01.005] [Medline: 24709016]
9. Zinken J, Zinken K, Wilson JC, Butler L, Skinner T. Analysis of syntax and word use to predict successful participation in guided self-help for anxiety and depression. Psychiatry Res 2010 Sep 30;179(2):181-186. [doi: 10.1016/j.psychres.2010.04.011] [Medline: 20483481]
10. 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]
11. Blei D, Ng A, Jordan M. Latent dirichlet allocation. J Mach Learn Res 2003;3(1):30.
12. Atkins DC, Rubin TN, Steyvers M, Doeden MA, Baucom BR, Christensen A. Topic models: a novel method for modeling couple and family text data. J Fam Psychol 2012 Oct;26(5):816-827 [FREE Full text] [doi: 10.1037/a0029607] [Medline: 22888778]
13. Atzil-Slonim D, Juravski D, Bar-Kalifa E, Gilboa-Schechtman E, Tuval-Mashiach R, Shapira N, et al. Using topic models to identify clients' functioning levels and alliance ruptures in psychotherapy. Psychotherapy (Chic) 2021 Jun;58(2):324-339. [doi: 10.1037/pst0000362] [Medline: 33734743]
14. Imel ZE, Steyvers M, Atkins DC. Computational psychotherapy research: scaling up the evaluation of patient-provider interactions. Psychotherapy (Chic) 2015 Mar;52(1):19-30 [FREE Full text] [doi: 10.1037/a0036841] [Medline: 24866972]
15. Wasserman L. All of Nonparametric Statistics. New York: Springer; 2006.
16. Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Hollon SD. Component studies of psychological treatments of adult depression: a systematic review and meta-analysis. Psychother Res 2019 Dec;29(1):15-29. [doi: 10.1080/10503307.2017.1395922] [Medline: 29115185]
17. Cuijpers P, Reijnders M, Huibers MJH. The role of common factors in psychotherapy outcomes. Annu Rev Clin Psychol 2019 May 07;15:207-231. [doi: 10.1146/annurev-clinpsy-050718-095424] [Medline: 30550721]
18. Hoogendoorn M, Berger T, Schulz A, Stolz T, Szolovits P. Predicting social anxiety treatment outcome based on therapeutic email conversations. IEEE J Biomed Health Inform 2017 Sep;21(5):1449-1459. [doi: 10.1109/JBHI.2016.2601123] [Medline: 27542187]
19. Chen Z, Doss H. Inference for the number of topics in the latent Dirichlet allocation model via Bayesian mixture modeling. J Computational Graphical Stat 2019 Feb 27;28(3):567-585. [doi: 10.1080/10618600.2018.1558063]
20. Spitzer RL, Kroenke K, Williams JBW, 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]

XSL•FO
RenderX

21. Rosenström TH, Ritola V, Saarni S, Joffe G, Stenberg J. Measurement invariant but non-normal treatment responses in guided internet psychotherapies for depressive and generalized anxiety disorders. Assessment 2021 Dec 14:10731911211062500 [FREE Full text] [doi: 10.1177/10731911211062500] [Medline: 34905968]
22. Stochl J, Fried EI, Fritz J, Croudace TJ, Russo DA, Knight C, et al. On dimensionality, measurement invariance, and suitability of sum scores for the PHQ-9 and the GAD-7. Assessment 2022 Apr 03;29(3):355-366. [doi: 10.1177/1073191120976863] [Medline: 33269612]
23. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
24. Perry PO. corpus: Text Corpus Analysis. 2021. URL: https://CRAN.R-project.org/package=corpus [accessed 2021-09-01]
25. Silge J, Robinson D. Text Mining With R: A Tidy Approach. Boston: O'Reilly; 2017.
26. Jones T, Doane W, Attbom M. textmineR: Functions for Text Mining and Topic Modeling. 2021. URL: https://CRAN.R-project.org/package=textmineR [accessed 2021-09-01]
27. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press; 2007.
28. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2022. URL: https://www.R-project.org/ [accessed 2021-09-01]
29. Brockmeyer T, Zimmermann J, Kulessa D, Hautzinger M, Bents H, Friederich H, et al. Me, myself, and I: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. Front Psychol 2015;6:1564 [FREE Full text] [doi: 10.3389/fpsyg.2015.01564] [Medline: 26500601]
30. Mor N, Winquist J. Self-focused attention and negative affect: a meta-analysis. Psychol Bull 2002 Jul;128(4):638-662. [doi: 10.1037/0033-2909.128.4.638] [Medline: 12081086]
31. Ingram RE. Self-focused attention in clinical disorders: review and a conceptual model. Psychol Bull 1990 Mar;107(2):156-176. [doi: 10.1037/0033-2909.107.2.156] [Medline: 2181521]
32. Roehrle B, Strouse J. Influence of social support on success of therapeutic interventions: a meta-analytic review. Psychotherapy 2008;45(4):464-476. [doi: 10.1037/a0014333] [Medline: 22122534]
33. Wang J, Mann F, Lloyd-Evans B, Ma R, Johnson S. Associations between loneliness and perceived social support and outcomes of mental health problems: a systematic review. BMC Psychiatry 2018 May 29;18(1):156 [FREE Full text] [doi: 10.1186/s12888-018-1736-5] [Medline: 29843662]
34. Monteregge S, Tsagkalidou A, Cuijpers P, Spinhoven P. The effects of different types of treatment for anxiety on repetitive negative thinking: a meta - analysis. Clin Psychol Sci Pract 2020 Jun;27(2):e12316. [doi: 10.1037/h0101777]
35. Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. Behav Ther 2016 Sep;47(5):755-772. [doi: 10.1016/j.beth.2016.05.002] [Medline: 27816086]
36. Pihlaja S, Lahti J, Lipsanen JO, Ritola V, Gummerus E, Stenberg J, et al. Scheduled telephone support for internet cognitive behavioral therapy for depression in patients at risk for dropout: pragmatic randomized controlled trial. J Med Internet Res 2020 Jul 23;22(7):e15732 [FREE Full text] [doi: 10.2196/15732] [Medline: 32706658]

Abbreviations

GAD-7: Generalized Anxiety Disorder–7 item
HUS-iCBT: internet-delivered cognitive behavioral therapy manufactured and delivered by the HUS Helsinki University Hospital
iCBT: internet-delivered cognitive behavioral therapy
LDA: latent Dirichlet allocation

Edited by R Kukafka; submitted 22.04.22; peer-reviewed by N Esfandiari, BT Tulbure; comments to author 26.07.22; revised version received 08.08.22; accepted 28.09.22; published 09.11.22.

Please cite as:
Mylläri S, Saarni SE, Ritola V, Joffe G, Stenberg JH, Solbakken OA, Czajkowski NO, Rosenström T
Text Topics and Treatment Response in Internet-Delivered Cognitive Behavioral Therapy for Generalized Anxiety Disorder: Text Mining Study
J Med Internet Res 2022;24(11):e38911
URL: https://www.jmir.org/2022/11/e38911
doi:10.2196/38911
PMID:36350678

XSL•FO
RenderX

©Sanna Mylläri, Suoma Eeva Saarni, Ville Ritola, Grigori Joffe, Jan-Henry Stenberg, Ole André Solbakken, Nikolai Olavi Czajkowski, Tom Rosenström. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Prevalence of Poisoned Google Search Results of Erectile Dysfunction Medications Redirecting to Illegal Internet Pharmacies: Data Analysis Study

András Fittler[1], PharmD, PhD; Péter Paczolai[1], PharmD; Amir Reza Ashraf[1], PharmD; Amir Pourhashemi[1]; Péter Iványi[2], PhD

[1]Department of Pharmaceutics, Faculty of Pharmacy, University of Pécs, Pécs, Hungary

[2]Department of Systems and Software Technologies, Faculty of Engineering and Information Technology, University of Pécs, Pécs, Hungary

Corresponding Author:
András Fittler, PharmD, PhD
Department of Pharmaceutics
Faculty of Pharmacy
University of Pécs
Rókus street 2
Pécs, 7624
Hungary
Phone: 36 205566509
Email: fittler.andras@pte.hu

Abstract

Background: Illegal online pharmacies function as affiliate networks, in which search engine results pages (SERPs) are poisoned by several links redirecting site visitors to unlicensed drug distribution pages upon clicking on the link of a legitimate, yet irrelevant domain. This unfair online marketing practice is commonly referred to as search redirection attack, a most frequently used technique in the online illegal pharmaceutical marketplace.

Objective: This study is meant to describe the mechanism of search redirection attacks in Google search results in relation to erectile dysfunction medications in European countries and also to determine the local and global scales of this problem.

Methods: The search engine query results regarding 4 erectile dysfunction medications were documented using Google. The search expressions were “active ingredient” and “buy” in the language of 12 European countries, including Hungary. The final destination website legitimacy was checked at LegitScript, and the estimated number of monthly unique visitors was obtained from SEMrush traffic analytics. Compromised links leading to international illegal medicinal product vendors via redirection were analyzed using Gephi graph visualization software.

Results: Compromised links redirecting to active online pharmacies were present in search query results of all evaluated countries. The prevalence was highest in Spain (62/160, 38.8%), Hungary (52/160, 32.5%), Italy (46/160, 28.8%), and France (37/160, 23.1%), whereas the lowest was in Finland (12/160, 7.5%), Croatia (10/160, 6.3%), and Bulgaria (2/160, 1.3%), as per data recorded in November 2020. A decrease in the number of compromised sites linking visitors to illegitimate medicine sellers was observed in the Hungarian data set between 2019 and 2021, from 41% (33/80) to 5% (4/80), respectively. Out of 1920 search results in the international sample, 380 (19.79%) search query results were compromised, with the majority (n=342, 90%) of links redirecting individuals to 73 international illegal medicinal product vendors. Most of these illegal online pharmacies (41/73, 56%) received only 1 or 2 compromised links, whereas the top 3 domains with the highest in-degree link value received more than one-third of all incoming links. Traffic analysis of 35 pharmacy specific domains, accessible via compromised links in search engine queries, showed a total of 473,118 unique visitors in November 2020.

Conclusions: Although the number of compromised links in SERPs has shown a decreasing tendency in Hungary, an analysis of the European search query data set points to the global significance of search engine poisoning. Our research illustrates that search engine poisoning is a constant threat, as illegitimate affiliate networks continue to flourish while uncoordinated interventions by authorities and individual stakeholders remain insufficient. Ultimately, without a dedicated and comprehensive effort on the part of search engine providers for effectively monitoring and moderating SERPs, they may never be entirely free of compromised links leading to illegal online pharmacy networks.

XSL•FO
RenderX

(J Med Internet Res 2022;24(11):e38957) doi:10.2196/38957

KEYWORDS

internet pharmacies; search engine redirection; compromised websites; illegal medicines; patient safety; Europe; erectile dysfunction medications

Introduction

Background

The inherent practicality and convenience of online shopping are proving increasingly influential in consumer's behavior worldwide. Based on the 2020 e-commerce statistics published by Eurostat [1], 89% of all European Union (EU) citizens used the internet within the last 12 months, and 65% of individuals made an online purchase in the same period. Nonprescription medicine or dietary supplements accounted for 28% of these transactions, demonstrating consumers' growing trust in online health- and well-being–related purchases [1]. A large-scale study [2] of changes in information-seeking behavior showed that the most frequently mentioned content is "product information" and "purchase" (30% of all responses in 1997 and 2019), followed by "Health" (18% of all responses in 1997 and 19% in 2019) [2]. Notably, user behavior had been remarkably consistent in the span of 22 years [2].

The use of internet pharmacies and the number of individuals obtaining medications and various health products online are increasing [2]. Several advantages including perceived anonymity, cost savings, and convenience motivate individuals to purchase medications online [3]. Furthermore, the lack of a valid prescription required by legal online and offline vendors is a strong driving force toward illegal online drug purchases [3]. However, several patient safety risks are linked to the procurement of medicines outside the traditional supply chain, including questionable sourcing, poor product quality, substandard and falsified medicines, improper storage, and transportation [4]. Risks are augmented by rogue internet pharmacies considered as a primary source of substandard and falsified medical products in developed countries [5-7].

The widespread availability of search engines and increased public interest in obtaining medicines online imply a major dilemma, whether consumers aiming to purchase medications from the internet are starting their online activity from relevant web pages (eg, a national authority website), or simply searching using their search engine of choice. Most likely the latter is the case. Search engines refer consumers to relevant online resources quickly. Their significance is illustrated by the fact that most trackable website traffic originates from search engines [8], and typically from Google as this platform is handling more than 90% of search queries worldwide. Online distributors choose to use several digital marketing techniques to attract customers via search engines. Website operators apply various search engine optimization (SEO) techniques to improve the visibility of their websites, a practice that is accepted and supported by search engines [9]. SEO is a complex and time-consuming procedure, especially in the international marketplace in which country- and language-specific optimization is required to reach a high-ranking position among organic query results.

For illegal medicine sellers, conventional SEO is neither cost- nor time-effective, as they are constantly threatened with regulatory closure [10]. Furthermore, paid advertisements offering prescription drugs without a prescription by unauthorized pharmacies cannot appear in any of the major paid search advertising services [11,12]. Therefore, alternative dishonest digital marketing methods including web spamming, forum abuse, and additional "black hat" SEO techniques are used by illegal drug distribution websites to promote their links in the unpaid search engine results pages (SERPs) to gain favorable search engine rankings [13,14].

As a result, the user's query on a search engine may contain both "normal" domains (ie, those related to the query) and "compromised/deceptive" domains (ie, ones that are unrelated to the query). The latter domains are promoted in the rank using "black hat" SEO methods, undermining the value proposition of search engines, as search results are presented with deceptive views of a website with inflated relevance to selected search terms. Individuals (search engine users) are referred to low-quality content or malicious websites when clicking on a deceptive search result. Consequently, the deceptive web pages practically "poison" the search result; therefore, this technique is termed as "search engine poisoning" or "search redirection attack" [9,15].

Manipulation of search results for erectile dysfunction medications was published nearly a decade ago by Leontiadis et al [15,16] and Wang et al [17]. Sildenafil was the first commercially available phosphodiesterase type 5 (PDE5) inhibitor available since 1998, followed by vardenafil, tadalafil, and avanafil [18]. Increasing prevalence of erectile dysfunction and widespread use of PDE5 inhibitors as the first-line oral treatment worldwide [19] have resulted in growing demand, which illegal online vendors have been taking advantage of [20].

Objectives

The major aim of our study is to introduce the relatively unknown but significant and persistent issue of poisoning of search engine results (SERs) of erectile dysfunction medications in European countries. Furthermore, the study is meant to measure the scale of the problem and illustrate the redirection networks referring users (patients) to illegal internet pharmacies. Public health significance of the problem is illustrated by the estimation of the likelihood of consumers clicking on poisoned search results and the number of monthly visitors redirected to illicit pharmacy networks. Our utmost aim is to warn the general public and raise the awareness of authorities and law enforcement agencies, thus facilitating long-awaited countermeasures.

XSL•FO
RenderX

Methods

Mechanism of Search Engine Poisoning and Redirection

A search engine poisoning attack begins with an attacker hacking into a vulnerable web page. Common targets are outdated, vulnerable, or complex content management and blogging systems (eg, WordPress; see Figure 1, part 1). Once the attacker has access to the system, a new code is injected, and the hacked website will "interrupt" all incoming HTTP requests to the original web page and respond to these requests differently from the original operation [15]. Typically, users are redirected through a redirection chain, consisting of intermediate pages to a final page. The destination is the illegal pharmacy website most users are unwillingly visiting. However, users do not see the original content of the compromised website after clicking on the search results, because they are presented with the unwanted final page, as hacked websites redirect the web browsers within milliseconds. Redirection attacks—identifiable in various search engines such as Google, Bing, and Yahoo!—disregard term relevance constraints and target search terms of the actual search; however, at the same time, the original content of the hacked website (domain) becomes irrelevant to the search terms used (see Figure 1, part 2).

Figure 1. Illustrative figure of how users pass through a redirection chain from the search result page to the final destination illegal online pharmacy website.

In the case of search engine poisoning attack, it is important that compromised websites look differently, depending on the visitor, due to the so-called cloaking method [13]. The original content stuffed with keywords and links to increase page rank is shown to the automated agent/crawler (eg, Google bot), meanwhile the redirected illegitimate online vendor is displayed to the customer (see Figure 1, part 3) [16]. Currently no efficient technique capable of identifying all spam web pages is available [13]. Because of the cloaking method used by the illegitimate pharmacy operators, the automation of the content evaluation of SERs is difficult and precise detection requires manual assessment or checking.

Obtaining and Evaluating SERs in National and International Data Sets

Search engine query results and links were documented and manually evaluated to simulate and evaluate what consumers see while browsing. Manual data acquisition was necessary as automatic search queries are prohibited by search engine providers and cloaking is difficult to identify automatically. The focus of the research was on erectile dysfunction

XSL•FO
RenderX

medications as a popular category affected by illegal online trade and potential source of substandard and falsified medicinal products [20,21]. Consequently, the search queries represent purchase intent (buying prescription medications online), rather than informative types of search (looking for product information). The 4 primary active pharmaceutical ingredients (APIs), sildenafil, tadalafil, vardenafil, and avanafil, were searched for using Google, the most popular search engine. Country-specific data were obtained by individualizing national search using the search terms of the "API" and the "buy" words in the language of the given country (eg, "comprar sildenafil" for Spain). Furthermore, search settings in Google have been adjusted to the preferred region. To track the evolution of the phenomena, the first 20 organic SERs were evaluated during 3 consecutive years: August and October 2019, August 2020, and November 2021 for the national data set. Meanwhile, the first 40 SERs were included in the international data set evaluated in November 2020. Accordingly, we conducted our research on 2 data sets: a long-term evaluation of Hungarian SERs and an international sample in Hungary and an additional 11 other countries (Bulgaria, Croatia, Estonia, Finland, France, Greece, Italy, Romania, Spain, Sweden, and United Kingdom) from different regions of Europe. As most (88%) users click on results appearing in the top 10 SER positions [22], by documenting the top 20 results we consider our findings representative for online queries at the time of evaluation. SER links of websites offering medicinal products for sale were included for evaluation; nonrelevant query results were excluded from our evaluation.

The documented search result data included date, country, search language, API, search phrase, URL and domain name, SER ranking, destination website URL for redirections, and website category. Two figures were used to describe the significance of the phenomena regarding search engine redirection attacks in SERs: (1) prevalence of hacked links in SERPs and (2) cumulative click-through rate (CTR). Both measures correlate with the likelihood of users—intentionally or unintentionally—visiting illegal pharmacies. Prevalence is calculated by dividing the number of infected links by the total number of evaluated links in SERPs. Based on Google's organic search ranking, CTR is a probability value of clicking on a given link assigned to each measured SER position. On the first page of the search (Google) result, 1-10 CTR per ranking values were determined based on the analysis by Sistrix [22], while further CTRs for 11-40 SER positions were computed with the equation of the exponential trend line connecting the first 1-10 SERP datapoints ($y=26{,}76e^{-0.258x}$, where y is the predicted CTR and x is the SER rank; $R^2=0.927$). Cumulative CTRs express the sum of CTR values regarding all documented positions in SERPs.

Compromised sites redirecting to international illegal medicine retailers have been classified into 3 categories referencing the redirection's life cycle based on Leontiadis et al [16]. First, the compromised site is likely a future redirect (hacked website content with or without links; however, no automatic redirection is yet observed). Second, active redirection to an international illegal medicinal product vendor via a compromised site. Lastly, inactive redirection, that is, sites used to be redirecting, but no longer redirecting, because they are not accessible at the time of evaluation, displaying 404 error code, or similar.

Graph Visualization, Legitimacy, and Traffic Analysis Regarding Destination Websites

Compromised SERP links leading to international illegal medicinal product vendors via redirection (active links) were evaluated and networks have been generated with Gephi [23], an open-source graph visualization and analysis tool. The national and international data sets were visualized as directed graphs illustrating the source and destination website domains. Multiple links from the same domain accounted for increased weight of the edge. The average degree (average number of edges per node in the graph), the in-degree (number of connecting edges), and the page rank (importance score of a node within a directed graph) of nodes were computed.

Destination websites offering products for sale in the national data set were categorized as follows: legitimate online pharmacies, illegal medicine retailers (rogue online pharmacies), or dietary supplement seller (nonpharmacy web shops). Destination website categories were not defined for EU countries, so only links with redirection to illegal online sellers were documented regarding the international data set. Destination website legitimacy was checked at LegitScript [24] and categorized as approved, unlicensed, or rogue (illegitimate). The estimated number of monthly unique visitors of the root domain for all regions at the time of evaluation is provided by SEMrush traffic analytics [25].

Data were analyzed using SPSS Statistics 26 for Windows (IBM Corp.) and MS Excel (Microsoft Inc.).

Ethical Considerations

There were no ethical issues, as only publicly available data obtained from SEs and websites were documented and evaluated. Furthermore, no customer or personal data were measured, recorded, or stored in this study.

Results

Compromised Websites Among SERPs of Medications for Treating Erectile Dysfunction in Hungary Between 2019 and 2021

The results show that during our 3-year observation period, there were no legitimate internet pharmacy websites among the evaluated SERPs. A decrease in the number of compromised sites linking visitors to illegitimate medicine sellers has been observed during our study period, while inaccessible broken links have increased. Similarly, the number of national rogue online pharmacies has increased in SERs up through 2021. All active ingredients have been affected by poisoning, with avanafil showing a somewhat diminished prevalence (Table 1).

XSL•FO
RenderX

Table 1. Top 20 search engine results page link categories for 4 erectile dysfunction medications.

Link category	August 2019, n (%)	October 2019, n (%)	August 2020, n (%)	October 2021, n (%)
Legitimate online pharmacy (n=80)[a]	0 (0)	0 (0)	0 (0)	0 (0)
National illegal medicinal product seller (n=80)	8 (10)	12 (15)	16 (20)	34 (43)
International illegal medicinal product vendor via compromised site and redirection (active; n=80)	43 (54)	33 (41)	25 (31)	4 (5)
Avanafil (n=20)	9 (45)	5 (25)	3 (15)	0 (0)
Sildenafil (n=20)	12 (60)	9 (45)	6 (30)	1 (5)
Tadalafil (n=20)	12 (60)	9 (45)	8 (40)	1 (5)
Vardenafil (n=20)	10 (50)	10 (50)	8 (40)	2 (10)
Compromised site without redirection (n=80)	5 (6)	3 (4)	1 (1)	0 (0)
Not accessible (eg, 404) at the time of evaluation (n=80)	2 (3)	7 (9)	9 (11)	15 (19)
Dietary supplement web shop (n=80)	9 (11)	10 (13)	14 (18)	8 (10)
Other sites not offering products for sale (n=80)	13 (16)	15 (19)	15 (19)	19 (24)

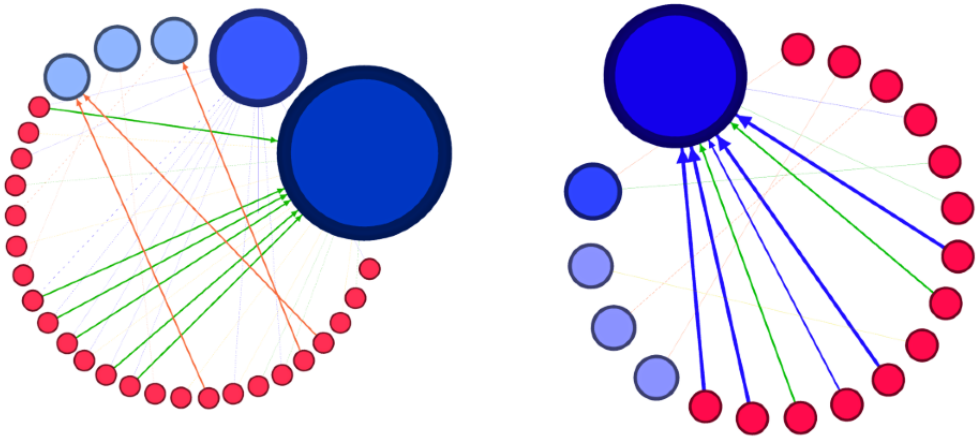
[a]According to national regulations, legitimate online pharmacies in Hungary cannot offer prescription medications—including oral medications for erectile dysfunction—via the internet.

Although most of the compromised websites were "true redirects" transferring individuals to international online sellers, we occasionally came across hacked sites without redirection. For example, in these cases, the rogue online pharmacy was operating under a subpage of the hacked domain, or the medication-related text was filled with keywords and links (so-called keyword stuffing and link building), indicating "black-hat" SEO techniques.

Such pages are likely to rank higher in search engines and develop redirects as time passes. In other instances, the web page we were looking for did not exist on the website's server. Pages not accessible (eg, 404 error) at the time of evaluation could be related to website administrators identifying the malicious redirect code inserted into a website. According to our observation, hacking is followed by the malicious redirection life cycle, which consists of future (inactive pages ready to become active), active, and finally inactive stages.

The complexity of the graphs decreased (the average degree changed from 1.17 to 0.667), between August 2019 and October 2021 (Figure 2). A majority (11/14, 79%) of the evaluated online pharmacies were categorized as rogue by LegitScript. We identified 5 destination online pharmacy websites in the link network at each evaluation date, except for October 2021. Initially, destination domains (eg, acs-pharmacy.com and evo-pharmacy.com) received numerous incoming links from SERs and played a central role in the network. By the end of the 3-year evaluation period, illegal pharmacy websites in-degree and page rank values underwent substantial reduction (Table 2). Website traffic analytics by SEMrush indicated a high number of monthly visitors (range 370-155,400) for important nodes with high page-rank values within the graph. This value illustrates the destination site's global visitor count in the given month of evaluation.

Figure 2. Visual graph of SERP links of compromised websites and illegal online medicine vendors accessed via search redirection attack visited in August 2019 (left) and August 2020 (right). SERP: search engine results page.



XSL•FO
RenderX

Table 2. Graph statistics, legitimacy rating, and traffic history regarding referred illegal medicine vendors for Hungarian erectile dysfunction medication search queries.

Domain accessed following search redirection attack	Date	In-degree[a]	Page rank[b]	Legitimacy rating (LegitScript)	Number of unique visitors per month (SEMrush)[c]
acs-pharmacy.com	August 2019	16	0.209	Rogue[d]	155,400
acs-pharmacy.com	October 2019	16	0.332	Rogue	117,000
1-pharm.com	August 2019	12	0.140	Rogue	11,000
specialmedassortment.com	August 2019	2	0.054	Rogue	3600
myworldpharma.com	August 2019	2	0.054	Not in database	4000
pharmpillsonline.com	August 2019	2	0.054	Rogue	800
herbsandmeds.com	October 2019	2	0.061	Rogue	5200
pharmrx-1.com	October 2019	2	0.051	Rogue	6500
cheap-pharma.com	October 2019	1	0.042	Rogue	5100
big-pharmacy.com	October 2019	1	0.032	Rogue	15,600
evo-pharmacy.com	August 2020	9	0.279	Rogue	83,400
evo-pharmacy.com	October 2021	2	0.574	Rogue	30,400
eu-pharm.de	August 2020	2	0.087	Not in database	370
ezshopremedieshere.com	August 2020	1	0.059	Not in database	Not in database
canadarx24h.com	August 2020	1	0.059	Rogue	5200
medsalltheworld.com	August 2020	1	0.059	Rogue	3100

[a]In-degree value shows the number of links adjacent to a domain.

[b]The page rank algorithm measures the importance of each node within the graph.

[c]The estimated number of monthly unique visitors of the root domain for all regions at the time (month) of evaluation provided by SEMrush traffic analytics.

[d]Rogue: online pharmacy website engaged in illegal activity; a rating determined by LegitScript.

International Relevance of Compromised SERPs in Europe 2020

A total of 1920 search results were evaluated in November 2020, in accordance with the results of the aforementioned 4 APIs listed in the top 40 results on the SERP pages throughout 12 European countries. Of those, 380 (19.79%) search query results were compromised, with a majority (n=342, 90%) of the links of the 230 infected source domains redirecting individuals to 73 international illegal medicinal product vendors. The remaining SER links were leading to compromised sites without redirection (6/380, 1.6%) or not accessible web pages/sites (32/380, 8.4%). Descriptive graph statistics of the international data set, website legitimacy category, and traffic history regarding destination online pharmacies with at least five referring links are depicted in Table 3.

The most influential destination domain in the international redirection graph was “ezshopremedieshere.com,” with 79 referring links from search queries in most (8/12, 66%) of the evaluated European countries, and 61,400 unique global visitors in November 2020. Although several destination websites had numerous incoming links, the average in-degree value was 1.11, as most nodes had only 1 (30/79, 38%) or 2 (12/79, 15%) compromised referrals from search engines (Figure 3). The number of monthly global visitors per domain was the highest for “forecastarrays.us,” “cheapshopmed.com,” and “haiyuanpenguan.com,” attaining 566,100, 135,100, and 128,300 visitors, respectively, according to SEMrush traffic analytics. Interestingly, these high-traffic domains had only a small number (1-3) of incoming links from SERs and only 1 European country was affected in each case (Finland, Estonia, and Croatia, respectively). The “cheapshopmed.com” domain is a rogue online pharmacy in the LegitScript database. However, the “forecastarrays.us” and “haiyuanpenguan.com” domains contain compromised pages, including their intended content, and they can be accessed after redirection with an embedded online pharmacy content, so the visitor count of these domains is likely to include nonmedicinal purchase intention also. Website traffic estimation was available for 40 destination domains, with 35 having pharmacy-specific domain names (including terms, such as Rx, pharm, meds, pills). These 35 active online pharmacy domains, accessible from 12 European countries via compromised links in search engine queries, included a total of 473,118 unique visitors during November 2020.

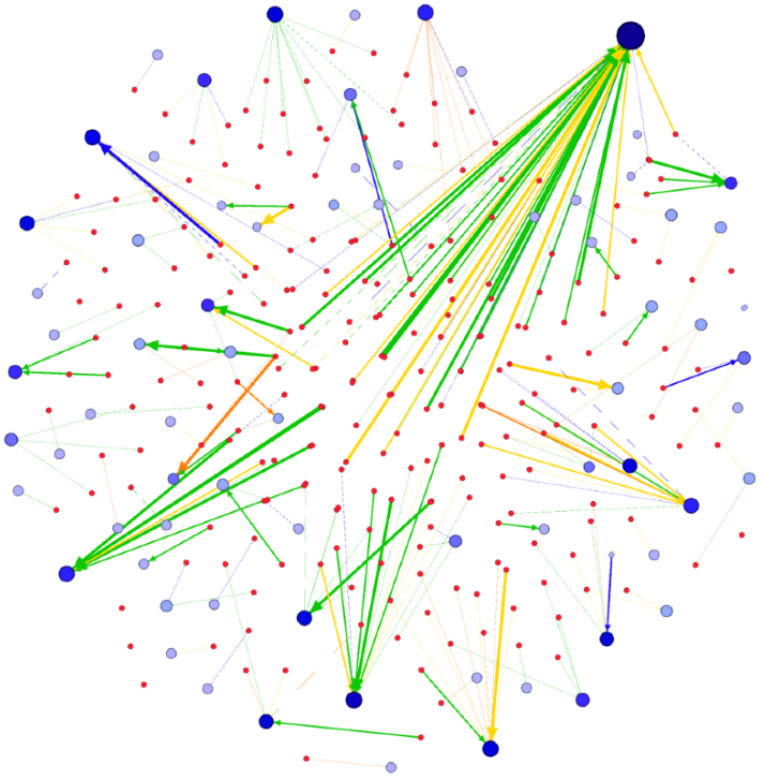
XSL•FO
RenderX

Table 3. Graph statistics, legitimacy rating, and traffic history regarding selected referred illegal medicine vendors for erectile dysfunction medication search queries in 12 European countries (November 2020).

Domain accessed following search redirection attack	In-degree	Page rank	Countries affected	Legitimacy rating (LegitScript)	Number of unique visitors per month (SEMrush)
ezshopremedieshere.com	79	0.080	Croatia, Estonia, France, Greece, Hungary, Italy, Spain, Sweden	Not in database	61,400
evo-pharmacy.com	20	0.017	Hungary	Rogue	Not in database
rx-qualityshop.com	19	0.023	Croatia, Estonia, Finland, Romania, Sweden	Rogue	Not in database
your-meds-store.com	14	0.013	Croatia, Estonia, Finland, Greece, Italy, Romania, Spain	Rogue	4600
onlinepharmacyhub.com	13	0.018	Croatia, UK, Estonia, Romania	Not in database	2300
overnightpharm.com	11	0.015	UK, Estonia, France, Italy, Spain, Sweden	Rogue	321
rx-24-online.com	10	0.018	UK, Sweden	Rogue	Not in database
hot-med.com	9	0.017	Estonia, Spain	Rogue	21,500
usamedicineget.com	8	0.005	Croatia, Estonia, Romania	Rogue	5000
igohealth365.com	8	0.012	UK, France, Italy, Spain	Rogue	Not in database
qualitypillsprovider.com	7	0.007	Hungary, Spain, Sweden	Rogue	519
meds-store-24h.com	7	0.010	Finland, Greece, Italy, Spain	Rogue	7800
pills-group.com	6	0.010	Italy	Not in database	Not in database
vipcanadianstore.com	6	0.008	France, Italy, Sweden	Rogue	Not in database
online-secure-shop24h.com	6	0.009	Bulgaria, Greece, Italy, Spain	Rogue	8400

XSL•FO
RenderX

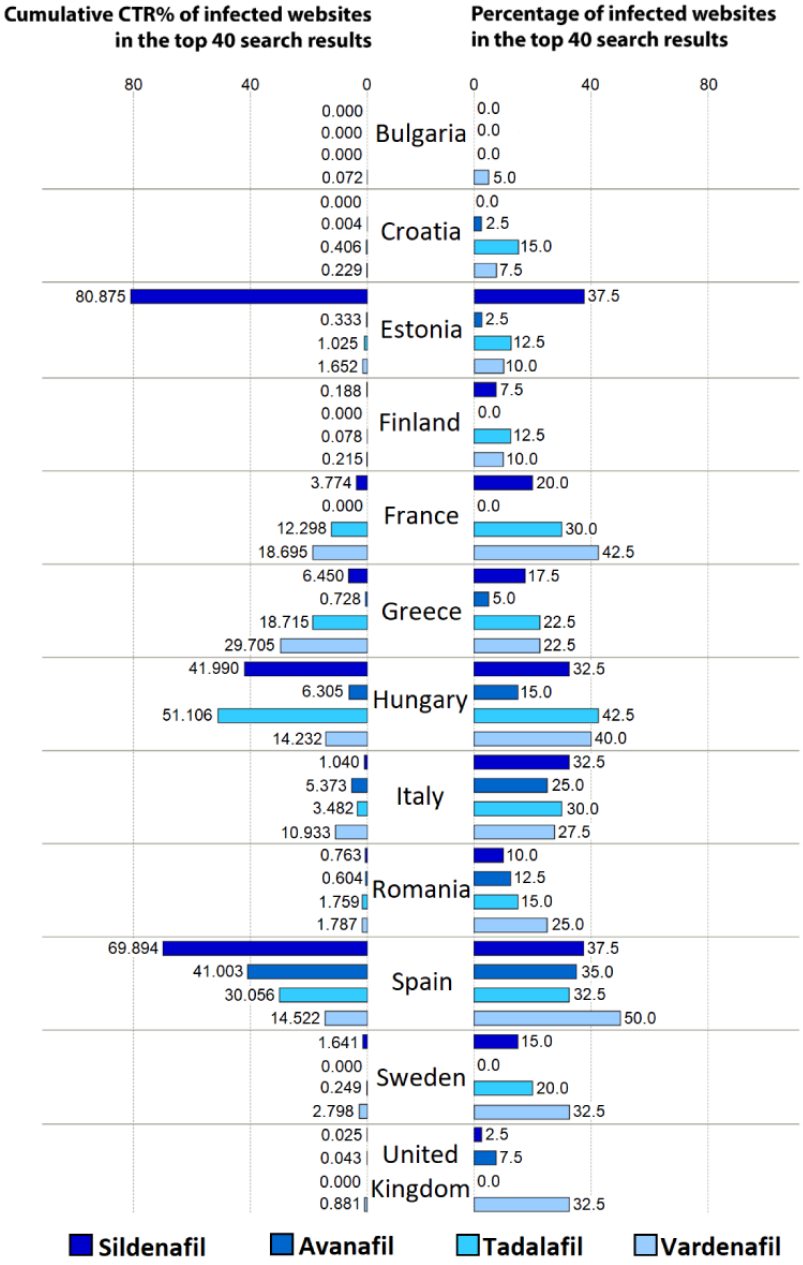
Figure 3. Graph of compromised websites (n=230) and illegal online medicine vendors (n=73) accessed via search redirection attack in 12 European countries visited in November 2020. Node size—represented by circles—illustrate the in-degree property of a domain in the graph. Small red nodes show compromised website domains in SERs and destination websites are labeled with blue. The edge—representing links—are colored based on the API name used in search queries (blue for sildenafil, green for vardenafil, yellow for tadalafil, and orange for avanafil). API: active pharmaceutical ingredient; SER: search engine result.



The EU countries are affected differently by redirection links within SERPs, leading to illegitimate online pharmacy websites (Figure 4). In the "Methods" section, we proposed 2 metrics to illustrate the magnitude of the problem manifested throughout European countries. The proportion of the hacked pages as a percentage of the total search query results and the cumulative CTR percentages were calculated to illustrate the issue of the compromised websites in a complex manner in each country's SERP. It is important to view cumulative CTR and the number of compromised websites as both unique and complementary factors. To state an example, if a country's SERP has several websites lower down the list, the cumulative CTR will be minimal. However, these websites pose a potential risk of rising surreptitiously quickly through the ranks and gaining higher CTRs.

XSL•FO
RenderX

Figure 4. Cumulative click-through rate (CTR) prevalence of redirection links within search engine result pages leading to illegitimate online pharmacy websites search queries in 12 European countries.



Compromised links redirecting to active online pharmacies were present in search query results of all evaluated countries. The prevalence of compromised links in national SERs was the highest in Spain (62/160, 38.8%), Hungary (52/160, 32.5%), Italy (46/160, 28.8%), and France (37/160, 23.1%), whereas it was the lowest in Finland (12/160, 7.5%), Croatia (10/160, 6.3%), and Bulgaria (2/160, 1.3%). Cumulative CTR values computed for APIs indicated the highest potential impact and danger of search engine redirection attacks for avanafil in Spain (41.0%), sildenafil in Estonia (80.9%), tadalafil in Hungary (51.1%), and vardenafil in Greece (29.7%). Prevalence and cumulative CTR metrics were relatively high for all APIs in Hungary and Spain, indicating a larger number of infected SER links with relatively high-ranking positions in search queries. Accordingly, consumers searching for erectile dysfunction medications online are more likely affected by online medicine purchase opportunities presented by illegal online pharmacies applying search engine redirection attack as a marketing technique in these countries. Although SERs in Romania, Finland, and Greece contain a substantial number of compromised links, because of low rankings, the cumulative CTR vales are low, indicating that consumers are less likely to click on compromised links leading to the destination illegal online pharmacy websites. The complete redirection network is illustrated in Figure 3.

Hacked websites are not specialized in active ingredients and target domains. Of the observed 230 infected source domains, many (n=65, 28.3%) promote various APIs. Although the majority (160/230, 69.6%) of source infections drive traffic to a single destination, many redirect individuals to various online pharmacy websites (range 1-6; mean 1.49 redirection links of independent destination domains).

XSL•FO
RenderX

Discussion

Principal Findings

The evolution of online advertising methods and specialization have led to the development of affiliate networks, an established method for legitimate merchants in which sponsors pay a commission to advertisers delivering traffic to their websites. Unfortunately, illegal online pharmacies are also a typical example of affiliate networks and search engine poisoning is a tool linked to affiliates to convert visitors from search engines. A robust number of independent affiliates, acting as advertisers or traffic brokers, received high (30%-40%) commissions for promoting illegal medication vendors and delivering traffic to the sponsor websites in which medications are sold to customers [14]. This affiliate program business model has numerous advantages for its participants. Sponsors (destination illegal pharmacy websites) do not have to heavily invest in marketing campaigns. Even more advantageous is that they free themselves from direct exposure to the criminal risks associated with large-scale advertising. Affiliates generate sales for sponsors by only focusing on attracting customers without developing web shops, customer service, etc. Online pharmaceutical sales are one of the oldest and largest affiliate program markets, with an estimated turnover of 500,000-600,000 customers, 700,000 billed orders, and US $73,000,000-85,000,000 revenue per 3-year period (2007-2010) analyzed by McCoy et al [14] referencing 2 major affiliate networks (Glavmed and SpamIt). By evaluating the change of new customer acquisitions, the authors concluded that affiliate programs attract new customers at a steady rate (approximately 3300/week). Thus, the market of counterfeit pharmaceuticals was not saturated, suggesting latent customer demand [14]. Furthermore, the same data set provides evidence for customer loyalty and satisfaction regarding online pharmacies, as repeat purchases constitute more than 20% of overall revenue. Our previous findings also indicate that a vast number of online pharmacies operate illegally and offer medicines to buyers in the long run [10].

It has been estimated that the number of men experiencing erectile dysfunction worldwide can reach 332 million by 2025 [19]. Erectile dysfunction medications containing PDE5 inhibitors are highly prone to falsification with proven potential health risk for patients. Analytical investigation of these products often shows the presence of dangerous excipients of nonpharmaceutical origin or quality, more than 1 undeclared PDE5s, and active ingredient amounts higher than declared values often surpassing the maximum therapeutic dose [5]. Previous research [26] regarding patient safety risks assessment of the online market of medicinal products revealed that Google search results include several suspicious links. By clicking on these SERs, the visitor is apparently redirected to an unlicensed drug distribution page by initially clicking on the link of a legitimate, yet irrelevant domain. This unfair online marketing of search redirection attack is thought to play a decisive role in the illegal internet pharmaceutical marketplace. Although search engine redirection attacks leading visitors to illegal online pharmacy networks have been previously published [9,16], we did not find relevant publications in medical informatics journals during the past decade. Admittedly, search engine redirection attacks are not limited to Google, the most popular search engine. The same phenomena could be identified in Microsoft Bing and Yahoo!. Seemingly, this unsolved issue has sunk into oblivion. This study was aimed to describe, map, and highlight its national and international significance.

Nearly half of search results were redirecting individuals to illicit medicine vendor sites during our national results obtained in 2019, with compromised websites being dominant in SERPs. This finding correlates with a previous study by Leontiadis et al [16], highlighting how redirections constitute the most significant proportion of results for the query set implemented in this study. Although the prevalence of compromised links in SERs and the complexity of the graphs have decreased in our national data set between August 2019 and October 2021, the danger has not dissipated. Consumers searching for ivermectin during the COVID-19 pandemic were more likely to find links redirecting to illegal medicine retailers that represent 73.3% of SER links within the first 30 search results in Google in March 2021 [26]. Despite the attempts to prevent this “black hat” SEO technique proposed a decade ago, limited success can be observed [9], and we are facing a constant issue that has not been solved for a relatively lengthy period.

Our international search query data set obtained from a representative sample of SERs among 12 European countries illustrates the international significance of search engine poisoning. All evaluated countries are affected, as at least one of four active ingredients for the treatment of erectile dysfunction was offered for sale via compromised links. The overall prevalence of hacked links in SERs was highest among Spain, Hungary, Italy, and France. Among 1920 manually evaluated links, we documented 380 compromised results from a total of 230 websites (domains) leading to 73 illegal online medicine vendors. The majority of these illegal online pharmacies (41/73, 56%) received only 1 or 2 compromised links. Meanwhile, the top 3 domains with the highest in-degree property received more than one-third of all incoming links. These findings support earlier studies stating that illicit advertising business is dominated by only a handful of big-league players [16].

An important implication regarding our findings is that search-redirection attackers use a complex system with potentially vulnerable elements to convert traffic to their illegitimate destination websites. We conclude that such practices can be disrupted by various stakeholders in a number of ways (Textbox 1).

Most likely, if any 1 or more than 1 of the aforesaid measures are considered, the redirection network collapses, and infected source websites will not appear, nor will they rank high in the search results. Lastly, they will not actively redirect to illegitimate online pharmacy domains.

A common feature of the aforesaid measures is the undisrupted continuity of the system, as it most likely requires time to build up such a complex network among numerous stakeholders. Findings of previously published literature suggest that the median survival time of a source infection is 19 days; however, some claim a lot lengthier time (17% of infections lasted at least six months, while 8% survived for more than 1 year) [16]. Our

XSL•FO
RenderX

findings also corroborate this, as 4 compromised pages in our national data set remained in the top 20 results for more than 2 years, between August 2019 and October 2021.

Textbox 1. Possible solutions to overcome search-redirection poisoning redirecting to illegal internet pharmacies.

- Search providers and authorities can identify compromised links by monitoring popular medicinal product–related search terms (eg, brand or active ingredient name of prescription medications), as infected websites contain numerous relevant keywords and links to rank high in search engine results pages (SERPs) for popular queries and to publicize themselves.
- In addition to manual evaluation of SERPs, previously published link-based and content-based algorithms as well as tailor-made automatic detection and classification engines can be used as benchmarks in the effective identification of pharma scam campaigns [27].
- Search engine providers play a decisive role in monitoring and moderating SERPs. Without their dedicated and comprehensive effort, SERPs may never be free of compromised links leading to illegal online pharmacy networks. Automated URL-based classification methods, similar to deSEO [28] proposed in 2011, can only be applied if search engine providers provide search query logs to authority or academic parties.
- If operators fail to identify the infection, compromised websites remain among the top results and maintain the functionality of redirecting. Consequently, the operators of vulnerable legitimate domains should be notified so that they can take action to improve content management system security and remove hacked pages.
- The intermediate redirection chain elements need to remain operational for effective redirection and search engine optimization, so when the webmaster removes the infection triggering the redirection, or any intermediary page, the redirection chain ceases to function.
- The destination illegitimate online pharmacies must stay online to remain operational. Therefore, drug authorities and law enforcement agencies can shut down final destination domains of rogue online pharmacies with a high number of incoming links and unique visitors.

As the number of infected websites appearing in SERPs and all other compromised websites within the redirection chain is considerably high and the number of destination websites are relatively low, it is reasonable to take measures against the latter by shutting down websites and domains. However, the efficacy of this intervention does not seem to be efficient enough, considering the fact that the Operation Pangea coordinated by Interpol has taken down more than 150,000 websites between 2008 and 2020. Despite this large-scale removal, an extremely large number of links (113,020 websites and online marketplaces) were subsequently closed down in 2021 [29,30], demonstrating the substantial scale and recurrence of this issue, which remains unresolved.

Limitations

Admittedly, our study bears several limitations, for instance, the search query results of only 1 search engine have been summarized; however, we believe that the validity of our methodology can be explained by the dominant market share of the search engine. Furthermore, as opposed to brand-name queries, API-based search may offer varied results; however, Google's complex algorithm is likely to provide results for related searches. API was used because our aim was to find all relevant websites, regardless of their original and generic names, varying from country to country, including unapproved generics and falsified medicines. Legitimacy of all final destination websites cannot be evaluated objectively, as there is no reliable database to evaluate all websites. However, we assumed all online medicine vendors using search engine redirection attack to attract customers and offer prescription medicines for sale most likely bear malicious intent and can be categorized as illegitimate online pharmacies.

In conclusion, our results illustrate that the phenomena of search engine poisoning have been persistent during the past decade and affiliate networks linked to illegitimate online pharmacies are flourishing. This supports the presumption that uncoordinated interventions aiming at ceasing illicit medicinal online purchases by authorities and individual stakeholders are not yet sufficient. It is a problem that has not been solved for more than a decade. Importantly, uncontrolled illegal sale of medications has many unfavorable consequences for the health of consumers and the safety of the pharmaceutical supply chain. Detecting and eliminating malicious links promoting illegal online pharmacies in search engines are of great importance with regard to cybersecurity and patient safety.

Acknowledgments

The research is supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (BO/00238/20/5), co-financed by the European Social Fund (EFOP-3.6.1.-16-2016- 00004). The authors thank Dr. Ramiz Haghighi for his valuable contribution in the initial phase of this study. The authors are grateful for the opportunity to use website analytical services provided by SEMrush.

Authors' Contributions

AF was responsible for conceptualization, methodology, writing of original draft, and supervision. PP was responsible for study investigation. ARA performed formal analysis, writing of original draft, and visualization. AP was responsible for study visualization. PI contributed to conceptualization and writing—review and editing.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Eurostat. Internet purchases - goods or services (2020 onwards) Internet. Eurostat. 2022 Mar 30. URL: https://ec.europa.eu/eurostat/databrowser/view/ISOC_EC_IBGS/default/table?lang=en&category=isoc.isoc_i.isoc_iec [accessed 2022-10-24]
2. Fittler A, Vida RG, Káplár M, Botz L. Consumers Turning to the Internet Pharmacy Market: Cross-Sectional Study on the Frequency and Attitudes of Hungarian Patients Purchasing Medications Online. J Med Internet Res 2018 Aug 22;20(8):e11115 [FREE Full text] [doi: 10.2196/11115] [Medline: 30135053]
3. Orizio G, Merla A, Schulz PJ, Gelatti U. Quality of online pharmacies and websites selling prescription drugs: a systematic review. J Med Internet Res 2011 Sep 30;13(3):e74 [FREE Full text] [doi: 10.2196/jmir.1795] [Medline: 21965220]
4. Fittler A, Vida RG, Rádics V, Botz L. A challenge for healthcare but just another opportunity for illegitimate online sellers: Dubious market of shortage oncology drugs. PLoS One 2018 Aug 28;13(8):e0203185 [FREE Full text] [doi: 10.1371/journal.pone.0203185] [Medline: 30153304]
5. Gaudiano MC, Manna L, Rodomonte AL, Bartolomei M, Bertocchi P, Gallinella B, et al. A survey on illegal and counterfeit medicines for the treatment of erectile dysfunctions in Italy. J Sex Med 2012 Aug;9(8):2130-2137. [doi: 10.1111/j.1743-6109.2012.02770.x] [Medline: 22612848]
6. Blackstone E, Fuhr J, Pociask S. The health and economic effects of counterfeit drugs. Am Health Drug Benefits 2014 Jun;7(4):216-224 [FREE Full text] [Medline: 25126373]
7. Mackey TK, Nayyar G. Digital danger: a review of the global public health, patient safety and cybersecurity threats posed by illicit online pharmacies. Br Med Bull 2016 Jun;118(1):110-126 [FREE Full text] [doi: 10.1093/bmb/ldw016] [Medline: 27151957]
8. Goodwin D. 71 Mind-Blowing Search Engine Optimization Stats Internet. searchenginejournal. URL: https://www.searchenginejournal.com/seo-guide/seo-statistics/#close [accessed 2022-10-24]
9. Lu L, Perdisci R, Lee W. SURF: detecting and measuring search poisoning. In: CCS '11: Proceedings of the 18th ACM conference on Computer and communications security. New York, NY: ACM Press; 2011 Presented at: CCS'11: the ACM Conference on Computer and Communications Security; October 17-21, 2011; Chicago, IL p. 467-476. [doi: 10.1145/2046707.2046762]
10. Fittler A, Bősze G, Botz L. Evaluating aspects of online medication safety in long-term follow-up of 136 Internet pharmacies: illegal rogue online pharmacies flourish and are long-lived. J Med Internet Res 2013 Sep 10;15(9):e199 [FREE Full text] [doi: 10.2196/jmir.2606] [Medline: 24021777]
11. The Center for Safe Internet Pharmacies. The Internet Pharmacy Market in 2016: Trends, Challenges and Opportunities. The Center for Safe Internet Pharmacies. 2016. URL: https://safemedsonline.org/wp-content/uploads/2016/01/The-Internet-Pharmacy-Market-in-2016.pdf [accessed 2022-10-24]
12. National Association of Boards of Pharmacy. Internet Drug Outlet Identification Program: Progress Report for State and Federal Regulators Internet. National Association of Boards of Pharmacy. 2017 Aug. URL: https://nabp.pharmacy/wp-content/uploads/2016/08/Internet-Drug-Outlet-Report-August-2017.pdf [accessed 2022-10-24]
13. Shahzad A, Mahdin H, Mohd N. An Improved Framework for Content-based Spamdexing Detection. IJACSA 2020;11(1):409-420 [FREE Full text] [doi: 10.14569/ijacsa.2020.0110151]
14. McCoy D, Pitsillidis A, Jordan G, Weaver N, Kreibich C, Krebs B, et al. PharmaLeaks: Understanding the business of online pharmaceutical affiliate programs. In: Security'12: Proceedings of the 21st USENIX conference on Security symposium. Berkeley, CA: USENIX Association; 2012 Aug Presented at: Security'12: The 21st USENIX conference on Security symposium; August 8-10, 2012; Bellevue, WA p. 1.
15. Leontiadis N, Moore T, Christin N. Measuring and Analyzing Search-Redirection Attacks in the Illicit Online Prescription Drug Trade. In: SEC'11: Proceedings of the 20th USENIX conference on Security. Berkeley, CA: USENIX Association; 2011 Aug Presented at: SEC'11: The 20th USENIX conference on Security; August 8-12, 2011; San Francisco, CA p. 1. [doi: 10.5555/2028067.2028086]
16. Leontiadis N, Moore T, Christin N. A Nearly Four-Year Longitudinal Study of Search-Engine Poisoning. In: CCS '14: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. New York, NY: ACM; 2014 Nov Presented at: CCS'14: 2014 ACM SIGSAC Conference on Computer and Communications Security; November 3-7, 2014; Scottsdale, AZ p. 930-941. [doi: 10.1145/2660267.2660332]
17. Wang D, Savage S, Voelker G. Cloak and dagger: dynamics of web search cloaking. In: CCS '11: Proceedings of the 18th ACM conference on Computer and communications security. New York, NY: ACM Press; 2011 Oct Presented at: CCS'11: The ACM Conference on Computer and Communications Security; October 17-21, 2011; Chicago, IL p. 477-490. [doi: 10.1145/2046707.2046763]
18. Madeira CR, Tonin FS, Fachi MM, Borba HH, Ferreira VL, Leonart LP, et al. Efficacy and safety of oral phosphodiesterase 5 inhibitors for erectile dysfunction: a network meta-analysis and multicriteria decision analysis. World J Urol 2021 Mar 09;39(3):953-962. [doi: 10.1007/s00345-020-03233-9] [Medline: 32388784]

XSL•FO
RenderX

19. Ayta IA, McKinlay JB, Krane RJ. The likely worldwide increase in erectile dysfunction between 1995 and 2025 and some possible policy consequences. BJU Int 1999 Jul;84(1):50-56. [doi: 10.1046/j.1464-410x.1999.00142.x] [Medline: 10444124]
20. Ahmed J, Modica de Mohac L, Mackey TK, Raimi-Abraham BT. A critical review on the availability of substandard and falsified medicines online: Incidence, challenges and perspectives. J Med Access 2022 Feb 15;6:23992026221074548 [FREE Full text] [doi: 10.1177/23992026221074548] [Medline: 36204527]
21. Orizio G, Schulz P, Domenighini S, Caimi L, Rosati C, Rubinelli S, et al. Cyberdrugs: a cross-sectional study of online pharmacies characteristics. Eur J Public Health 2009 Aug;19(4):375-377. [doi: 10.1093/eurpub/ckn146] [Medline: 19151103]
22. Beus J. Why (almost) everything you knew about Google CTR is no longer valid Internet. sistrix. 2020. URL: https://www.sistrix.com/blog/why-almost-everything-you-knew-about-google-ctr-is-no-longer-valid/ [accessed 2022-10-24]
23. Gephi - The Open Graph Viz Platform. Gephi. URL: https://gephi.org [accessed 2022-10-24]
24. LegitScript. Website Status Checker Internet. LegitScript. URL: https://www.legitscript.com/ [accessed 2022-10-24]
25. Traffic Analytics Internet. SEMrush. URL: https://www.semrush.com/analytics/traffic/ [accessed 2022-10-24]
26. Fittler A, Adeniye L, Katz Z, Bella R. Effect of Infodemic Regarding the Illegal Sale of Medications on the Internet: Evaluation of Demand and Online Availability of Ivermectin during the COVID-19 Pandemic. Int J Environ Res Public Health 2021 Jul 13;18(14):7475 [FREE Full text] [doi: 10.3390/ijerph18147475] [Medline: 34299920]
27. Corona I, Contini M, Ariu D, Giacinto G, Roli F, Lund M, et al. PharmaGuard: Automatic identification of illegal search-indexed online pharmacies. New York, NY: IEEE; 2015 Presented at: 2015 IEEE 2nd International Conference on Cybernetics (CYBCONF); June 24-26, 2015; Gdynia, Poland p. 324-329. [doi: 10.1109/cybconf.2015.7175954]
28. John PJ, Yu F, Xie Y, Krishnamurthy A, Abadi M. deSEO: Combating Search-Result Poisoning. Berkeley, CA: USENIX; 2011 Aug Presented at: 20th USENIX Security Symposium; August 8-12, 2011; San Francisco, CA URL: https://www.usenix.org/legacy/events/sec11/tech/full_papers/John.pdf
29. Lee KS, Yee SM, Zaidi STR, Patel RP, Yang Q, Al-Worafi YM, et al. Combating Sale of Counterfeit and Falsified Medicines Online: A Losing Battle. Front Pharmacol 2017;8:268. [doi: 10.3389/fphar.2017.00268] [Medline: 28559845]
30. Thousands of fake online pharmacies shut down in INTERPOL operation. Interpol. 2021. URL: https://www.interpol.int/en/News-and-Events/News/2021/Thousands-of-fake-online-pharmacies-shut-down-in-INTERPOL-operation [accessed 2022-10-24]

Abbreviations

API: active pharmaceutical ingredient
CTR: click-through rate
EU: European Union
PDE5: phosphodiesterase type 5
SEO: search engine optimization
SER: search engine result
SERP: search engine results page

Edited by G Eysenbach; submitted 23.04.22; peer-reviewed by W Ceron, H Ayatollahi, A Hamed, C Weerth; comments to author 18.07.22; revised version received 31.07.22; accepted 12.08.22; published 08.11.22.

Please cite as:
Fittler A, Paczolai P, Ashraf AR, Pourhashemi A, Iványi P
Prevalence of Poisoned Google Search Results of Erectile Dysfunction Medications Redirecting to Illegal Internet Pharmacies: Data Analysis Study
J Med Internet Res 2022;24(11):e38957
URL: https://www.jmir.org/2022/11/e38957
doi:10.2196/38957
PMID:36346655

©András Fittler, Péter Paczolai, Amir Reza Ashraf, Amir Pourhashemi, Péter Iványi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding the Social Mechanism of Cancer Misinformation Spread on YouTube and Lessons Learned: Infodemiological Study

Ho Young Yoon[1*], PhD; Kyung Han You[2*], PhD; Jung Hye Kwon[3,4,5], MD, PhD; Jung Sun Kim[3], MD, MSc; Sun Young Rha[6,7], MD, PhD; Yoon Jung Chang[8], MD, PhD; Sang-Cheol Lee[9], MD, PhD

[1]Division of Communication and Media, Ewha Womans University, Seoul, Republic of Korea

[2]Department of Media and Communication Studies, Jeonbuk National University, Jeonju, Republic of Korea

[3]Division of Hemato-Oncology, Department of Internal Medicine, Chungnam National University Sejong Hospital, Sejong-Si, Republic of Korea

[4]Division of Hematology and Oncology, Department of Internal Medicine, College of Medicine, Chungnam National University, Daejeon, Republic of Korea

[5]Daejeon Regional Cancer Center, Daejeon, Republic of Korea

[6]Division of Medical Oncology, Yonsei University College of Medicine, Seoul, Republic of Korea

[7]Yonsei Cancer Center, Yonsei University Health System, Seoul, Republic of Korea

[8]Division of Cancer Control and Policy, National Cancer Center Korea, Department of Cancer Control and Population Health, National Cancer Center Graduate School of Cancer Science and Policy, Goyang, Republic of Korea

[9]Division of Hematology and Oncology, Department of Internal Medicine, Soonchunhyang University Hospital Cheonan, Cheonan, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jung Hye Kwon, MD, PhD
Division of Hemato-Oncology
Department of Internal Medicine
Chungnam National University Sejong Hospital
20, Bodeum 7-Ro
Sejong-Si, 30099
Republic of Korea
Phone: 82 44 995 4781
Email: Kwonjhye.onco@gmail.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/11/e44334/

Abstract

Background: A knowledge gap exists between the list of required actions and the action plan for countering cancer misinformation on social media. Little attention has been paid to a social media strategy for disseminating factual information while also disrupting misinformation on social media networks.

Objective: The aim of this study was to, first, identify the spread structure of cancer misinformation on YouTube. We asked the question, “How do YouTube videos play an important role in spreading information about the self-administration of anthelmintics for dogs as a cancer medicine for humans?” Second, the study aimed to suggest an action strategy for disrupting misinformation diffusion on YouTube by exploiting the network logic of YouTube information flow and the recommendation system. We asked the question, “What would be a feasible and effective strategy to block cancer misinformation diffusion on YouTube?”

Methods: The study used the YouTube case of the self-administration of anthelmintics for dogs as an alternative cancer medicine in South Korea. We gathered Korean YouTube videos about the self-administration of fenbendazole. Using the YouTube application programming interface for the query “fenbendazole,” 702 videos from 227 channels were compiled. Then, videos with at least 50,000 views, uploaded between September 2019 and September 2020, were selected from the collection, resulting in 90 videos. Finally, 10 recommended videos for each of the 90 videos were compiled, totaling 573 videos. Social network visualization for the recommended videos was used to identify three intervention strategies for disrupting the YouTube misinformation network.

XSL•FO
RenderX

Results: The study found evidence of complex contagion by human and machine recommendation systems. By exposing stakeholders to multiple information sources on fenbendazole self-administration and by linking them through a recommendation algorithm, YouTube has become the perfect infrastructure for reinforcing the belief that fenbendazole can cure cancer, despite government warnings about the risks and dangers of self-administration.

Conclusions: Health authorities should upload pertinent information through *multiple* channels and should exploit the existing YouTube recommendation algorithm to disrupt the misinformation network. Considering the viewing habits of patients and caregivers, the direct use of YouTube hospital channels is more effective than the indirect use of YouTube news media channels or government channels that report public announcements and statements. Reinforcing through multiple channels is the key.

(J Med Internet Res 2022;24(11):e39571) doi:10.2196/39571

KEYWORDS

cancer misinformation; social media health misinformation; fenbendazole; self-administration; complex contagion; YouTube; social media factual information delivery strategy

Introduction

Cancer Misinformation and Social Media

Cancer misinformation on social media has piqued the interest of health care professionals. Misinformation is defined as information that causes people to "hold inaccurate factual beliefs" [1]. Studies found that nearly one-third of cancer content on social media was misinformation [2,3] and that 76.9% of it contained harmful information [3]. Cancer misinformation can significantly reduce a patient's chances of survival. It may lead to a delay in receiving appropriate treatment, the pursuit of unproven alternative therapies, or the rejection of the currently prescribed treatment [2]. Beyond the disintegration of trust between patients, carers, and physicians, the detrimental impacts of disinformation extend to psychological and mental health [2-5].

Indeed, misinformation is one of the most serious negative consequences of social media. False news spreads much "further, faster, deeper, and more broadly" than the truth on social media [6]. The novelty of misinformation grabs the immediate attention of humans, resulting in human-mediated message delivery. The echo chamber induced by the recommendation algorithm amplifies the spread of inaccurate information among individuals with similar interests [7]. Given that caregivers and patients' family members frequently associate the information community with social support and develop a strong interest in new treatment information [8-10], social media provides the optimal environment for information dissemination, combining human and nonhuman algorithm-driven exposure and connections. This is of particular concern for physicians, as several studies have shown that people are more receptive to misinformation than accurate information [2,4,10].

Thus, it is imperative that public health agencies and health professionals advocate for proactive measures to offset the detrimental consequences of social media misinformation [10-13]. Numerous solutions have been proposed, including digital literacy education [4], accurate information provision [11,14], and media campaigning [15]. While these solutions seek to provide "true" information, rarely have studies been done on concrete and effective ways to disrupt the flow of misinformation on social media, particularly based on an analysis of the current social media information flow network. As a result, a knowledge gap exists between policy implementation and the list of required actions [16].

This may be attributed to the communication pattern of cancer misinformation. Misinformation is frequently mixed with true information in everyday health communication [17]. This mode of communication has rendered it incapable of scrutinizing new information, because information is not only the mixture of true and false but of old and new. As a result, studying the spread of cancer misinformation requires a clearly visible case that can track and reconstruct the pattern of communication.

Complex Contagion of Health Behaviors in Social Media

Adoption of behaviors frequently manifests as a complex contagion: the spread of collective behaviors requiring contact with multiple sources of activation [18]. Complex contagion occurs when social reinforcement influences the transmission of behaviors, beliefs, and attitudes. Unlike simple contagion, in which exposure results in immediate transmission, complex contagion requires social legitimation, credibility, and behavioral complementarity due to uncertainty [18,19]. Studies found that adopting health-related behaviors, such as smoking, exercise, and antivaccine beliefs, are proven to follow the complex contagion mechanism [20-22]. The key point here is not repeated exposure to a single source, but the exposure to multiple sources for social confirmation and reinforcement [18].

Studies have discovered that social media is one of the perfect environments for observing complex contagion [23-25]. Clustered communities in social media serve as peer-to-peer communication and information dissemination networks [25-27]. Social media research frequently reveals patterns of shared exposure to common stimuli [27], but not all patterns are identical. If complex contagion is viewed as the direct link between social media users, it is social cohesion that binds people together in a networked group, often seen in Facebook [28,29]. If the contagion is based on exposure between individuals, regardless of connection between users, then it is contagion via the network structure that enables access to the same source of exposure, typically found in YouTube through its recommendation algorithm [30,31]. The continuous flow of relevant and engaging YouTube videos, as well as the linking of video content via YouTube's recommendation system [17,32],

XSL•FO
RenderX

creates the environment conducive to social reinforcement, a necessary requirement for complex contagion.

Research Context: Fenbendazole Self-administration on YouTube

On September 3, 2019, a Korean-language YouTube video portraying Joe Tippens, who claimed to be entirely cured of his cancer after self-administration of fenbendazole, was uploaded. This video had a total of 2.4 million view counts and 33,702 likes as of September 21, 2021, and was an instant hit among cancer patients and caregivers in South Korea [33]. On September 20, 2019, news outlets reported that pharmacies were out of stock of fenbendazole tablets [34]. According to the Korean government, fenbendazole sales totaled approximately 229,000 tablets in September 2019, which was 5 times the quantity sold from January to August 2019. In November 2019, around 403,000 tablets were sold [35]. The media also grabbed public attention by reporting about a patient with cancer who was a comedian who publicly proclaimed that he would disclose his status following self-administration of fenbendazole. The Ministry of Food and Drug Safety of South Korea issued a caution on the use of fenbendazole for cancer treatment on October 28, 2019 [36]. The National Cancer Center Korea indicated in January 2020 that no clinical trials with fenbendazole were planned, adding that “it is not worthwhile to pursue” [37]. However, YouTube videos describing personal experiences with self-administration of fenbendazole were continuously released across patient and caregiver communities

Research Questions

Given the research context, this study attempted to address the following research questions: (1) What are the characteristics of the YouTube cancer misinformation network regarding the self-administration of fenbendazole? More specifically, (2) is there evidence to support the complex contagion process in social media? (3) Does the networking pattern of the YouTube misinformation network for fenbendazole provide us with useful insights as to why the conventional campaign of disrupting misinformation through news media is less effective? (4) Is there a structural separation of the network between news media providers and user-generated content?

Methods

Study Design

We have undertaken unstructured data analysis by collecting data from YouTube. Following the collection of YouTube video data, the data were analyzed using timeline trend analysis, content analysis, and network analysis. These are common strategies applied to unstructured big data analysis. Indeed, mining unstructured online data and analyzing them to discover patterns is a basic technique for big data analytics [38]. The detailed process of data collection, network data processing, and data analysis is depicted in Multimedia Appendix 1.

YouTube Data Collection

We used Google’s YouTube application programming interface (API) to retrieve a list of search query and recommended videos [39] by using a Python program. The API provides the meta information, such as a video title, channel name, the date and time of upload, and number of views, likes, and comments [40].

Our data collection strategy was as follows. First, we used the term “fenbendazole” in Korean to download the list of the videos from the YouTube API. Second, videos were included in the analysis if the video views were more than 50,000, uploaded between September 2019 and September 2020. The number 50,000 has been frequently used in the literature to filter relevant and popular YouTube videos [41,42]. Third, we compiled a list of the top 10 recommended videos for each video we searched. Figure 1 depicts the recommended video interface. The top 10 is the number that has been used in the literature for examining recommendation effects [43,44]. Note that this list was not the same list as our API search query for fenbendazole. YouTube’s recommendation algorithm analyzes viewers’ data regarding their viewing habits and uses the data to make recommendations. This does not imply that the YouTube API reflects the preferences of the API key holder. The YouTube API returns the data to the developer that matches the query parameter, such as video, channel, and playlist [39]. It is known that the YouTube API offers results based on popularity rather than individual user desires [45].

Figure 1. YouTube recommended video interface. Recommended videos are highlighted on the right side. For privacy protection, the screenshot is blurred.

XSL•FO
RenderX

Video Content Analysis

To comprehend the content of the collected fenbendazole YouTube videos, we manually examined the videos that met our selection criteria. Two PhD media researchers participated as coders. The two coders have never viewed fenbendazole YouTube videos. Detailed information about the YouTube videos, such as channel name, video title, and view count, was provided as well as the YouTube video URLs.

The coders coded each video's channel holders and the viewpoint or attitude toward the self-administration of fenbendazole for cancer treatment. Prior to the coders watching the videos, the channel holders were coded as individuals or institutions. The coders selected subcategory codes for each type of holder after viewing a few videos and engaging in discussion with the research team. We coded individuals as patients, health care professionals, caregivers, and others. The coders further coded whether the holder self-administered fenbendazole if the holder was a patient. Regarding institutions, we categorized the holders as news media, hospitals, and others.

We classified the viewpoint or attitude toward the self-administration of fenbendazole into four subcategories:

1. The "positive" subcategory code was used when the account holder mentioned that she had self-administered fenbendazole or had recommended or sometimes guided others on how to take fenbendazole without physician consultation.
2. The "reserved" subcategory code was used when the account holder mentioned that doctor consultation for fenbendazole self-administration was necessary or that fenbendazole may only be used for people with specific health conditions, such as patients with terminal-stage cancer or when no treatment options remained.
3. The "neutral" subcategory code was used when the video only introduced the fenbendazole case.
4. The "negative" subcategory code was used when the viewpoint was the opposite of the positive viewpoint (ie, the account holder was against the use of fenbendazole).

Finally, our coders reviewed the comments on the videos in order to understand how viewers responded to the videos and to validate their coding by determining whether viewers interpreted the videos as they had been coded.

Influencer and YouTube Video Recommendation Network

A variety of techniques were used to decipher the misinformation network. First, the timeline of uploaded videos was analyzed, and the influencers, in terms of view count and location in the core of the network, were identified through the network analysis index, *k*-core. Second, network diagrams were drawn in order to locate the influencers in the network. We used the edge list format to convert the list of video searches and recommendations into a network matrix of relational data. In this process, duplicate videos were counted as the value of network ties. The NetDraw program (Analytic Technologies) was then used to draw and to analyze the network data [46]. Although the analysis unit of the collected data was a video, we depicted the network diagram at the channel level for intuitive understanding, as we were particularly interested in who formed the core of the network rather than the role of each video. As the recommendation algorithm of YouTube reflects the quality of video measured by user appreciation, personal preference, and diversity [40], it is logical that the channels at the network's core exert the most influence on information flow. Through *k*-core analysis and multidimensional scaling (MDS), we were able to determine the network core. *k*-core is an index that identifies a highly cohesive region of the entire graph [47]. When the *k*-core property is combined with the MDS property that clusters network nodes with comparable relationships to other nodes, the *k*-core nodes tend to locate at the center of the network (ie, the core of network). As the majority of YouTube videos are viewed as a result of recommendations [44], the core of the recommendation network is the center of the information cascade in the dissemination of misinformation.

Third, we investigated the network by examining the ego networks of institutions' channels, such as hospitals and news organizations, to see if they were in the core of the network. The ego network or egocentric network refers to "a network based on a particular individual...comprised of all the relations that a focal agent has with others" [47]. We think the connection to news media and hospitals in the network is important, as trust in these institutions is related to the spread of misinformation [9]. In addition, the news media serve as fact-checking institutions, and hospitals provide scientific health information to counter misinformation. Thus, a comparison of both institutions' ego networks can help us better comprehend and analyze the networking pattern between traditional media and YouTube viewing habits.

Ethical Considerations

This study used publicly available data, which are exempt from Institutional Review Board (IRB) approval as stated in Article 16 of the Enforcement Rule of Bioethics and Safety Act of South Korea and Article 13 of its Enforcement Decree. The legislation specifies that publicly available data are waived from IRB review, unless they include collection and recording of sensitive personal information regulated by the Personal Information Protection Act of South Korea. This study did not collect any personal information.

Results

Searched and Collected Data

The YouTube API search for the term "fenbendazole" in Korean returned 702 videos from 227 channels. In total, 90 videos received over 50,000 views. The videos with more than 50,000 views accounted for 98% of the total view traffic from 702 videos. Regarding recommended videos, 573 videos were collected because of the recommendation overlaps between the recommendations from 90 videos. The total number of overlapping videos was 36.3% (n=327) of the theoretical maximum number of videos (ie, 90 videos × 10 recommendations each = 900) without any overlap between recommended videos. The overlap percentage between the searched results and the suggested videos was 25.6%, as the suggested videos were not restricted to fenbendazole videos but reflected the users' other viewed subjects.

XSL•FO
RenderX

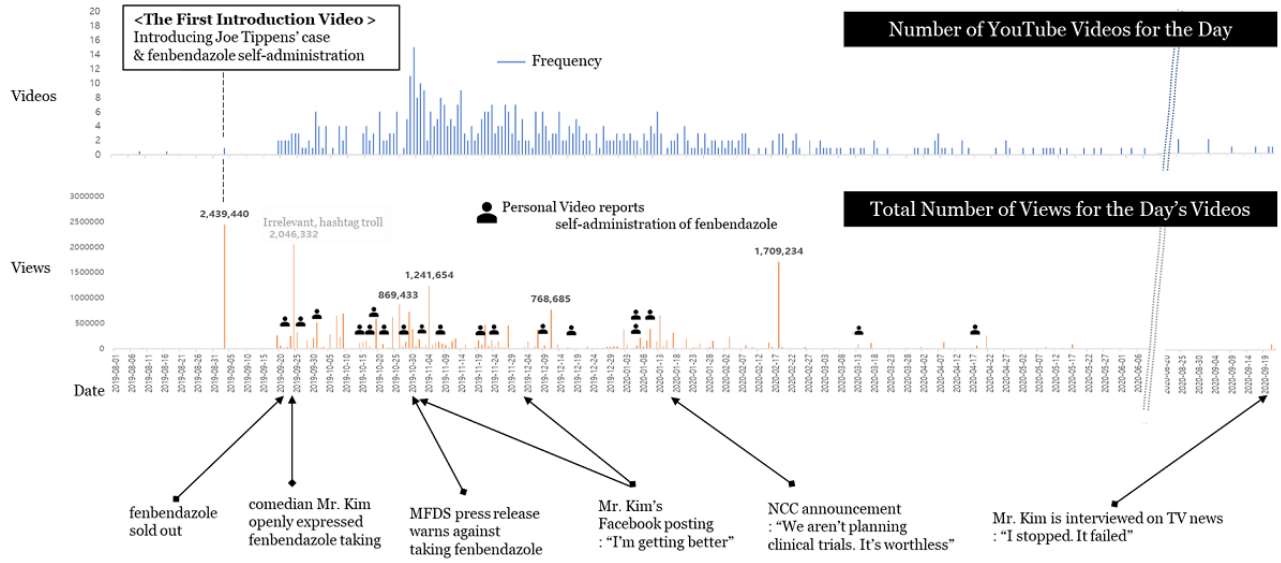
Self-administered Fenbendazole Videos and Opinions From Health Professionals

The number of YouTube videos uploaded for the day and the total number of views for the day's videos are depicted in Figure 2. These videos are the result of an API search query. The graph indicates that a significant number of personal videos talking about self-administration of fenbendazole were uploaded during the first 6 months after Joe Tippens' case was introduced.

Interestingly, individuals who self-administered fenbendazole continued to appear on YouTube following the news reports about fenbendazole selling out. Although the number of personal videos dealing with the body's reaction after self-administration was not large, the consistency of video uploads was of crucial importance; it was not repeated exposure to a single source that confirmed people's perspectives but multiple exposure to diverse sources that confirmed their views. Furthermore, these videos have garnered considerable attention. The top three videos showing responses to self-administered fenbendazole received an average of 215,256 views, which is 4 times the average view count of 50,326 for all 573 videos in the analysis data set. As these videos reported a favorable self-assessment of fenbendazole for pain relief and a fall in tumor markers, the information has virtually become reliable information to viewers. The top 20 most-viewed videos are listed in Table 1; 8 of these are personal testimonials about how fenbendazole improved their symptoms.

In addition, some health professionals on YouTube have not strongly criticized fenbendazole use. This signals to viewers that it would be worth trying because at least it would not be toxic. The comments in these videos referred to these health professionals as "true doctor" [48]. Out of 3 professionals, 2 recommended that patients should consult their doctors before taking fenbendazole (Table 1), but they also described the role of other catalysts, such as vitamins, in aiding in the absorption of the fenbendazole components. As these health professionals are active physicians—one is a physician practicing internal medicine and holds a PhD in chemistry, and the other is the director of the Division of Hematology and Medical Oncology in a general hospital—people interested in the self-administration of fenbendazole interpreted this as a positive signal for taking the medication. For example, the comments by users in the video by internal medicine physicians included the following: "an excellent video to know more details. It helped me making decision between confusing opinions" (ID: ***** Kim, anonymized for privacy protection), "...it was hard to trust physicians but by listening your heartfelt comments, my negative trauma is gone today" (ID: ** tree, anonymized), and "thank you for encouragement as a person taking fenbendazole with colorectal cancer at stage 4" (ID: ***flower, anonymized). One other professional—a clinic owner and radiologist—even advocated for the use of fenbendazole.

Figure 2. The timeline of fenbendazole YouTube videos. MFDS: Ministry of Food and Drug Safety; NCC: National Cancer Center Korea.



XSL•FO
RenderX

Table 1. The top 20 YouTube fenbendazole videos.

Date (year-month-day)	Channel	Description[a]	Title summary	Viewpoint[b]	View count[c], n	Like count[c], n
2019-09-03	World Village Magazine TV	First video about fenbendazole effect	Introduction of Joe Tippens	Positive	2,439,440	33,702
2020-02-18	VIDEOMUG	News media	News Report of Issue	Neutral	1,708,628	11,829
2019-11-04	Dr. Ezra TV	Health professional	Albendazole Possibility	Reservation	1,151,984	37,337
2019-10-26	JIGUIN	Self-administration	Final-Stage Cancer Symptom	Positive	869,433	9196
2019-12-11	Dr. Ezra TV	Health professional	Parasites That Transmit Cancer	Reservation	660,420	26,729
2020-01-13	Changbal Testing TV	Hashtag troller	Parasite Shock Visual	Positive	600,257	9529
2019-10-29	Dr. Ezra TV	Health professional	Personal Thought	Reservation	521,487	16,738
2019-10-07	JIGUIN	Self-administration	Postfenbendazole Personal Review	Positive	477,383	10,320
2020-01-02	KimCell	Patient taking albendazole (2 tablets)	Metastatic Cancer and Fenbendazole	Positive	387,156	—[d]
2019-10-09	Ahn	Self-administration	Postfenbendazole Personal Review	Positive	382,838	6694
2019-10-30	JIGUIN	Self-administration	Postfenbendazole Personal Review	Positive	382,818	9652
2020-01-10	Sanchae Story	Self-administration	Self-administration of Fenbendazole and Cured	Positive	362,256	6909
2019-10-01	JIGUIN	Self-administration	Postfenbendazole Personal Review	Positive	325,478	5562
2019-11-21	Dr. Ezra TV	Health professional	US Government is Testing Fenbendazole	Partially positive or reservation	299,936	13,376
2019-10-09	MitoTV (doctor)	Health professional	Anticancer Anthelmintics?	Positive	278,706	8120
2019-10-05	JIGUIN	Self-administration	Reply to Comments	Positive	278,395	8838
2020-04-21	Segye Ilbo (newspaper)	News media	News Report on Mr. Kim	Neutral	261,328	1534
2019-10-24	JIGUIN	Self-administration	Cancer Expert View	Positive	258,664	7543
2019-11-28	Changbal Testing TV	Hashtag troller	Parasites and Cancer	Not applicable	248,907	7009
2019-09-23	Dr. Nah's Medical Talk	Health professional	Expert View on Fenbendazole	Negative or reservation	239,643	3777

[a]The description category is based on the classification of the channel holder.

[b]The viewpoint category is based on the attitude toward the self-administration of fenbendazole.

[c]The count as of September 21, 2021, the date of data collection.

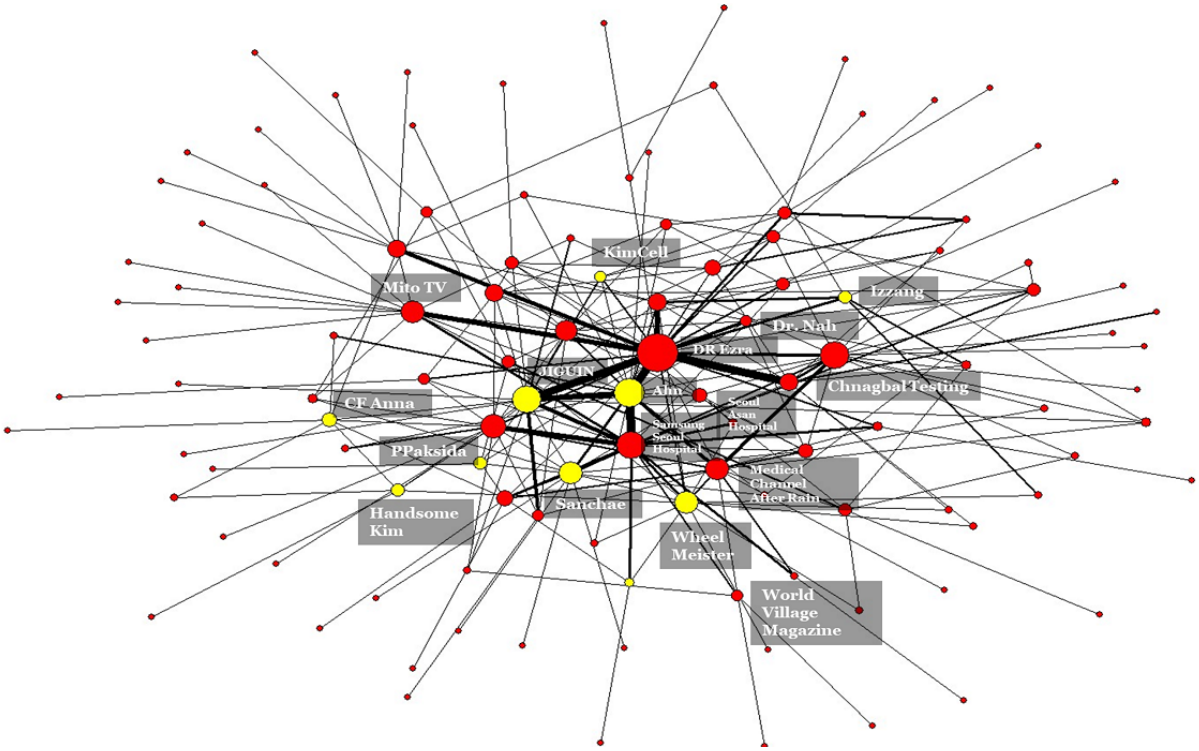
[d]Not reported.

YouTube Recommendation Network: Creating a Spiral Circle of Positivity for Fenbendazole Self-administration

Figure 3 displays the recommendation video network during the first 6-month period; the network during the whole year is displayed in Multimedia Appendix 1. The network demonstrates that personal videos after the self-administration of fenbendazole were continuously posted and connected through the recommendation algorithm.

XSL•FO
RenderX

Figure 3. Fenbendazole YouTube video recommendation network between September 2019 and February 2020. The default node color is red. The yellow nodes are the channels where fenbendazole self-administration was conducted. The node size reflects the degree of a node, namely, the number of connections with other nodes. The tie strength reflects the number of repeated recommendation linkages.



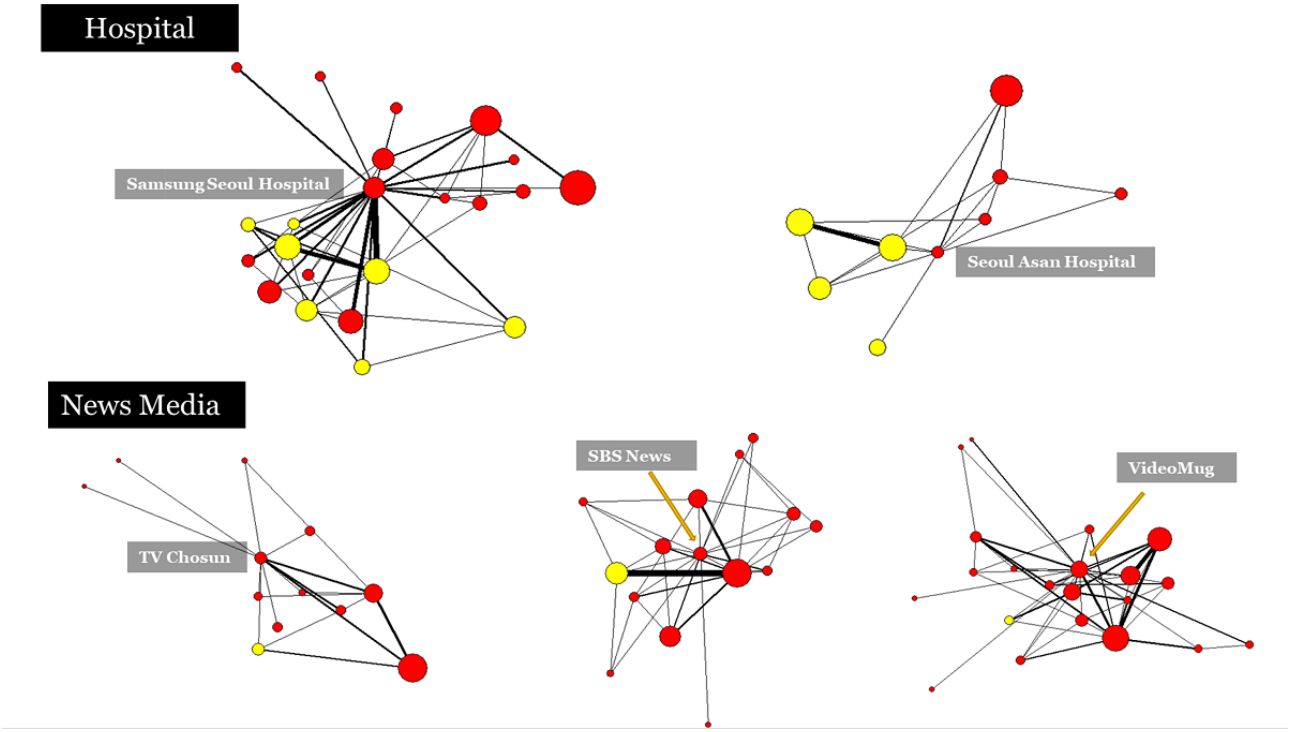
The recommendation network reveals that positive evaluation videos by individual patients and videos by health care professionals synergistically promoted the spread of favorable discourse about fenbendazole. In Figure 3, the first video explaining the positive effect of fenbendazole self-administration on cancer is located outside the network. The central position was taken by one professional video channel—Dr. Ezra TV—which stated that it is preferable for the Korean government to allow people to take fenbendazole in order to collect feasible data rather than warning people, and that it is a sign of capitalism in a capitalist society that a cheap affordable medication would not be invested in for cancer treatment. The Dr. Ezra TV channel was linked to seven video channels about self-administered fenbendazole in a single step, which means that personal videos on the body's reaction to fenbendazole were directly recommended by the YouTube recommendation system. The channels with the top 20–highest view counts were inextricably linked. For example, the Ahn channel and the JIGUIN channel, along with the Dr. Erza TV channel, were a tightly connected network group, namely, a network clique. These two patient channels reported positive outcomes after the self-administration of fenbendazole in terms of pain reduction and positive blood test results; albeit they did acknowledge that it was unknown whether the effect was primarily due to fenbendazole or was accompanied by other treatments and procedures.

By contrast, the network of recommendations did not include any government or authoritative medical channels. Within the network core, there were two hospital channels; however, it appears that the channels just mirrored users' YouTube viewing habits. These channels did not include any content on fenbendazole. In fact, these two hospitals are the top two hospitals in terms of the number of cancer patients they treat. News channels played no role in disseminating "true" pertinent information. This becomes obvious when the recommendation network is enlarged to a 1-year time range (Multimedia Appendix 1) in order to visualize the links between the spiral of information and news media, as the initial 6-month recommendation network did not reveal many news media linkages.

The ego network of two hospitals and three news media outlets is depicted in Figure 4. The top two diagrams demonstrate how hospital channels were immediately connected to numerous video channels about self-administration of fenbendazole via a single recommendation, whereas news channels were rarely connected to these. In other words, even if accurate information is distributed, patients and caregivers engaged in self-administration are unlikely to be connected to government and other authentic YouTube medical channels.

XSL•FO
RenderX

Figure 4. Ego networks of the fenbendazole YouTube recommendation network by institutions between September 2019 and February 2020. The default node color is red. The yellow nodes are the channels where fenbendazole self-administration was conducted. The node size reflects the degree of a node, namely the number of connections with other nodes. The tie strength reflects the number of repeated recommendation linkages.



Discussion

Principal Findings

This study delved into one of the cancer misinformation networks on YouTube. By analyzing the data from searched and recommended videos, we found that personal videos about self-administered fenbendazole were continuously uploaded and accumulated over time as if showing promising evidence for the use of fenbendazole as a cancer treatment. In addition, the recommended content network of fenbendazole has become the infrastructure for confirming the audience's belief and hope in fenbendazole as an alternative cancer medicine. Patients are actively seeking health information over the internet, thereby increasing their self-efficacy in making treatment decisions and altering provider-patient interactions [49-52]. As such, the appearance of supportive professional videos stating that the use of fenbendazole for cancer is scientifically unknown, but possibly helpful, may inspire hope and belief among cancer patients and caregivers who are thinking about self-administering fenbendazole; these videos may also lead patients and caregivers to disregard announcements from the National Cancer Center Korea and the Korean Medical Association.

Unfortunately, the effectiveness of fenbendazole has not been established, and other major adverse effects have been observed [53]. During our investigation, we also found that the YouTube recommendation network was unrelated to credible medical knowledge content. Although some hospital channels were, indeed, connected to the network, it seemed to reflect people's viewing habits rather than topics related to fenbendazole.

In summary, while the YouTube content and recommendation network served as a substantial information source for complex contagion, medical institutions and government entities were excluded from the network, and no dialogue from them was discovered. This resulted in a breakdown of communication between patients and caregivers, resulting in enormous sales of fenbendazole tablets.

Strategies to Fight Social Media Cancer Misinformation

Given YouTube's role as a hub for complex contagion, three strategies to fight against social media cancer misinformation networks are recommended. First, health authorities need to upload a variety of pertinent information through multiple channels. This does not necessarily mean the authorities require multiple channels. They can incorporate existing influencers and other channels. The objective is to have numerous sources of exposure in order to disrupt the cascade of misinformation. A single source will not be sufficient to break the complex contagion dynamic.

Second, it is imperative that health authorities take into account YouTube's recommendation system, current viewing habits, and information flow network between patients and caregivers. As illustrated in Figure 4, prominent hospital channels may be an option, as stakeholders are already engaged in the channels' content. Adding a new channel would be ineffective, as it would need to build an audience from scratch.

Third, relying on the news media does not resolve the issue: health authorities must take an active role in resolving social media misinformation. The news media, on the other hand, is frequently constrained by mechanical objectivity and is prone to report both sides of an argument. Furthermore, those who

XSL•FO
RenderX

follow the news media are not the target audience for the health authorities' message. For example, social media is not the first preference of individuals who intend to learn from the news.

Limitations

While our investigation exposed YouTube's cancer misinformation network, it is not without limits. To begin, the study's data were not collected in real time. The real-time data collection process may have resulted in more concrete information flow dynamics between YouTube videos. However, one of the most difficult areas of health communication is the real-time monitoring system for cancer misinformation. It is difficult to classify misinformation in advance, before it spreads on social media, unless a global censorship system is in place to monitor every conversation. The fenbendazole case in this study, for example, demonstrates how the unexpected introduction of a case from a foreign country, such as that of Joe Tippens, prompted reactions among Korean cancer patients and caregivers.

Second, the study is country specific and focused on fenbendazole as a case study. Although we discovered the insight that the spread of cancer misinformation follows complex contagion logic in social media, more evidence is necessary to augment the study's conclusions. Interviews with people who have taken fenbendazole, in particular, would be helpful in determining the intensity of social media influence.

Third, we concentrated on the videos that had the highest number of views. This excludes individuals who self-administered fenbendazole but received little attention. Fenbendazole was extensively used to the extent that it was sold out at the national level. A future large-scale national-level investigation will further enrich our understanding of social media misinformation.

Conclusions

This study contributes to the body of knowledge by including practice strategies for combating social media cancer misinformation. By studying the content on YouTube, this study has attempted to close the knowledge gap between what to do and how to do it when it comes to delivery of accurate information via social media. The study proposed a way for campaigning against misinformation and educating people by having health policy authorities use the current network of information flow on YouTube. The study focused on involvement with the actual information flow network rather than implementing a conventional one-way communication strategy, even on social media. This action plan will offer valuable information, especially for people who rely solely on online resources, such as social media, and have limited means of accessing expert health knowledge and information. The study's recommendation may not be a comprehensive strategy for combating misinformation, but it may be one of the most successful methods for increasing trust between health care practitioners and stakeholders, such as patients and caregivers.

Acknowledgments

This study was supported by the National R&D Program for Cancer Control through the National Cancer Center (NCC) funded by the Ministry of Health & Welfare, Republic of Korea (HA21C0048).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study design and network diagram.
[PDF File (Adobe PDF File), 370 KB - jmir_v24i11e39571_app1.pdf]

References

1. Kuklinski J, Quirk P, Jerit J, Schwieder D, Rich R. Misinformation and the currency of democratic citizenship. J Polit 2000 Aug;62(3):790-816 [FREE Full text] [doi: 10.1111/0022-3816.00033]
2. Gage-Bouchard EA, LaValley S, Warunek M, Beaupin LK, Mollica M. Is cancer information exchanged on social media scientifically accurate? J Cancer Educ 2018 Dec;33(6):1328-1332 [FREE Full text] [doi: 10.1007/s13187-017-1254-z] [Medline: 28721645]
3. Johnson SB, Parsons M, Dorff T, Moran MS, Ward JH, Cohen SA, et al. Cancer misinformation and harmful information on Facebook and other social media: A brief report. J Natl Cancer Inst 2022 Jul 11;114(7):1036-1039 [FREE Full text] [doi: 10.1093/jnci/djab141] [Medline: 34291289]
4. Warner EL, Waters AR, Cloyes KG, Ellington L, Kirchhoff AC. Young adult cancer caregivers' exposure to cancer misinformation on social media. Cancer 2021 Apr 15;127(8):1318-1324 [FREE Full text] [doi: 10.1002/cncr.33380] [Medline: 33368215]
5. Miller TT, Maurer SH, Felker JT. Searching for a cure on Facebook: Patterns of social media use amongst caregivers of children with brain tumors. Cancer Med 2022 Sep 28;11(17):3323-3331 [FREE Full text] [doi: 10.1002/cam4.4693] [Medline: 35343641]
6. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]

XSL•FO
RenderX

7. Törnberg P. Echo chambers and viral misinformation: Modeling fake news as complex contagion. PLoS One 2018;13(9):e0203958 [FREE Full text] [doi: 10.1371/journal.pone.0203958] [Medline: 30235239]
8. Taylor J, Pagliari C. The social dynamics of lung cancer talk on Twitter, Facebook and Macmillan.org.uk. NPJ Digit Med 2019;2:51 [FREE Full text] [doi: 10.1038/s41746-019-0124-y] [Medline: 31304397]
9. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
10. Himelboim I, Han JY. Cancer talk on Twitter: Community structure and information sources in breast and prostate cancer social networks. J Health Commun 2014;19(2):210-225. [doi: 10.1080/10810730.2013.811321] [Medline: 24111482]
11. Mollica MA, Chou WS, Tonorezos ES, Smith AW. Young adult caregivers' perceptions of cancer misinformation on social media: Response to Warner et al. Cancer 2021 Apr 15;127(8):1177-1179 [FREE Full text] [doi: 10.1002/cncr.33382] [Medline: 33368152]
12. Peterson JS, Swire-Thompson B, Johnson SB. What is the alternative? Responding strategically to cancer misinformation. Future Oncol 2020 Sep;16(25):1883-1888. [doi: 10.2217/fon-2020-0440] [Medline: 32564627]
13. Vraga EK, Bode L. Correction as a solution for health misinformation on social media. Am J Public Health 2020 Oct;110(S3):S278-S280. [doi: 10.2105/ajph.2020.305916]
14. Biancovilli P, Makszin L, Csongor A. Breast cancer on social media: A quali-quantitative study on the credibility and content type of the most shared news stories. BMC Womens Health 2021 May 15;21(1):202 [FREE Full text] [doi: 10.1186/s12905-021-01352-y] [Medline: 33992111]
15. Conley C, Otto A, McDonnell G, Tercyak K. Multiple approaches to enhancing cancer communication in the next decade: Translating research into practice and policy. Transl Behav Med 2021 Nov 30;11(11):2018-2032. [doi: 10.1093/tbm/ibab089] [Medline: 34347872]
16. Sylvia Chou W, Gaysynsky A. A prologue to the special issue: Health misinformation on social media. Am J Public Health 2020 Oct;110(S3):S270-S272. [doi: 10.2105/ajph.2020.305943]
17. Tang L, Fujimoto K, Amith M, Cunningham R, Costantini RA, York F, et al. "Down the rabbit hole" of vaccine misinformation on YouTube: Network exposure study. J Med Internet Res 2021 Jan 05;23(1):e23262 [FREE Full text] [doi: 10.2196/23262] [Medline: 33399543]
18. Centola D, Macy M. Complex contagions and the weakness of long ties. Am J Sociol 2007 Nov;113(3):702-734. [doi: 10.1086/521848]
19. Guilbeault D, Becker J, Centola D. Complex contagions: A decade in review. In: Lehmann S, Ahn YY, editors. Complex Spreading Phenomena in Social Systems. Cham, Switzerland: Springer; 2018:3-25.
20. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008 May 22;358(21):2249-2258 [FREE Full text] [doi: 10.1056/NEJMsa0706154] [Medline: 18499567]
21. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: A randomized controlled trial. Prev Med Rep 2016 Dec;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
22. Salathé M, Bonhoeffer S. The effect of opinion clustering on disease outbreaks. J R Soc Interface 2008 Dec 06;5(29):1505-1508 [FREE Full text] [doi: 10.1098/rsif.2008.0271] [Medline: 18713723]
23. Borge-Holthoefer J, Banos RA, Gonzalez-Bailon S, Moreno Y. Cascading behaviour in complex socio-technical networks. J Complex Netw 2013 Apr 22;1(1):3-24. [doi: 10.1093/comnet/cnt006]
24. Bakshy E, Rosenn I, Marlow C, Adamic L. The role of social networks in information diffusion. In: Proceedings of the 21st International Conference on World Wide Web. 2012 Presented at: The 21st International Conference on World Wide Web; April 16-20, 2012; Lyon, France p. 519-528. [doi: 10.1145/2187836.2187907]
25. Harrigan N, Achananuparp P, Lim E. Influentials, novelty, and social contagion. Soc Networks 2012 Oct;34(4):470-480. [doi: 10.1016/j.socnet.2012.02.005]
26. Centola D. The spread of behavior in an online social network experiment. Science 2010 Sep 03;329(5996):1194-1197. [doi: 10.1126/science.1185231] [Medline: 20813952]
27. Centola D. Social media and the science of health behavior. Circulation 2013 May 28;127(21):2135-2144. [doi: 10.1161/CIRCULATIONAHA.112.101816] [Medline: 23716382]
28. Friedkin NE. Social cohesion. Annu Rev Sociol 2004 Aug 01;30(1):409-425. [doi: 10.1146/annurev.soc.30.012703.110625]
29. Friedkin NE. Structural cohesion and equivalence explanations of social homogeneity. Sociol Methods Res 2016 Jun 30;12(3):235-261. [doi: 10.1177/0049124184012003001]
30. Burt RS. Cohesion versus structural equivalence as a basis for network subgroups. Sociol Methods Res 2016 Jun 29;7(2):189-212. [doi: 10.1177/004912417800700205]
31. Burt RS. Social contagion and innovation: Cohesion versus structural equivalence. Am J Sociol 1987 May;92(6):1287-1335. [doi: 10.1086/228667]
32. Park SJ, Lim YS, Park HW. Comparing Twitter and YouTube networks in information diffusion: The case of the “Occupy Wall Street” movement. Technol Forecast Soc Change 2015 Jun;95:208-217. [doi: 10.1016/j.techfore.2015.02.003]

XSL•FO
RenderX

33. Nam JH. Dog anthelmintic sells out after rumors of cancer treatment...Food and Drug Administration warns "Do not take" [Document in Korean]. SBS News. 2019. URL: https://news.sbs.co.kr/news/endPage.do?news_id=N1005449578&plink=ORI [accessed 2022-02-20]
34. Choi H. Unverified YouTube clip: Cancer is cured with dog anthelmintic...sold out [Document in Korean]. JTBC News. 2019. URL: https://news.jtbc.joins.com/article/article.aspx?news_id=NB11882980 [accessed 2022-02-20]
35. Kang S. Animal anthelmintic "fenbendazole" has no anticancer effect, but still popular [Document in Korean]. Hit News. 2020. URL: http://www.hitnews.co.kr/news/articleView.html?idxno=30389 [accessed 2022-02-20]
36. Anthelmintic for animals. It's only allowed for animals [Document in Korean]. Ministry of Food and Drug Safety. 2019 Oct 28. URL: https://www.mfds.go.kr/brd/m_99/down.do?brd_id=ntc0021&seq=43763&data_tp=A&file_seq=2 [accessed 2022-03-02]
37. Lee E. Anthelmintic for dogs as anticancer medicine – The National Cancer Center says, "it's not worth it to conduct clinical trials" [Document in Korean]. The JoongAng. 2020. URL: https://www.joongang.co.kr/article/23678676#home [accessed 2022-03-03]
38. Das TK, Kumar PM. Big data analytics: A framework for unstructured data analysis. Int J Eng Sci Technol 2013;5(1):155-158 [FREE Full text]
39. Data API. Search: list. YouTube. 2020. URL: https://developers.google.com/youtube/v3/docs/search/list [accessed 2021-03-30]
40. Davidson J, Liebald B, Liu J, Nandy P, Van Vleet T, Gargi U, et al. The YouTube video recommendation system. In: Proceedings of the 4th ACM conference on Recommender Systems. 2010 Presented at: The 4th ACM conference on Recommender Systems; September 26-30, 2010; Barcelona, Spain p. 293-296. [doi: 10.1145/1864708.1864770]
41. Burgess J, Matamoros-Fernández A. Mapping sociocultural controversies across digital media platforms: One week of #gamergate on Twitter, YouTube, and Tumblr. Commun Res Pract 2016 Apr 25;2(1):79-96. [doi: 10.1080/22041451.2016.1155338]
42. Aggrawal N, Arora A, Anand A, Irshad MS. View-count based modeling for YouTube videos and weighted criteria–based ranking. In: Ram M, Davim JP, editors. Advanced Mathematical Techniques in Engineering Sciences. Boca Raton, FL: CRC Press; 2018.
43. Papadamou K, Papasavva A, Zannettou S, Blackburn J, Kourtellis N, Leontiadis I, et al. Disturbed YouTube for kids: Characterizing and detecting inappropriate videos targeting young children. In: Proceedings of the 14th International AAAI Conference on Web and Social Media. 2020 Presented at: The 14th International AAAI Conference on Web and Social Media; June 8-11, 2019; Virtual p. 522-533 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7320
44. Zhou R, Khemmarat S, Gao L. The impact of YouTube recommendation system on video views. In: Proceedings of the 10th Annual Conference on Internet Measurement. 2010 Presented at: The 10th Annual Conference on Internet Measurement; November 1-30, 2010; Melbourne, Australia p. 404-410. [doi: 10.1145/1879141.1879193]
45. Bärtl M. YouTube channels, uploads and views. Convergence 2018 Jan 10;24(1):16-32. [doi: 10.1177/1354856517736979]
46. Borgatti SP, Everett MG, Freeman LC. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies; 2002. URL: https://sites.google.com/site/ucinetsoftware/home [accessed 2022-10-26]
47. Scott J. Social Network Analysis. 4th edition. London, UK: SAGE Publications Ltd; 2017.
48. Lee H. "The controversy of YouTube doctors – they mentioned how to self-administer fenbendazole" [Document in Korean]. Young Doctors. 2019 Nov 18. URL: https://www.docdocdoc.co.kr/news/articleView.html?idxno=1074269 [accessed 2022-08-15]
49. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: A qualitative study. BMJ 2004 Mar 06;328(7439):564 [FREE Full text] [doi: 10.1136/bmj.328.7439.564] [Medline: 15001506]
50. McMullan M. Patients using the internet to obtain health information: How this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
51. Lleras de Frutos M, Casellas-Grau A, Sumalla EC, de Gracia M, Borràs JM, Ochoa Arnedo C. A systematic and comprehensive review of internet use in cancer patients: Psychological factors. Psychooncology 2020 Jan;29(1):6-16. [doi: 10.1002/pon.5194] [Medline: 31385400]
52. Bass SB, Ruzek SB, Gordon TF, Fleisher L, McKeown-Conn N, Moore D. Relationship of internet health information use with patient behavior and self-efficacy: Experiences of newly diagnosed cancer patients who contact the National Cancer Institute's Cancer Information Service. J Health Commun 2006 Mar;11(2):219-236. [doi: 10.1080/10810730500526794] [Medline: 16537289]
53. Oh K. Mr. Kim said "I failed. Never recommend anthelmintics for cancer treatment" [Document in Korean]. Chosun Ilbo. 2020 Sep 22. URL: https://www.chosun.com/culture-life/health/2020/09/22/II7X43CUAZFEHO4BMLQKYCATUU [accessed 2022-03-30]

Abbreviations

API: application programming interface

XSL•FO
RenderX

IRB: Institutional Review Board
MDS: multidimensional scaling

Edited by A Mavragani; submitted 01.06.22; peer-reviewed by L Hao Yen, A Benis, A Singleton; comments to author 12.08.22; revised version received 30.08.22; accepted 20.10.22; published 14.11.22.

Please cite as:
Yoon HY, You KH, Kwon JH, Kim JS, Rha SY, Chang YJ, Lee SC
Understanding the Social Mechanism of Cancer Misinformation Spread on YouTube and Lessons Learned: Infodemiological Study
J Med Internet Res 2022;24(11):e39571
URL: https://www.jmir.org/2022/11/e39571
doi:10.2196/39571
PMID:36374534

©Ho Young Yoon, Kyung Han You, Jung Hye Kwon, Jung Sun Kim, Sun Young Rha, Yoon Jung Chang, Sang-Cheol Lee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of the Hashtag #DataSavesLives on Twitter: Exploratory and Thematic Analysis

Piotr Teodorowski[1*], MA, MRES; Sarah E Rodgers[1*], PhD; Kate Fleming[2*], PhD; Lucy Frith[3*], PhD

[1]Department of Public Health, Policy & Systems, University of Liverpool, Liverpool, United Kingdom

[2]NHS Digital, Liverpool, United Kingdom

[3]Centre for Social Ethics and Policy, University of Manchester, Manchester, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Piotr Teodorowski, MA, MRES
Department of Public Health, Policy & Systems
University of Liverpool
Whelan Building
Liverpool, L69 3GL
United Kingdom
Phone: 44 0151 794 5329
Email: p.teodorowski@liverpool.ac.uk

Abstract

Background: "Data Saves Lives" is a public engagement campaign that highlights the benefits of big data research and aims to establish public trust for this emerging research area.

Objective: This study explores how the hashtag #DataSavesLives is used on Twitter. We focused on the period when the UK government and its agencies adopted #DataSavesLives in an attempt to support their plans to set up a new database holding National Health Service (NHS) users' medical data.

Methods: Public tweets published between April 19 and July 15, 2021, using the hashtag #DataSavesLives were saved using NCapture for NVivo 12. All tweets were coded twice. First, each tweet was assigned a positive, neutral, or negative attitude toward the campaign. Second, inductive thematic analysis was conducted. The results of the thematic analysis were mapped under 3 models of public engagement: deficit, dialogue, and participatory.

Results: Of 1026 unique tweets available for qualitative analysis, discussion around #DataSavesLives was largely positive (n=716, 69.8%) or neutral (n=276, 26.9%) toward the campaign with limited negative attitudes (n=34, 3.3%). Themes derived from the #DataSavesLives debate included ethical sharing, proactively engaging the public, coproducing knowledge with the public, harnessing potential, and gaining an understanding of big data research. The Twitter discourse was largely positive toward the campaign. The hashtag is predominantly used by similar-minded Twitter users to share information about big data projects and to spread positive messages about big data research when there are public controversies. The hashtag is generally used by organizations and people supportive of big data research. Tweet authors recognize that the public should be proactively engaged and involved in big data projects. The campaign remains UK centric. The results indicate that the communication around big data research is driven by the professional community and remains 1-way as members of the public rarely use the hashtag.

Conclusions: The results demonstrate the potential of social media but draws attention to hashtag usage being generally confined to "Twitter bubbles": groups of similar-minded Twitter users.

(J Med Internet Res 2022;24(11):e38232) doi:10.2196/38232

KEYWORDS

consumer involvement; patient participation; stakeholder participation; social media; public engagement; campaign; big data; research; trust; tweets; Twitter; perception; usage; users; data sharing; ethics; community; hashtag

XSL•FO
RenderX

Introduction

Background

Well-established ways for sharing knowledge with the general public by researchers include academic publications, presentations, and media engagement (to name a few). However, previous research has raised concerns that the communication between scientists and the public needs to be more accessible and interactive than traditional engagement activities [1-3]. Public engagement, when it is a 2-way process of sharing, promoting, and disseminating research to the public [4,5], can improve trust between researchers and the public [6]. The growth of social media platforms, such as Twitter, a microblogging platform (up to 280 characters per post) [7], offers a more interactive way to engage with the public and can be particularly useful in promoting engagement around controversial topics. Twitter provides a less formal and more dynamic interaction among its users. Posts (tweets) are open to read for everyone, but only Twitter users can post (tweet) them (but Twitter is free and easy to sign up to). Users can reshare original tweets (retweet) with their audience (followers). Researchers are already active on Twitter to communicate their work as they can reach the public [8], colleagues in their field [9], policymakers, and practitioners [10].

One of the key issues in big data research and one subject to a prolonged public debate is the reuse of medical data for research. Often called big data, it has the potential to provide novel health solutions and improve health inequalities [11,12]. Non(re)use of data can negatively impact health services and research [13]. However, some public members are concerned about how their medical data are stored, controlled, (pseudo)anonymized, and reused [14,15]. Public trust and support are needed for big data projects to continue [16]. However, there remains little public understanding of big data research [14].

"Data Saves Lives" is a public engagement campaign that highlights the benefits of big data research, showing how patient data can be used securely to improve health care [17]. The campaign tries to build trust between researchers and the public. It was started by the University of Manchester's Health eResearch Centre in 2014. Since then, it has expanded outside the United Kingdom, and in 2019, it was launched in Europe. The Data Saves Lives European initiative is a multipartner project led by the European Patients' Forum and the European Institute for Innovation through Health Data [18]. The campaign activities target social media, especially Twitter, using the hashtag #DataSavesLives. Hashtags allow the linkage of all posts on the same subject. Any user can use hashtags on Twitter, and to gain broad coverage, it is recommended to get as many Twitter users as possible to use the hashtag. However, this also means that the hashtag's originators do not control by whom and how it is used. This can lead to highjacking of the hashtag by other users, who may use it for a different purpose than initially intended [19,20].

In 2021, the UK government and its agencies adopted the hashtag #DataSavesLives to support their plans to set up a new national database holding National Health Service (NHS) users' medical data, which could be, in some circumstances, available for sharing with third parties [21]. The idea was driven by the COVID-19 pandemic and the recognition that data have the power to shape and improve health care services [22]. The plan was to collect 55 million patients' pseudonymized data in England to be reused (eg, to support services and research). This received heavy criticism from activists regarding lack of transparency around informed consent and confidentiality [23]. Patients would have only limited time to opt out of the scheme, and their consent was mentioned only once in the initial governmental policy documents [24]. The plan's legality was challenged, and there were concerns that medical professionals would refuse to comply by not sharing their patients' data [25]. Poor communication resulted in public concerns around this new scheme. British media outlets from the *Independent* to the *Daily Mail* described the plan as "controversial" [26,27]. These attitudes were not new, as a similar (but not linked) project was abandoned in the past due to negative public opinion [23,28]. Medical professionals had raised concerns about building trust with the public regarding new government plans. The British Medical Association and the Royal College of General Practitioners called for a better public engagement campaign to alleviate public fears [29]. One and half million people initially opted out of the scheme [30]. The government deferred the deadline for the public to opt out of the new database scheme due to public concerns [31]. Later, the policy was reviewed to discuss building trust with the public further [32]. The new governmental policy was published in June 2022 [33]. In contrast, there have been no such controversies in Europe or the adoption of #DataSavesLives by European public institutions.

Previous studies have explored public perceptions of big data research, but few have examined how online public engagement campaigns could promote the benefits of big data research. One paper discussed #DataSavesLives on Twitter, but its coverage was from September 2016 to August 2017 [34]. Our study expands on previous research and explores how the campaign's hashtag was used when the UK government decided to adopt the hashtag in its campaign strategy. Thus, we cover the period of April-July 2021, when there was an ongoing discussion in news headlines around the newly proposed scheme.

Models of Public Engagement

Science communication as a research area emerges from diverse fields and offers theoretical underpinnings for how researchers can engage with the public [3], where the public is understood as any person in society [35]. We use the terms "public" and "public members" in this paper as people who do not have a background in health care or big data research—laypeople. Three theoretical models of how researchers can engage with the public exist in the literature: deficit, dialogue, and participatory [36,37]. These differ in where they locate researchers or the public in the process of engagement [37].

Deficit Model

The deficit model is the oldest and nowadays heavily criticized model for being too passive a form of communication [35]. It is also known as the knowledge transmission model [38] as it assumes that the public has a limited understanding of the research, and through engagement, researchers can educate the

XSL•FO
RenderX

public and explain the complexity of their work, promoting a researcher-centered model [2,39]. The model theorizes that if the public is not supportive of the ongoing research, researchers only need to explain it better to the public [39,40]. Thus, the underpinning problem is the public's lack of understanding [3]. The weakness of this model is the ongoing need to educate the public, which can be only done through a top-down (and usually 1-way) approach, with researchers giving the public information and telling them how they should understand the issues. Empirical evidence has shown that the deficit model of engagement does not change public views toward science [41].

Dialogue Model

The dialogue model was developed in response to the mistrust the public had in research in general (but particularly in medical research) and the perceived failure and passivity of the deficit model to tackle that challenge successfully [40]. The public and researchers may have different perspectives and can interpret the same things differently [39]. The dialogue model recognises the need for an active exchange between researchers and the public, ensuring 2-way communication [37]. This communication can improve understanding among both groups as they can see different perspectives on the same issue. The dialogue model moves away from researcher-centredness in the communication process and invites public views on the research. Public understanding of science is no longer perceived as limited or inferior to researchers' (as it was in the deficit model), but rather, it is perceived to offer a unique view. The model theorizes that the dialogue can further improve trust if researchers listen and implement public feedback. The public will not only understand the researchers' perspectives better but also be more willing to act upon on their advice [42]. For example, it might be more willing to take a new medicine or participate in research.

Participatory Model

Shifting further the power balance between researcher and public, the participatory model argues for public-centredness in communication. Researchers and the public discuss the research agenda, and in contrast to the dialogue model, they also jointly find solutions. This democratization of the process has been argued to have the potential to improve the quality of information and reaching the public [43]. Both groups have something to gain from this cooperation [37]. In health research, it would come under the definition of public involvement, where work is being done together *with* the public rather than *for* it [44]. Growing research shows that public contributors (eg, lay members) are successfully involved in developing and shaping engagement of health care services [45].

Research Questions

Underpinned by the (deficit, dialogue, and participatory) models of public engagement, this study aims to answer the following research questions:

- How was the hashtag #DataSavesLives used on Twitter as the UK government adapted the hastag in its campaign strategy?
- What were the attitudes toward the campaign among Twitter users using #DataSavesLives?

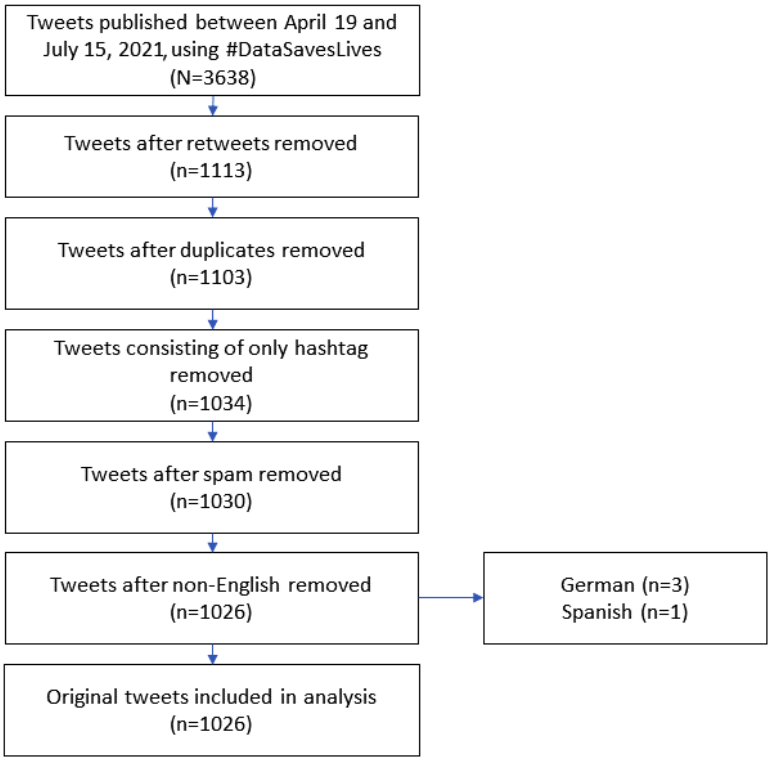
Methods

Data Collection

Tweets were recorded using NCapture software for Google Chrome. This web browser extension collects social media data, such as tweets (including retweets), and imports them directly to NVivo 12 (QSR International) for analysis. Only public tweets from the previous week could be recorded. NCapture does not guarantee that all tweets will be captured at once, as this depends on Twitter; thus, we captured tweets twice per week (Tuesday and Thursday) to get maximum coverage. If an individual tweet is captured twice, NVivo 12 uploads it into the data set only once. Tweets using the hashtag #DataSavesLives were captured for 3 months from April 27 to July 15, 2021. This covered tweets that were posted between April 19 and July 15, 2021. A total of 3638 tweets (including retweets) were collected. We cleaned the data set in NVivo 12 (see Figure 1). All retweets, duplicates, tweets consisting only of hastags, spam, and tweets in languages other than English were removed. After cleaning the data set, 1026 (28.2%) tweets were used in the qualitative analysis. Data saturation was deemed to have been reached. This assumption is based on previous research, which successfully conducted a qualitative analysis of fewer than 1000 tweets and provided novel insights into the online discussion through Twitter hashtags [46-48].

XSL•FO
RenderX

Figure 1. Process of cleaning data sets for qualitative analysis.



Analysis

The analysis was conducted using NVivo 12. NCapture downloaded tweets as 1 data set to NVivo 12 software, and this enabled us to visualize the collected data data.

Descriptive statistics were used to summarise the top 40 user locations, the most active accounts, and the top hashtag used alongside #DataSavesLives and identify the most prominent tweet (based on the number of retweets). We included both tweets and retweets in this analysis to get a broader picture of all Twitter users using the hashtag.

To understand attitudes toward the campaign aims among Twitter users, each original tweet was manually assigned a category as having a positive, neutral, or negative attitude toward the campaign and big data research. The principles and techniques found in content analysis guided this process [49]. We jointly created a short description of each category and then conducted a pilot coding of a sample of tweets during the team meeting. Based on these discussions, an experienced Twitter researcher (author PT) systematically coded all remaining tweets.

Second, we undertook an inductive thematic analysis [50,51]. PT manually coded all tweets, and the team met to identify, review, and refine themes and choose the quotes representing them. Our research team is interdisciplinary, we work in and outside big data research, and 1 author (KF) is based outside the university, ensuring that we have both insider and outsider perspectives. Further analysis was carried out by mapping the thematic analysis results to the public engagement models, which offer insights into how Twitter users used the hashtag #DataSavesLives. Previous research has shown that the engagement techniques can be successfully mapped under these 3 engagement models [36].

Ethical Considerations

The University of Liverpool Ethics Committee (approval no. 9815) granted ethical approval. All captured data are publicly available online. Following established practice [52,53], when we used a direct quote, authors (excluding organizations) were informed and given an option to opt out. No one asked to opt out, and 1 person requested a copy of the published paper. We did not include pictures, links, and emoticons.

Results

Descriptive Statistics

Of all tweets (N=3638) published in this period, the top 40 locations (excluding “unknown”) were from the United Kingdom, showing that the use of the hashtag is still mostly based in the United Kingdom. Other countries included the United States, Australia, Germany, Spain, and Belgium (see Table 1). The discussion was dominated by professionals. Of the 10 most active accounts using the hashtag (which represents n=1746, 48%, of all tweets), all were nonindividual accounts, such as organizations, networks, or public bodies. All public body accounts were linked to the UK’s NHS (see Table 2).

The most prominent tweet had 56 retweets, and it discussed a new webinar on big data research and concerns around data privacy. Some organizations, such as the Health Data Research UK, regularly promoted the benefits of big data research using the hashtag [54].

Most of the hashtags used alongside the campaign were neutral or positive. The top 10 included #healthdata (n=239, 65.8%),

XSL•FO
RenderX

#covid19 (n=134, 3.7%), #nhs (n=102, 2.8%), #ai (n=101, 2.8%), #healtac2021 (n=91, 2.5%), #digitalhealth (n=89, 2.4%) #health (n=88, 2.4%), #testmining (n=84, 2.3%) #research (n=81, 2.2%), and #data (n=65, 1.8%). The negative anticampaign hashtag #DataGrab, which was used by Twitter users accusing the UK government of trying to sell their medical data, appeared 9 times in the whole data set and 5 times in original tweets, thus rarely appearing alongside #DataSavesLives, showing little cross-over between these 2 hashtags.

Table 1. Locations of Twitter users using #DataSavesLives (N=3638 tweets).

Country	Tweets, n
United Kingdom	2247
European Union (including Spain, Germany, and Belgium)	76
United States	56
Australia	44

Table 2. The 10 most active Twitter accounts using #DataSavesLives.

Twitter account	Tweets using #DataSavesLives, n (%)	Type of organization running the account
@hdr_uk	480 (13.2)	Nonprofit organization
@usemydata	353 (9.7)	Nonprofit organization
@nhsx	261 (7.2)	Public body
@nhsdigital	132 (3.6)	Public body
@datasaveslives	125 (3.4)	Nonprofit organization
@apha_analysts	97 (2.7)	Network
@uk_healtex	85 (2.3)	Network
@economics_unit	68 (1.9)	Public body
@medconfidential	66 (1.8)	Campaign group
@pioneer_hub	63 (1.7)	Nonprofit organization

Attitudes

Discussion around #DataSavesLives was largely positive (n=716, 69.8%) or neutral (n=276, 26.9%) toward the campaign. There was some sarcasm in the negative attitudes (n=34, 3.3%) but no dark humor or personal attacks, which has been found in some other Twitter studies. This shows that the debate was generally conducted in a professional fashion, contrary to many politicized social media discussions [28,55,56].

Positive comments included reporting on successful, ongoing, or future projects that had benefitted the public when using big data.

> *The University is partnering with experts from across the UK to launch a £2m data hub for mental health. The hub promises to speed up research into mental health and improve inclusiveness for disadvantaged groups #MentalHealth #DataSavesLives* [EdinburghUni]

This evidence of public benefit can be seen in examples of how big data helped the response to the COVID-19 pandemic.

> *When the pandemic hit in 2020 we urgently looked at whether we could use routine data feeds to produce a more rapid cancer data set that would help quantify the impact of COVID-19 on cancer services. This is one example of how that work is now being used #DataSavesLives @PHE_uk https://t.co/4Eu1QgxXGm* [EllissBrookes]

Twitter users often emphasised how important or relevant was their work around big data research, thus linking it to the campaign's underpinning rationale of showing that the reuse of medical data can change and even, indeed, save people's lives.

> *Our Hubs are working to improve health data so that researchers & innovators are better able to use it to enable discoveries that improve people's lives! #DataSavesLives Find out more: https://t.co/ZKQoaUWSos* [HDR_UK]

Often, organizations would quote stakeholders (eg, public members) to support these statements. There were calls for more public involvement and better data linkage.

Neutral tweets shared job opportunities, information about upcoming conferences, webinars, or new publications and asked people to participate in surveys or studies on big data research.

> *Hear from a super panel of speakers on Tues 25 May 10:00 -11:30 - A researcher's journey to accessing patient data. #datasaveslives #admindata* [SCADR_data]

Negative tweets did not always take issue with the campaign itself but raised concerns about the lack of public trust in the opt-out deadline for the new UK database scheme. Others picked

XSL•FO
RenderX

up on wording used in the hashtag and pointed out that the hashtag only appeals to professionals, not the public, and uses emotions to try to generate public support.

> *It's the wholly presumptuous nature of this scheme that is so abhorrent in my mind #DataSavesLives' the classic 'appeal to emotion' rolled out time and again as dogma in an attempt to upend logic #DataAsAsset is clearly much closer to reality* [griffglen]

Thematic Analysis

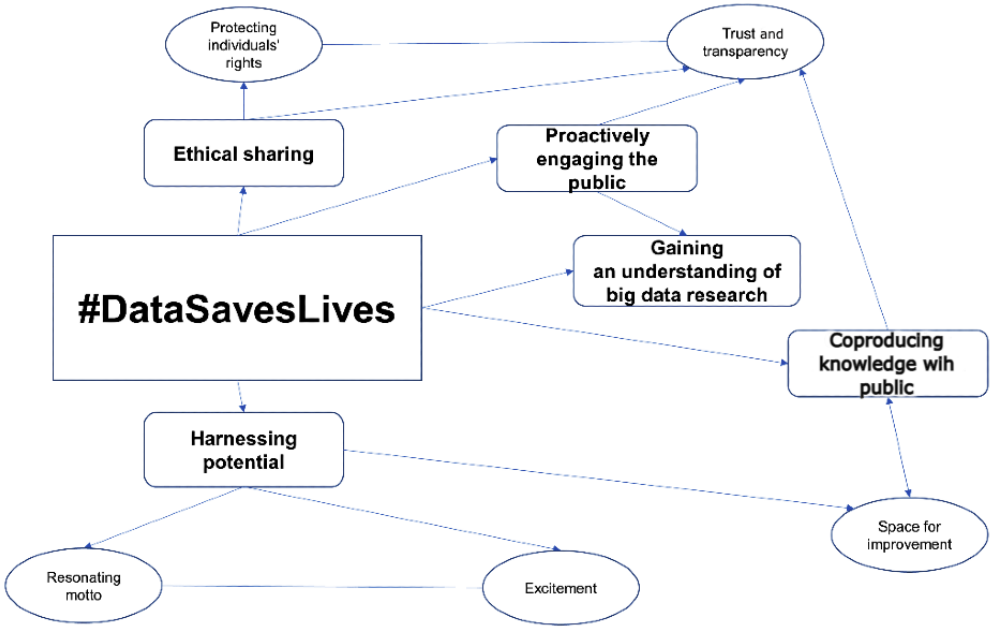
We constructed 5 interlinked themes divided into 5 subthemes (Table 3) to illustrate how the debate around #DataSavesLives appears on Twitter. Figure 2 presents these key connectors and relationships between subthemes. We present the themes under the public engagement models of deficit, dialogue, and participatory.

Table 3. Themes and subthemes derived from the #DataSavesLives debate on Twitter through reflexive thematic analysis.

Themes	Subthemes
Ethical sharing	• Trust and transparency • Protecting individuals' rights
Proactively engaging the public	N/A[a]
Co-producing knowleadge with public	N/A
Harnessing potential	• Excitement • Space for improvement • Resonating motto
Gaining an understanding of big data research	N/A

[a]N/A: not applicable.

Figure 2. Key connectors and relationships between themes and subthemes.



Deficit Model

Harnessing the Potential

Tweet authors on the whole thought that big data has the potential to offer benefits to patients (eg, in the development of new medicines). Harnessing the potential of big data is visible in the following 3 subthemes: excitement, space for improvement, and resonating motto. The COVID-19 pandemic is present here but only as an additional argument for the claim that big data research is helpful for tackling new challenges.

Excitement

Tweet authors were often excited to announce new research projects and share study results (especially when showing how it has made some fundamental change or had the potential for real-life impact). Some of the tweets included authors who are passionate about the subject and others who were excited to participate in new studies. The researchers' success was recognized and noticed by the broader research community (eg, receiving an award). Other tweets refer to upcoming events where authors were publicizing their next presentation (this refers to both single events or conferences).

XSL•FO
RenderX

> *This is one of the most exciting pilots I've seen up close. How we can link patient data, what the analysis tells us and how we can provide evidence to make change for patient benefit. #datasaveslives* [SarahM_Research]

Resonating Motto

Underpinning the campaign's motto is the argument that linking data and big data research saves and improves people's lives. This was a resonating motto, with many tweets about how the usage of medical data made an impact and provided new solutions. Tweets were either generic (relating to the benefits of big data research in general) or referred to specific research projects (both completed and ongoing).

> *'Data makes the unknown known' @margaretgrayson @useMYdata @NHSConfed #NHSReset #datasaveslives* [ConyersRebecca]

Space for Improvement

Tweets also argued for some changes to ensure the maximum benefit of big data. There were calls for more investment in big data research infrastructure, showing that big data research is still developing.

> *Predictive data modelling could lead to better humanitarian outcomes, but we are missing half the data needed. Time to act! #DataSavesLives.* [Enovacom_en]

Twitter users also recognized that some of these changes had to happen soon to offer more benefits from research.

> *Ahead of a crucial @G7, @NMRPerrin argues for the urgent need for better coordination across the global data sharing landscape https://t.co/aw8Apgw5Ku #datasaveslives @GS_Humphreys @royalsociety @GloPID_R* [ICODA_research]

Gaining an Understanding of Big Data Research

This theme is about reaching others (including the public but primarily other professionals, policymakers and researchers) and offering an opportunity to learn more about individual projects.

The hashtag offered an opportunity to call people to action, to apply for job openings (mostly research related), and to welcome new team members. Some tweets asked other researchers to support big data research or answer ongoing consultations or surveys.

> *Only a few days left to apply for this! Working with a great team enabling the #HealthData infrastructure to support #COVID19 #research. Secondments welcome, remote working too so location flexible. #HealthData #DataSavesLives* [LaraEdw001]

This illustrates how the hashtag was used among similar-minded people to publicize new opportunities and events.

Tweets also allow readers to learn more about big data projects, attend events, follow online chats, and read recent blogs or papers. This is mostly passive and focused on dissemination rather than engaging.

> *Check out this thread from @HDR_UK with examples of how #DataSavesLives being added throughout June* [NIHRresearch]

Dialogue Model

Ethical Sharing

The need for ethical, safe, and lawful sharing of data in big data research and the importance of doing it right were a prominent theme in the data. Two subthemes deal with key aspects of achieving these aims: protecting individuals' rights, and trust and transparency.

Protecting Individuals' Rights

There is agreement that big data research offers new opportunities for innovation. However, the impact on individual rights remains the main concern. This was particularly around how the data are used, who has access, how secure it is, and whether patients could be identified. Many organizations attempt to reassure people by telling them that any data usage is secure and transparent.

There were concerns that health data could be sold to private companies to make a profit. Some tweets linked that concern with people's decisions to opt out in the United Kingdom from using their medical data for research. Some admitted that the public has not been properly or sufficiently engaged around and about these issues.

> *@Axelheitmueller, you're completely correct, the benefits of data sharing are immense for the health of our nation. For some reason there's a narrative that we intend to make a profit from data. This is simply not the case. We do not, and we will not sell data! #datasaveslives* [simonrbolton]

Trust and Transparency

Associated with individual rights are trust and transparency, which underpin public support for big data research. Tweet authors argued that public trust is essential for big data research to succeed and that the processes of data sharing have to be transparent and follow well-established principles. Otherwise, it risks undermining public support as the public will lose confidence. There have been comments within the UK context that recent political events have undermined that trust, which is also shown by the hashtag #DataGrab. Trust and transparency are perceived as the building blocks of successful research projects and are often the rationale that underpins public engagement.

> *Sharing my data can aid research needed to improve health care for myself and others with chronic illness. However, there does need to be clearer reassurance that data won't be misused so that individuals can make an informed choice. #GPDPR #nhsdataoptout #DataGrab #datasaveslives* [LucindaH19]

XSL•FO
RenderX

Engaging the Public

There was a push in the tweets to have better engagement with the public and encourage conversations about big data research. Some approaches to this included avoiding jargon and ensuring that events are free to attend. There was also some media engagement as Twitter users shared links where researchers took part in media interviews. In addition, media outlets were tagged as Twitter users tried to catch their attention. These engagement activities are intended to help the public understand the value of big data research better. However, if they limited themselves to only explaining big data research to the public, they could be seen as following a deficit model of engagement, with its associated limitations.

> *Health data research can be confusing sometimes and full of buzzwords and jargon. This article clearly explains how health data is used and why it's so important. If you donate your data to health research you could help improve future health care. #DataSavesLives #DataScience* [genscot]

Participatory Model

Coproducing Knowledge With the Public

Public contributors could be successfully involved in big data research. These are public members who actively contribute to research projects, ensuring that research is conducted *with* and not *to* or *about* them. Views on how much the public should be involved differed. Some tweets explore the active role of the public in studies as public contributors, whereas others focus only on reaching people and showing them the benefits of big data research (as shown in the previous theme, proactively engaging the public).

Tweets refer to involving public members in big data projects. In this theme, there is a call for more public involvement. Tweet authors showed examples of how involving the public as active contributors had a positive impact on their research.

There were calls for more public control, thanking patients for sharing their medical data for research (not opting out), and recruitment calls for new public contributors in big data projects.

> *None of this would be possible without our Data Trust Committee – the diverse and inclusive group of patients and members of the public, who review every data access request and make decisions based on the Five Safes and, ultimately, the public's best interest. #datasaveslives* [useMYdata]

> *It is also important to involve patients in developing registries or data collections. Also citizens, because they produce the data and therefore, as owner of the data, they should have a seat on the "Datatable" too. #patientsinvolved #datasaveslives #MTF2021* [Birgitpower]

Discussion

Principal Findings

This study explored how #DataSavesLives was used on Twitter. The findings clearly show that the debate was mostly positive toward the campaign. This is not surprising as most participants were organizations, academics, and institutions that work in big data research. Our findings confirm previous research on the #DataSavesLives hashtag—that it is being used to identify similar-minded projects around big data and to spread positive messages toward big data research, particularly when there are public controversies [34].

We mapped the results of our thematic analysis into models of public engagement. This showed that the largest number of themes were within the deficit and dialogue models and only 1 theme was included in the participatory model. Each model has its uses, and a hierarchy is not necessarily the most useful way to understand them [37]. The public engagement campaign can be placed within all of these models [39]. However, if the campaign wants to improve trust with public members, more active exchange with the public is needed. This can be achieved by moving more campaign-related activities into activities that would conform with the dialogue or participatory models. One way of doing this is to engage more Twitter users to participate in active discussion online. Previous research has shown that Twitter can accommodate a vibrant debate around challenging topics [57]. How Twitter users used the hashtag #DataSavesLives is not a new phenomenon in Twitter discussions about science. For example, a study that explored science festivals found that organizations mostly focus on distributing information and only a smaller part of the Twitter activity is actually interactive [58].

The hashtag usage remains limited to similar-minded Twitter users—a Twitter bubble. The results indicate that communication around big data research is driven by the professional community and research remains 1-way because the public rarely uses the hashtag. This confirms previous research showing that government science organizations do not fully use the potential of social media to engage with the public [59]. Within this data set, there was only a limited appearance of negative hashtags, such as #DataGrab (n=5), which was used during the UK debate on the new database scheme. This elicits questions about how successful the campaign is in achieving its goals of engaging with the public. The campaign messages do not target any seldom-heard communities but rather focus on researchers and professionals. Twitter bubbles are not a new phenomenon, and Sunstein [60] describes them as an "echo chamber" that amplifies the already existing beliefs of Twitter. However, despite public members not using the hashtag themselves, it does not exclude the possibility that they are exposed to these messages, as research [61] has shown that researchers with over 1000 followers on Twitter have diverse followers (eg, media representatives and public members). The #DataSavesLives campaign shares many aspects of 1-way communication and remains in the deficit engagement model. However, many engagement campaigns have limited interaction with the public at the beginning but can improve over time [39]. Thus, based on previous research, the campaign has potential to develop.

The campaign was relaunched in Europe in 2019, but there were only 4 Tweets in languages other than English. Our findings indicate that the campaign remains UK centric as the most active Twitter accounts are based in the United Kingdom. The high

XSL•FO
RenderX

activity of the government-run UK organizations poses the question whether the hashtag and campaign could continue on Twitter without their involvement. The use of #DataSavesLives remains limited on Twitter. However, this can be explained by the type of messages published online. Most were positive or neutral toward the campaign, whereas the negative emotions on social media spread faster than the positive emotions [62]. This should not encourage Tweet authors to start appealing to negative emotions but rather recognize the limitations of the positive engagement campaign.

Ethical challenges and issues of trust and transparency around big data research remain a concern for the public [63]. In 2014, NHS England launched a promotional campaign showing how medical records would become part of a larger database. The project called Care.data was controversial, and a previous study explored the #caredata controversy on Twitter [28]. At that time, there was a distinct lack of public engagement or involvement in big data projects. There now seems to be a clear recognition that the public should be proactively engaged and involved in discussions about big data projects. There is an improvement in how professionals and organizations perceive public involvment. According to Tweet authors, the public can be involved at various points. Some suggest only explaining the benefits of big data research, while others call for and present examples of having public contributors involved in research (eg, governance). Limited the public understanding of the use of big data remains 1 of the largest challenges [64], and more engagement could, arguably, improve this situation.

Based on our research findings, PT participated in a Tweet chat hosted by the European Patients' Forum as part of their regular conversations around big data research on Twitter. We hoped that this would allow more online engagement within the dialogue model. The discussion considered the online movement and how social media is spreading the campaign's message [65]. We found it beneficial to present our research, discuss the emerging findings, and engage with Twitter users who had used the hashtag #DataSavesLives. This was an opportunity to talk to the people involved in running the campaign about what they thought the future of the campaign might be. The public member contributing to the discussion pointed out the need for more actively involving the public around big data research. This further confirmed our findings and the need for researchers to shift engagment to dialogue and participatory models.

Limitations

Organizations in the United Kingdom were the main authors of downloaded Tweets. This limits our understanding of how much the results of our study reflect public attitudes toward the campaign and questions whether the public is actually aware of it. Twitter offers limited demographics about its users. Some data, such as location, were unknown (eg, online location appeared as the third-most popular location, used by 7.6% of Twitter users) or included 2 or more countries. In addition, because some demographic data were unavailable, we cannot say whether the usage differs among different age groups or other attributes.

The activity of an automated Twitter account, a bot, can influence Twitter traffic. A bot aims to create tweets and retweets to expand the coverage of their messages. We manually coded the data set and did not notice this kind of activity, but this does not guarantee that it was not there.

Data collection took place when there were new database scheme controversies in the United Kingdom, which could have influenced some traffic and messages. Future research should check whether the Twitter discussion has shifted depending on the context. Our study explored only usage of #DataSavesLives in English, but it is also available in German as #DatenrettenLeben. Our study focused on Twitter, the main microblogging platform, where users often discuss contentious or political topics. However, the hashtag is also available on other social media (Facebook and Instagram), and future research could explore whether engagement there differs from Twitter. Other research could also focus on negative hahstags toward sharing routinely collected health data, such as the already mentioned #DataGrab.

Conclusion

This study shows how Twitter users used #DataSavesLives when the hashtag was adopted by the UK government and during the UK domestic controversies around data linkage and sharing. There are growing expectations from funders that researchers will engage with the public. Social media campaigns, such as #DataSavesLives, may offer an opportunity to further this goal. This study expands our understanding of the #DataSavesLives campaign. The results demonstrate the potential of social media and recognizes the need for engaging with a wider range of opinions and different Twitter constituencies. Thus, researchers need to identify new ways of actively engaging a wider range of the general public. There is a need to move engagement activities from a deficit model to dialogue and participatory models that include active 2-way engagment between researchers and public members and genuinely include the public in meaningful involvement. Future research could explore whether and how Facebook and Instagram users use the hashtag.

Acknowledgments

We would like to thank Estefania Cordero, project communications officer (European Patients' Forum) and Gözde Susuzlu Briggs, project coordinator (European Patients' Forum) for assisting us with the history and background of the #DataSavesLives campaign.

We also would like to show our gratitude to 2 anonymous reviewers who provided constructive feedback on the draft version of this paper.

XSL•FO
RenderX

This report is independent research funded by the National Institute for Health Research Applied Research Collaboration North West Coast (ARC NWC). The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care. SER is partly funded by the National Institute for Health Research (NIHR) Applied Research Collaboration North West Coast.

Authors' Contributions

All authors contributed to the study design. PT drafted the first draft of this paper, and SER, KF, and LF contributed to drafting and editing. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Kurath M, Gisler P. Informing, involving or engaging? Science communication, in the ages of atom-, bio- and nanotechnology. Public Underst Sci 2009 Sep 03;18(5):559-573. [doi: 10.1177/0963662509104723] [Medline: 20027772]
2. Simis MJ, Madden H, Cacciatore MA, Yeo SK. The lure of rationality: why does the deficit model persist in science communication? Public Underst Sci 2016 May 26;25(4):400-414. [doi: 10.1177/0963662516629749] [Medline: 27117768]
3. Priest S, Goodwin J, Dahlstrom MF. Ethics and Practice in Science Communication. Chicago, IL: University of Chicago Press; 2018.
4. INVOLVE. What Is Public Involvement in Research?. 2020. URL: https://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/ [accessed 2022-09-21]
5. National Co-ordinating Centre for Public Engagement. What Is Public Engagement?. 2021. URL: https://www.publicengagement.ac.uk/about-engagement/what-public-engagement [accessed 2022-11-17]
6. Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: experiences of public engagement in the Scottish Health Informatics Programme. Sci Public Policy 2016 Oct;43(5):713-723 [FREE Full text] [doi: 10.1093/scipol/scv075] [Medline: 28066123]
7. Twitter: About. URL: https://about.twitter.com/en [accessed 2022-07-15]
8. López-Goñi I, Sánchez-Angulo M. Social networks as a tool for science communication and public engagement: focus on Twitter. FEMS Microbiol Lett 2018 Feb 01;365(2):fnx246. [doi: 10.1093/femsle/fnx246] [Medline: 29165564]
9. Fuller M, Allen T. Let's have a Tweetup: the case for using Twitter professionally. Arch Pathol Lab Med 2016 Sep 1;140(9):956-957. [doi: 10.5858/arpa.2016-0172-sa]
10. Tripathy JP, Bhatnagar A, Shewade HD, Kumar AMV, Zachariah R, Harries AD. Ten tips to improve the visibility and dissemination of research for policy makers and practitioners. Public Health Action 2017 Mar 21;7(1):10-14 [FREE Full text] [doi: 10.5588/pha.16.0090] [Medline: 28775937]
11. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
12. Hemingway H, Asselbergs F, Danesh J, Dobson R, Maniadakis N, Maggioni A, Innovative Medicines Initiative 2nd Programme, Big Data for Better Outcomes, BigData@Heart Consortium of 20 Academic Industry Partners including ESC. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J 2018 Apr 21;39(16):1481-1495 [FREE Full text] [doi: 10.1093/eurheartj/ehx487] [Medline: 29370377]
13. Jones K, Laurie G, Stevens L, Dobbs C, Ford D, Lea N. The other side of the coin: harm due to the non-use of health-related data. Int J Med Inform 2017 Jan;97:43-51 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.09.010] [Medline: 27919394]
14. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
15. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2022 Jan 12;48(1):3-13 [FREE Full text] [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
16. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015 May;41(5):404-409 [FREE Full text] [doi: 10.1136/medethics-2014-102374] [Medline: 25617016]
17. Data Saves Lives. What Is Data Saves Lives?. 2020. URL: https://datasaveslives.eu/aboutdsl [accessed 2022-01-11]
18. European Patients Forum. 'DATA SAVES LIVES' UNVEILED AT EPF CONGRESS. 2019. URL: https://www.eu-patient.eu/news/latest-epf-news/2019/data-saves-lives-unveiled-at-epf-congress/ [accessed 2022-12-02]
19. Bradshaw AS. #DoctorsSpeakUp: exploration of hashtag hijacking by anti-vaccine advocates and the influence of scientific counterpublics on Twitter. Health Commun 2022 Apr 18:1-11. [doi: 10.1080/10410236.2022.2058159] [Medline: 35437069]
20. Jackson SJ, Foucault Welles B. Hijacking #myNYPD: social media dissent and networked counterpublics. J Commun 2015 Nov 05;65(6):932-952. [doi: 10.1111/jcom.12185]

XSL•FO
RenderX

21. Scott P, Emerson K, Henderson-Reay T. Data saves lives. BMJ 2021 Jul 08;374:n1694. [doi: 10.1136/bmj.n1694] [Medline: 34244207]
22. Data Saves Lives: Reshaping Health and Social Care with Data (draft), Updated 10 February 2022. 2020. URL: https://tinyurl.com/yeyrt4vk [accessed 2022-08-01]
23. O'Dowd A. New GP patient data extraction scheme raises concern. BMJ 2021 May 28;373:n1389. [doi: 10.1136/bmj.n1389] [Medline: 34049862]
24. Shaw D, Erren T. If data save lives, consent saves data and data donation. BMJ 2021 Aug 20;374:n2069. [doi: 10.1136/bmj.n2069] [Medline: 34417169]
25. Armstrong S. GP data scheme may face further delay as practices refuse to hand over patient information. BMJ 2021 Jul 14;374:n1777. [doi: 10.1136/bmj.n1777] [Medline: 34261673]
26. Shaun L. Misinformation on NHS data sharing could scupper vital health research, warn experts. The Independent. 2021. URL: https://www.independent.co.uk/news/health/nhs-data-sharing-research-health-b1862658.html [accessed 2022-08-11]
27. Craig E. Matt Hancock says 'data saves lives' and dismisses idea of sending letters to patients to tell them how they can opt out of controversial NHS data-sharing plan because 'hardly anyone writes to you these days'. The Daily Mail. 2021. URL: https://www.dailymail.co.uk/news/article-9712079/Matt-Hancock-dismisses-idea-writing-patients-tell-opt-NHS-data-sharing.html [accessed 2022-08-11]
28. Hays R, Daker-White G. The care.data consensus? A qualitative analysis of opinions expressed on Twitter. BMC Public Health 2015 Sep 02;15:838 [FREE Full text] [doi: 10.1186/s12889-015-2180-9] [Medline: 26329489]
29. Royal College of General Practitioners, British Medical Association. Letter to NHS Digital from RCGP and BMA re. General Practice Data for Planning and Research - Communications with the Public. 2021. URL: https://web.archive.org/web/20210808123959/https://www.rcgp.org.uk/-/media/Files/News/rcgp-bma-letter-gpdpr.ashx?la=en [accessed 2022-11-10]
30. Wise J. New data strategy for England aims to rebuild public trust. BMJ 2022 Jun 14;377:o1455. [doi: 10.1136/bmj.o1455] [Medline: 35700998]
31. NHS Digital. General Practice Data for Planning and Research (GPDPR). 2021. URL: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research#top [accessed 2021-11-17]
32. Goldacre B, Morley J. Better, Broader, Safer: Using Health Data for Research and Analysis. 2022. URL: https://tinyurl.com/bddw2msd [accessed 2022-07-27]
33. Data Saves Lives: Reshaping Health and Social Care with Data, Updated 15 June 2022. 2022. URL: https://tinyurl.com/2me5etrf [accessed 2022-08-02]
34. Hassan L, Nenadic G, Tully MP. A social media campaign (#datasaveslives) to promote the benefits of using health data for research purposes: mixed methods analysis. J Med Internet Res 2021 Feb 16;23(2):e16348 [FREE Full text] [doi: 10.2196/16348] [Medline: 33591280]
35. Burns TW, O'Connor DJ, Stocklmayer SM. Science communication: a contemporary definition. Public Underst Sci 2016 Aug 19;12(2):183-202. [doi: 10.1177/09636625030122004]
36. Metcalfe J. Comparing science communication theory with practice: an assessment and critique using Australian data. Public Underst Sci 2019 May 12;28(4):382-400. [doi: 10.1177/0963662518821022] [Medline: 30755086]
37. Trench B. Towards an analytical framework of science communication models. In: Communicating Science in Social Contexts. Netherlands: Springer; 2008:119-135.
38. Seethaler S, Evans JH, Gere C, Rajagopalan RM. Science, values, and science communication: competencies for pushing beyond the deficit model. Sci Commun 2019 May 13;41(3):378-388. [doi: 10.1177/1075547019847484]
39. Bucchi M. Of deficits, deviations and dialogues: theories of public communication of science. In: Handbook of Public Communication of Science and Technology. London, UK: Routledge; 2008:71-90.
40. Wynne B. Public engagement as a means of restoring public trust in science: hitting the notes, but missing the music? Community Genet 2006 Jun 2;9(3):211-220. [doi: 10.1159/000092659] [Medline: 16741352]
41. Miller S. Public understanding of science at the crossroads. Public Underst Sci 2001 Jan 1;10(1):115-120.
42. Lee R, Garvin T. Moving from information transfer to information exchange in health and health care. Soc Sci Med 2003 Feb;56(3):449-464 [FREE Full text] [doi: 10.1016/s0277-9536(02)00045-x]
43. Carney L, Jones L, Braddon F, Pullyblank A, Dixon A. A colorectal cancer patient focus group develops an information package. Ann R Coll Surg Engl 2006 Sep;88(5):447-449. [doi: 10.1308/003588406x114686]
44. National Institute for Health and Care Research (NIHR). Briefing Notes for Researchers - Public Involvement in NHS, Health and Social Care Research. URL: https://www.nihr.ac.uk/documents/briefing-notes-for-researchers-public-involvement-in-nhs-health-and-social-care-research/27371 [accessed 2022-07-14]
45. Mockford C, Staniszewska S, Griffiths F, Herron-Marx S. The impact of patient and public involvement on UK NHS health care: a systematic review. Int J Qual Health Care 2012 Feb;24(1):28-38 [FREE Full text] [doi: 10.1093/intqhc/mzr066] [Medline: 22109631]
46. 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]

XSL•FO
RenderX

47. Bogen KW, Mulla MMM, Haikalis M, Orchowski LM. Sexual victimization among men: a qualitative analysis of the Twitter hashtag #UsToo. J Interpers Violence 2022 May 04;37(9-10):NP7825-NP7849. [doi: 10.1177/0886260520967167] [Medline: 33146060]
48. Rashid MA, McKechnie D, Gill D. What advice is given to newly qualified doctors on Twitter? An analysis of #TipsForNewDocs tweets. Med Educ 2018 Jul 15;52(7):747-756. [doi: 10.1111/medu.13589] [Medline: 29766550]
49. Thelwall M. Webometrics and Social Web Research Methods. Wolverhampton, England: University of Wolverhampton; 2013:140.
50. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
51. Braun V, Clarke V. Thematic Analysis: A Practical Guide. London, UK: Sage; 2021.
52. Talbot CV, O'Dwyer ST, Clare L, Heaton J, Anderson J. How people with dementia use twitter: a qualitative analysis. Comput Hum Behav 2020 Jan;102:112-119. [doi: 10.1016/j.chb.2019.08.005]
53. Williams ML, Burnap P, Sloan L. Towards an ethical framework for publishing Twitter data in social research: taking into account users' views, online context and algorithmic estimation. Sociology 2017 Dec 26;51(6):1149-1168 [FREE Full text] [doi: 10.1177/0038038517708140] [Medline: 29276313]
54. Health Data Research UK. Call to Action: #DataSavesLives - We All Need to Be Represented. 2021. URL: https://www.hdruk.ac.uk/news/data-saves-lives-we-all-need-to-be-represented-call-to-action/#utm_source=hive&utm_medium=email [accessed 2021-11-15]
55. Teodorowski P. Spór migracyjny wśród użytkowników Tweetera w #BrexitDay. KSzPP 2020 Nov 18;4(28):83-104 [FREE Full text] [doi: 10.33119/kszpp/2020.4.5]
56. Karmegam D, Mapillairaju B. What people share about the COVID-19 outbreak on Twitter? An exploratory analysis. BMJ Health Care Inform 2020 Nov 19;27(3):e100133 [FREE Full text] [doi: 10.1136/bmjhci-2020-100133] [Medline: 33214193]
57. Richardson J, Grose J, Nelmes P, Parra G, Linares M. Tweet if you want to be sustainable: a thematic analysis of a Twitter chat to discuss sustainability in nurse education. J Adv Nurs 2016 May 29;72(5):1086-1096. [doi: 10.1111/jan.12900] [Medline: 26821875]
58. Su LY, Scheufele DA, Bell L, Brossard D, Xenos MA. Information-sharing and community-building: exploring the use of Twitter in science public relations. Sci Commun 2017 Oct 04;39(5):569-597. [doi: 10.1177/1075547017734226]
59. Lee NM, VanDyke MS. Set it and forget it: the one-way use of social media by government agencies communicating science. Sci Commun 2015 May 27;37(4):533-541. [doi: 10.1177/1075547015588600]
60. Sunstein C. # Republic. Princeton, NJ: Princeton University Press; 2018.
61. Côté IM, Darling ES. Scientists on Twitter: preaching to the choir or singing from the rooftops? FACETS 2018 Jan;3(1):682-694. [doi: 10.1139/facets-2018-0002]
62. Zhu X, Kim Y, Park H. Do messages spread widely also diffuse fast? Examining the effects of message characteristics on information diffusion. Comput Hum Behav 2020 Feb;103:37-47 [FREE Full text] [doi: 10.1016/j.chb.2019.09.006]
63. Stockdale J, Cassell J, Ford E. "Giving something back": a systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. Wellcome Open Res 2018 Jan 17;3:6 [FREE Full text] [doi: 10.12688/wellcomeopenres.13531.2] [Medline: 30854470]
64. Skovgaard LL, Wadmann S, Hoeyer K. A review of attitudes towards the reuse of health data among people in the European Union: the primacy of purpose and the common good. Health Policy 2019 Jun;123(6):564-571 [FREE Full text] [doi: 10.1016/j.healthpol.2019.03.012] [Medline: 30961905]
65. Reflecting on the Online Movement – How Social Media Is Spreading the Message of Data Saves Lives. 2021. URL: https://twitter.com/i/events/1464220352674996242 [accessed 2021-11-26]

Abbreviations

NHS: National Health Service.

Edited by R Kukafka; submitted 24.03.22; peer-reviewed by D Carvalho, L Hassan; comments to author 22.06.22; revised version received 16.08.22; accepted 16.09.22; published 15.11.22.

Please cite as:
Teodorowski P, Rodgers SE, Fleming K, Frith L
Use of the Hashtag #DataSavesLives on Twitter: Exploratory and Thematic Analysis
J Med Internet Res 2022;24(11):e38232
URL: https://www.jmir.org/2022/11/e38232
doi:10.2196/38232
PMID:36378518

XSL•FO
RenderX

©Piotr Teodorowski, Sarah E Rodgers, Kate Fleming, Lucy Frith. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Themes in TikTok Videos Featuring Little Cigars and Cigarillos: Content Analysis

Julia Vassey[1], MPH; Scott I Donaldson[1], PhD; Allison Dormanesh[1], PhD; Jon-Patrick Allem[1], PhD

Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Jon-Patrick Allem, PhD
Department of Population and Public Health Sciences
Keck School of Medicine
University of Southern California
3rd Floor, SSB 312D
1845 N Soto St
Los Angeles, CA, 90032
United States
Phone: 1 323 442 7921
Email: allem@usc.edu

Abstract

Background: Little cigars and cigarillos (LCCs) are popular tobacco products among youth (ie, adolescents and young adults). A variety of LCC-related promotional and user-generated content is present on social media. However, research on LCC-related posts on social media has been largely focused on platforms that are primarily text- or image-based, such as Twitter and Instagram.

Objective: This study analyzed LCC-related content on TikTok, an audio and video–based platform popular among youth.

Methods: Publicly available posts (N=811) that contained the LCC-related hashtags #swishersweets or #backwoods were collected on TikTok from January 2019 to May 2021. Metadata were also collected, including numbers of likes, comments, shares, and views per video. Using an inductive approach, a codebook consisting of 26 themes was developed to help summarize the underlying themes evident in the TikTok videos and corresponding captions. A pairwise co-occurrence analysis of themes was also conducted to evaluate connections among themes.

Results: Among the 811 posts, the LCC presence theme (ie, a visible LCC) occurred in the most prominent number of posts (n=661, 81.5%), followed by music (n=559, 68.9%), youth (n=332, 40.9%), humor (n=263, 32.4%), LCC use (n=242, 29.8%), flavors (n=232, 28.6%), branding (n=182, 22.4%), paraphernalia (n=137, 16.9%), blunt rolling (n=94, 11.6%), and price (n=84, 10.4%). Product reviews had the highest engagement, with a median 44 (mean 2857, range 36,499) likes and median 491 (mean 15,711, range 193,590) views; followed by product comparisons, with a median 44 (mean 1920, range 36,500) likes and median 671 (mean 11,277, range 193,798) views. Promotions had the lowest engagement, with a median 4 (mean 8, range 34) likes and median 78 (mean 213, range 1131) views. The most prevalent themes co-occurring with LCC presence were (1) music (434/811, 53.5%), (2) youth (264/811, 32.6%), (3) humor (219/811, 27%), (4) flavors (214/811, 26.4%), and (5) LCC use (207/811, 25.5%).

Conclusions: LCC-related marketing and user-generated content was present on TikTok, including videos showing youth discussing, displaying, or using LCCs. Such content may be in violation of TikTok's community guidelines prohibiting the display, promotion, or posting of tobacco-related content on its platform, including the display of possession or consumption of tobacco by a minor. Further improvement in the enforcement of TikTok community guidelines and additional scrutiny from public health policy makers may be necessary for protecting youth from future exposure to tobacco-related posts. Observational and experimental studies are needed to understand the impact of exposure to LCC-related videos on attitudes and behaviors related to LCC use among youth. Finally, there may be a need for engaging antitobacco videos that appeal to youth on TikTok to counter the protobacco content on this platform.

(J Med Internet Res 2022;24(11):e42441) doi:10.2196/42441

KEYWORDS

cigarillo; little cigar; social media; TikTok; video; cigar; cigarette; smoker; smoking; tobacco; social media; content analysis; youth; young adult; adolescent; user generated content

XSL•FO
RenderX

Introduction

Little cigars and cigarillos (LCCs) comprise over 97% of the US cigar market and are popular among youth (ie, adolescents and young adults) [1]. In 2021, LCCs were the third most-used tobacco product among US adolescents, with 5% reporting having used cigars in their lifetime [2]. The popularity of LCCs has been partly attributed to tobacco industry marketing that promotes a wide range of available flavors [3,4], a lack of minimum pack size requirements, price, and fewer sales restrictions than on cigarettes [4,5]. Research has shown that LCCs are associated with an increased risk of nicotine addiction and preventable chronic diseases, including cancer and heart disease [6,7].

A variety of content related to flavored LCCs exists on social media. Despite restrictions on marketing and sales of flavored cigar products in the United States [8] and restrictions on advertising of tobacco products on social media [9,10], platforms popular among adolescents [11,12] host direct ads from tobacco brands or promotional tobacco content from microinfluencers [13-16] (eg, models, bloggers, and brand ambassadors) who typically post on behalf of tobacco brands. Studies have demonstrated that promotional content from influencers may be perceived as more trustworthy than traditional advertising (eg, banner ads), since content created by influencers may be perceived as more authentic [17,18] than direct marketing. Exposure to tobacco-related content from such trusted sources may increase susceptibility to the use of tobacco products among social media users, including youth.

Prior research has demonstrated that LCCs are often discussed on social media in the context of music (eg, rap or hip-hop lyrics), humor, and other forms of entertainment; these are positive experiences that could normalize tobacco use among youth [5,19-22]. For example, research has shown that if adolescents observe friends, acquaintances, or influencers using tobacco products on social media and appearing to be happy and popular, they may perceive tobacco use as a behavior to emulate [23].

Research to date on LCC-related content on social media has mostly focused on text-based platforms like Twitter and image-based platforms like Instagram [5,19-22]. Missing from this literature is a comprehensive investigation of LCC-related content on emerging social media platforms, like TikTok. Only one study, to our knowledge, has analyzed promotional content for large cigars and Swisher Sweets on TikTok posted in 2016 through 2020 [16]. TikTok is one of the most popular video-sharing social-networking platforms among youth [12,24,25]. Posts on TikTok are short (typically less than a minute long) video clips often created and viewed by youth. Among TikTok users, 25% are between the ages of 10 to 19 years [24]. Due to its relatively younger user base, TikTok prohibits not only promoting but also posting any tobacco-related content or displaying the possession or consumption of tobacco by a minor [9]. However, despite these restrictions, the platform does contain promotional and user-generated tobacco-related content, including for LCCs [16,26], that could be accessible to youth. Although TikTok announced that content showing the use of tobacco products would not be eligible for recommendation by the platform's algorithm to its users [27], the effectiveness of this policy remains unknown.

The goal of this study was to collect and characterize recent (2019-2021) posts about LCCs on TikTok. We used a detailed codebook that included 26 themes to conduct this content analysis. We documented the presence of videos featuring adolescents and young adults using LCCs and compared the prevalence of and user engagement with themes normalizing risky behavior (ie, showing LCC use as a fun experience) versus themes criticizing or warning others about the consequences of risky behavior (ie, showing negative health consequences of LCC use). Lastly, we described the presence of LCC marketing themes, including the promotion of flavors. This study provides evidence to keep the public health community abreast of an important tobacco-related issue at the intersection of social media and tobacco control.

Methods

Data Collection

TikTok data were collected by scraping (ie, electronically copying) publicly available posts that contained the hashtags #swishersweets or #backwoods. Similar hashtags have been used in prior research [5,19,21] and were chosen based on their high audience engagement compared to other LCC-related hashtags on TikTok. Prior to data collection, #swishersweets had 1.6 million views and #backwoods had 2.5 million views, while #blackandmild had 345,000 views.

A total of 1570 URLs were initially stored in a csv file. Metadata, which included the numbers of likes, comments, shares, and views per video, were also recorded. The study period ranged from January 2019 through May 2021. A total of 91% (1429/1570) of videos were posted in 2020 or 2021 and 9% (141/1570) in 2019.

The research team worked collaboratively to become familiar with the data and developed a codebook based on an inductive approach. The goal of this approach was to condense the raw audiovisual data into a summary format and report the underlying themes that were evident in the data. The unit of analysis was the post (ie, the TikTok video and the corresponding caption).

To establish interrater reliability, 2 coders analyzed a subsample of Backwoods-related posts (n=200), with percentage agreement ranging from 78% to 100%. Two additional coders analyzed a subsample of Swisher Sweets–related posts (n=149), with percentage agreement ranging from 81% to 99%. The senior author (JPA) served as the arbitrator and resolved disagreements. Four coders then analyzed all videos in the data set (N=1570) to document the presence or absence of each theme. If a theme was present in a post (ie, displayed or mentioned in a video, text overlay, or corresponding caption), it was coded as "1"; otherwise, it was coded as "0." Each post could contain more than one LCC-related theme.

XSL•FO
RenderX

There were 358 inaccessible posts at the time of content analysis. To understand why such posts were inaccessible, the coders documented the apparent reason a post was no longer publicly available. These posts could have been taken down by TikTok's content moderators, as indicated by the message "Video currently unavailable." If the video was directly removed by TikTok, it was an indication that the platform was enforcing its community guidelines prohibiting tobacco-related content [28]. Posts could also have been made unavailable by the content creator due to changes in privacy settings, as indicated by the message "The account is private." These posts were not included in the analytic sample. Additionally, posts coded as non–tobacco-related (n=401), such as videos with the hashtag #backwoods that showed wood paint or clothing dye, were also excluded as being irrelevant to the study, resulting in a final analytic sample of 811 posts. In this sample, the research team examined the proportion of each LCC-related theme and engagement with it, defined as the average (ie, arithmetic mean) and median numbers of likes, shares, comments, and views. Since the numbers of shares and comments were low for all categories, only the numbers of likes and views are reported.

To evaluate connections across themes (ie, the frequency with which each pair of themes was present in a single post), with the goal of uncovering the contextual complexities of a post, pairwise co-occurrence analyses were used. To visualize a co-occurrence matrix containing the counts of co-occurrences of themes, the igraph package in R (version 1.2.6; R Software Foundation) was used. Themes related to positive or negative health effects and addiction, restrictions on LCC use by adolescents, and the "other" theme were either absent or only represented by <1% of the videos and were excluded from the co-occurrence analysis.

Ethics Approval

The University of Southern California Institutional Review Board approved all study procedures (UP-21-00135). Usernames of the content creators were not collected for this study.

Results

The final codebook contained 26 themes (described in Table 1) that included entertainment, LCC product presence or use, presence of youth (ie, individuals who appeared younger than 30 years) in the videos, LCC product characteristics (flavors, packaging, tobacco wrap, and price), marketing and sales, user testimony, positive or negative health effects of LCC use, and addiction. Examples for each theme, represented by paraphrased quotes or descriptions of the videos, are provided in Multimedia Appendix 1.

The prevalence of the themes in TikTok videos with the #backwoods and #swishersweets hashtags are shown in Table 2. The most prominent themes in the videos (N=811) were LCC presence (ie, LCC products were shown, as opposed to only discussed; n=661, 81.5%); music (n=559, 68.9%); youth displaying, discussing, or using LCCs (n=332, 40.9%); humor (n=263, 32.4%); LCC use (n=242, 29.8%); flavors (n=232, 28.6%); branding (ie, LCC logos; n=182, 22.4%); paraphernalia (n=137, 16.9%); blunt rolling (n=94, 11.6%); and price (n=84, 10.4%). The least prevalent themes (ie, restrictions on LCC use by adolescents, other, health warnings, and risk-taking) were present in <1% of the videos for each category. Themes related to positive or negative health effects and addiction to LCCs were absent in the analytic sample (since they were present in videos that were removed by TikTok content moderators or had been made private by users at the time of content analysis).

Among themes identified in more than 1% of the videos in the sample, product reviews had the highest engagement, with a median 44 (mean 2857, range 36,499) likes and median 491 (mean 15,711, range 193,590) views; followed by product comparison, with a median 44 (mean 1920, range 36,500) likes and median 671 (mean 11,277, range 193,798) views; negative sentiment, with a median 44 (mean 983, range 36,500) likes and median 429 (mean 7545, range 193,790) views; and positive sentiment, with a median 37 (mean 485, range 4095) likes and median 563 (mean 4128, range 27,898) views. Promotions had the lowest engagement, with a median 4 (mean 8, range 34) likes and median 78 (mean 213, range 1131) views (Table 2).

As shown in Figure 1, the prevalent co-occurring themes with the LCC presence theme were music (434/811, 53.5%), youth (264/811, 32.6%), humor (219/811, 27%), flavors (214/811, 26.4%), and LCC use (207/811, 25.5%). Figure 1 in high resolution is available in Multimedia Appendix 2.

XSL•FO
RenderX

Table 1. LCC-related themes identified in TikTok videos collected with the hashtags #swishersweets and #backwoods. Videos unrelated to tobacco were excluded from the analytic sample.

Theme	Definition
Entertainment	
Humor	Video that has a sarcastic tone, is satirical, or contains pranks, parody, or jokes (this may include memes—graphics or images that encapsulate a concept or catchphrase)
Music	Song including a chorus or instrumentals
Smoke tricks	Performing tricks with smoke
Pop culture	References to rap, hip-hop, or celebrity endorsements (eg, Snoop Dogg or Cardi B)
Product presence or use	
LCC[a] presence	Visible LCC product packaging, product wrappers, or products displayed on a table, with or without product *use*; other tobacco products may be present along with LCCs (videos without the *presence* of actual products, such as those with people discussing, but not showing, LCCs or smoke clouds, were not included in this category)
LCC use	Self-reported or visible LCC use, including puffing, hitting, rolling, and blowing smoke or an individual holding an LCC
Blunt rolling	Blunt making (ie, the hollowing out of an LCC and refilling it with cannabis)
Polysubstance use	Use of an LCC in combination with other substances, such as other tobacco products, alcohol, marijuana, or other illicit drugs
Youth	
Youth	Presence of young people (ie, a person in the video appears younger than 30 years)
Restrictions on LCC use by adolescents	Use by adolescents of LCCs during or after school hours on school grounds, possibly combined with disciplinary action from parents, schoolteachers, or other authority figures for using, or attempting to purchase, tobacco (before reaching the legal minimum age of 21 years)
Product characteristics	
Flavors	Visible flavored LCC products, including flavor names on product packaging
Product review	Unboxing or removing items from their packaging and showing them to viewers
Price	Showing or discussing LCC prices, including their affordability
Marketing and sales	
Branding	Visible LCC logos on merchandise, apparel, and accessories, including, but not limited to, matches, lighters, and ashtrays
Promotions	Discussing free products, discounts, or coupons (posts in this theme may contain URLs or provide a store's physical address or other contact information)
User testimony	
Product sentiment (negative)	Expressions of dissatisfaction with LCCs or disdain for their use (this may include descriptions of the product's quality, burning speed, or dislike of the flavors)
Product sentiment (positive)	Expressions of approval of an LCC (this may include descriptions of the product's quality, burning speed, or enjoyment of the flavors)
Product comparison	Comparisons between more than one LCC product (ie, one product being better or worse than another)
Positive or negative health effects and addiction	
Addiction	Discussions of addiction to nicotine because of LCC use (these posts may demonstrate a person expressing cravings for or a desire for LCCs)
Health warnings	Expressions of concern about the health effects or adverse effects of LCC use, including lung failure, tongue discoloration, and cancer (these videos may contain scenes of a physician or medical provider offering medical advice; they do not include videos showing the nicotine warning label often seen on tobacco packages)
Cessation	Quitting nicotine or giving up LCCs
Risk-taking	Participating in a dangerous behavior to obtain or use an LCC (this may include cliff diving into a lake, sticking an arm into a vacuum, or inducing vomiting or other forms of self-harm)
Other themes	
Paraphernalia	Visible LCC-related materials, such as matches, lighters, and ashtrays, that may or may not include LCC brand logos
Crowds/socializing	Videos showing groups of people (more than 3)
Other	Tobacco-related posts that do not naturally fit into the abovementioned themes

XSL•FO
RenderX

Theme	Definition
Nontobacco	Content unrelated to tobacco
Video currently unavailable	Post removed by TikTok
The account is private	Post made private by a content creator

[a]LCC: little cigars and cigarillos.

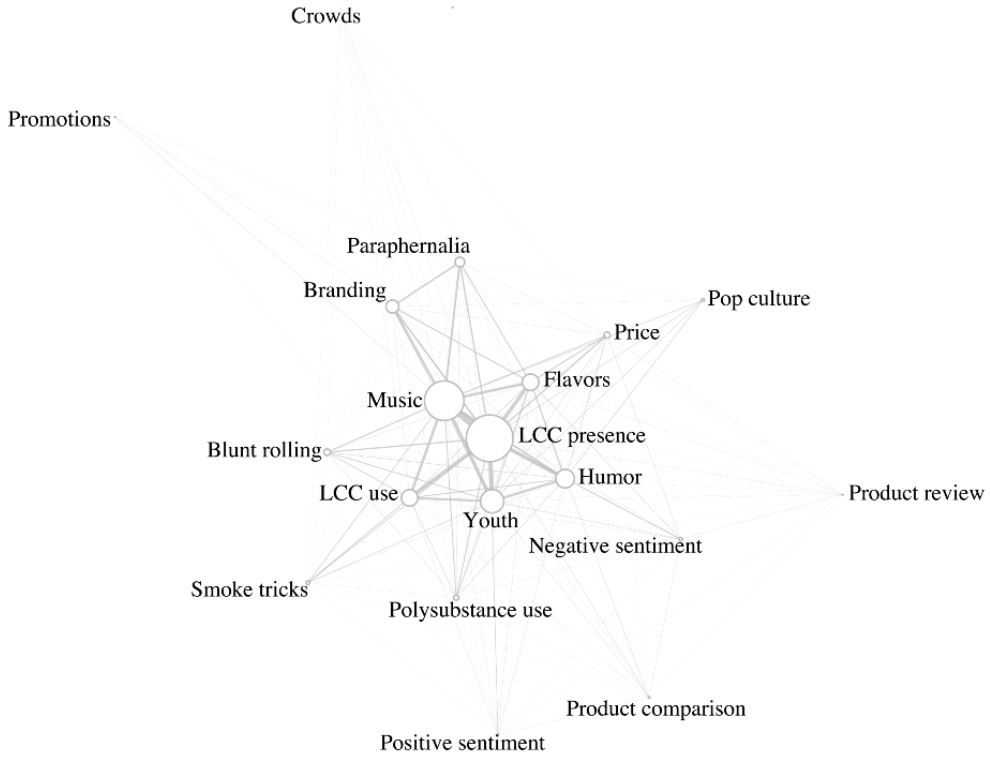
Table 2. Prevalence of themes featured in TikTok videos discussing the #backwoods and #swishersweets hashtags related to little cigars and cigarillos (N=811).

Themes	Videos, n (%)	Likes, median	Likes, mean	Views, median	Views, mean
Entertainment					
Music	559 (68.9)	15	257	228	2478
Humor	263 (32.4)	20	624	322	6732
Smoke tricks	54 (6.7)	12	694	189	3843
Pop culture	41 (5)	20	549	247	2822
Product presence or use					
LCC[a] presence	661 (81.5)	16	217	242	2529
LCC use	242 (29.8)	12	299	204	2365
Blunt rolling	94 (11.6)	16	280	225	4229
Polysubstance use	70 (8.6)	25	267	277	3438
Youth					
Youth	332 (40.9)	21	483	298	4917
Restrictions on LCC use by adolescents	4 (<1)	24	112	301	2474
Product characteristics					
Flavors	232 (28.6)	18	350	225	3708
Product review	13 (1.6)	44	2857	491	15,711
Price	84 (10.4)	23	564	264	3994
Marketing and sales					
Branding	182 (22.4)	15	953	204	6602
Promotions	13 (1.6)	4	8	78	213
Other themes					
Paraphernalia	137 (16.9)	13	1564	206	10,619
Crowds/socializing	10 (1.2)	8	28	229	946
Other	1 (<1)	11	11	386	386
User testimony					
Product sentiment (negative)	49 (6)	44	983	429	7545
Product sentiment (positive)	27 (3.3)	37	485	563	4128
Product comparison	23 (2.8)	44	1920	671	11,277
Positive or negative health effects and addiction					
Addiction	0 (0)	0	0	0	0
Health warnings	1 (<1)	16	16	187	187
Cessation	0 (0)	0	0	0	0
Risk taking	1 (<1)	107	107	7359	7359

[a]LCC: little cigar or cigarillo.

XSL•FO
RenderX

Figure 1. Co-occurrence of themes in little cigar or cigarillo–related videos on TikTok (N=811). The size of the circles represents the frequency with which the theme occurred. The proximity of the circles and the width of the lines represent the frequency of pairwise theme co-occurrences. LCC: little cigar or cigarillo.



Discussion

Principal Findings

We conducted a content analysis of TikTok videos featuring LCCs and found that imagery of LCCs was present in most of the analyzed videos. Many videos featuring LCCs had music and humor. Videos showing LCC use; youth displaying, discussing, or using these products; flavored LCCs; and LCC brand names and logos were also prevalent.

Entertainment-Focused LCC Videos Featuring Youth

Entertainment-related themes (eg, music or humor) have been documented in prior research on LCC-related content on social media [5,16,19-22]. In this study, we created a codebook with nuanced categories to assess the context in which LCCs were discussed, displayed, or used by individuals who appeared to be younger than 30 years. We demonstrated that young-looking individuals in TikTok videos often displayed, discussed, or used LCCs while singing, dancing, or making jokes, potentially normalizing LCC use on social media [23]. At the same time, videos focused on health-protective behaviors, such as health warnings or cessation messages, were rarely observed. Thus, TikTok appears to be a platform that commonly shows risky health behaviors in an entertainment-focused context, exposing youth to harmful tobacco-related content [3,4].

Flavored LCC Products

LCC flavors appeared quite frequently in the videos analyzed in this study. In April 2022, the US Food and Drug Administration proposed a product standard to prohibit all "characterizing flavors" (ie, flavors other than tobacco) in cigars to protect public health, decrease tobacco-related health disparities, and advance health equity [8]. However, flavored cigars continue to command a substantial share of the cigar market, and this study demonstrates that they are commonly depicted and advertised on social media, further exposing youth to harmful products.

Marketing and Sales of LCC-Related Content

Among themes related to marketing and sales, only branding with LCC logos was frequently observed in this study. Videos showing promotional LCC content, including product offers with discounts, coupons, or free samples, were less commonly observed, (unlike, for example, LCC-related content on Instagram [5], where promotional posts are frequent). Promotional videos also had low user engagement, with median and mean numbers of likes and views that were among the lowest in the data set. This finding can be explained by prior research on tobacco that demonstrated that sponsored content is perceived by social media users as less authentic, and consequently less engaging, than organic user-generated content [17].

TikTok Policy Related to Tobacco Content

TikTok is now one of the most-used platforms by adolescents and young adults, having surpassed Instagram in popularity among adolescents [12,24]. In 2021, 63% of Americans between the ages of 12 and 17 used TikTok on a weekly basis, compared to 57% for Instagram [11]. Despite TikTok's community guidelines prohibiting the display, promotion, or posting of tobacco-related content on its platform, including videos showing the possession or consumption of tobacco by a minor [9], this study demonstrated that videos showing LCCs were

XSL•FO
RenderX

present. The removal of posts that violate the platform's community guidelines is essential for protecting youth from exposure to harmful posts, including tobacco-related posts. This is accomplished through the flagging of suspected violations by content moderators [28]. In this study, 22.8% (358/1570) of the original posts collected during the study period were removed by TikTok, which indicates that its community guidelines are enforced to a degree. Further improvement in the enforcement of community guidelines is necessary.

Influence of LCC-Related Social Media Content on LCC Use Among Youth

Although this study found that young people were featured in LCC-related videos, little is known about the influence of LCC-related content on platforms such as TikTok or Instagram on LCC uptake among adolescents and young adults. This may be especially important for a comprehensive understanding of contributing factors to racial and ethnic disparities in health. For example, non-Hispanic Black populations have a higher rate of LCC use than non-Hispanic White populations [1-3,29,30]. While non-Hispanic Black populations are known to be disproportionately targeted by marketing material from LCC manufacturers [3,4], little research exists to demonstrate marketing strategies for tobacco products that target different racial or ethnic groups in the United States on social media. Such research could include, for example, a survey focused on assessing the effect of exposure to promotional LCC-related TikTok content on LCC use among adolescents with geographically, socioeconomically, and racially diverse backgrounds. This could contribute to a better understanding of the impact of social media marketing for LCCs on diverse youth populations in the United States.

Limitations

Videos with only 2 hashtags (#backwoods and #swishersweets) were collected. Given the degree of user engagement with these 2 hashtags, as described in the Methods section, and the fact that Backwoods and Swisher Sweets represent 2 of the most popular LCC brands in the United States [5], we believe the decision to limit our study to these 2 hashtags was justified. However, it is possible that the sample used in this study was not generalizable to other LCC brands linked to hashtags. Additionally, this study focused on TikTok posts, and the findings may not be generalizable to other social media platforms. The posts in this study were collected from a 29-month period (24 months in 2019-2020 and 5 months in 2021) and may not extend to other time periods. In addition, we assumed that user-engagement behavior (eg, liking, viewing, commenting, and sharing posts) was constant throughout the studied period, regardless of any external factors that might have affected this behavior. Data collection relied on the public availability of posts, which prevented the collection of posts from private accounts. Many posts that were initially collected in this study were made unavailable by users or removed by TikTok administrators. This loss of data may have biased the current findings. The findings may not be generalizable to all TikTok users, since we only collected English-language videos. The coders' assessment of the perceived age of the individuals who appeared to be younger than 30 in the videos did not allow for distinguishing between adolescents and young adults. In addition, this subjective assessment could have been biased. However, the 30-year age threshold was conservative enough to give a fairly accurate estimation of the prevalence of young-looking people featured in the videos [31]. Finally, while many videos we reviewed appeared to be user-generated, they might in fact have been sponsored and lacked the required sponsorship disclosures [13,32]. An analysis of the content creators was beyond the scope of this study.

Conclusions

LCC-related content, including videos featuring young individuals discussing, displaying, or using LCCs, might remain on TikTok in violation of the platforms' own community guidelines. Further improvement in the enforcement of TikTok community guidelines and additional scrutiny from public health policy makers is necessary for protecting youth from exposure to tobacco-related posts. Observational and experimental studies are needed to understand the impact of exposure to LCC-related videos on attitudes and behaviors related to LCC use among youth. Future research is also needed to assess LCC marketing strategies targeting different racial and ethnic groups on this platform. Finally, engaging, user-generated antitobacco content that appeals to youth and could be perceived by TikTok users as authentic may be needed to counter the protobacco content on TikTok. Such user-generated antitobacco content may complement the efforts of public health agencies and nonprofit organizations that create educational health campaigns to inform youth about the risks of using tobacco products [33].

Acknowledgments

The authors acknowledge the contributions of the undergraduate students Sarah Donahue, Heather Lopez, Oumayma Benyoussef, and Kyara Yapari in the Department of Population and Public Health Sciences, University of Southern California, for their assistance with coding the TikTok data used in this study.

This study was funded by the California Tobacco Control Branch, Center for Healthy Communities of the California Department of Public Health (contract 21-10032), as part of the Tobacco Industry Monitoring Evaluation project. The findings and conclusions in this article are those of the authors and do not necessarily represent the views or opinions of the California Department of Public Health or the California Health and Human Services Agency.

XSL•FO
RenderX

Data Availability

Data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation and in compliance with the internal review board protocol, to any qualified researcher.

Authors' Contributions

JPA had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. JPA contributed to the concept and design of the study. JV, JPA, AD, and SID contributed to acquisition, analysis, and interpretation of the data. JV contributed to the drafting of the manuscript. JV, JPA, AD, and SID contributed to critical revision and final approval of the manuscript. JV contributed to the statistical analysis. JPA obtained funding for the study.

Conflicts of Interest

JPA has received fees for consulting services in court cases pertaining to content on social media platforms. The authors report no other conflicts of interest.

Multimedia Appendix 1
Supplementary Table 1. Examples of themes featured in LCC-related TikTok videos (N=811).
[DOCX File , 16 KB - jmir_v24i11e42441_app1.docx]

Multimedia Appendix 2
High resolution Figure 1.
[PDF File (Adobe PDF File), 29 KB - jmir_v24i11e42441_app2.pdf]

References

1. Cornacchione Ross J, Reboussin DM, Noar SM, Wiseman KD, Sutfin EL. What do adolescents and young adults think a cigarillo is? implications for health communication. Int J Environ Res Public Health 2021 Mar 23;18(6):3311 [FREE Full text] [doi: 10.3390/ijerph18063311] [Medline: 33806872]
2. Gentzke AS, Wang TW, Cornelius M, Park-Lee E, Ren C, Sawdey MD, et al. Tobacco product use and associated factors among middle and high school students - National Youth Tobacco Survey, United States, 2021. MMWR Surveill Summ 2022 Mar 11;71(5):1-29 [FREE Full text] [doi: 10.15585/mmwr.ss7105a1] [Medline: 35271557]
3. Kong AY, Queen TL, Golden SD, Ribisl KM. Neighborhood disparities in the availability, advertising, promotion, and youth appeal of little cigars and cigarillos, United States, 2015. Nicotine Tob Res 2020 Dec 12;22(12):2170-2177 [FREE Full text] [doi: 10.1093/ntr/ntaa005] [Medline: 31917833]
4. Meernik C, Ranney LM, Lazard AJ, Kim K, Queen TL, Avishai A, et al. The effect of cigarillo packaging elements on young adult perceptions of product flavor, taste, smell, and appeal. PLoS One 2018 Apr 19;13(4):e0196236 [FREE Full text] [doi: 10.1371/journal.pone.0196236] [Medline: 29672604]
5. Allem JP, Escobedo P, Chu K, Boley Cruz T, Unger JB. Images of little cigars and cigarillos on instagram identified by the hashtag #swisher: thematic analysis. J Med Internet Res 2017 Jul 14;19(7):e255 [FREE Full text] [doi: 10.2196/jmir.7634] [Medline: 28710057]
6. Deeming tobacco products to be subject to the Federal Food, Drug, and Cosmetic Act, as amended by the Family Smoking Prevention and Tobacco Control Act; restrictions on the sale and distribution of tobacco products and required warning statements for tobacco products. Federal Register. 2016. URL: https://www.federalregister.gov/documents/2016/05/10/2016-10685/deeming-tobacco-products-to-be-subject-to-the-federal-food-drug-and-cosmetic-act-as-amended-by-the [accessed 2022-05-23]
7. Sterling KL, Fryer CS, Pagano I, Fagan P. Little cigars and cigarillos use among young adult cigarette smokers in the United States: understanding risk of concomitant use subtypes. Nicotine Tob Res 2016 Dec 06;18(12):2234-2242 [FREE Full text] [doi: 10.1093/ntr/ntw170] [Medline: 27613889]
8. Tobacco product standard for characterizing flavors in cigars. Federal Register. URL: https://www.federalregister.gov/documents/2022/05/04/2022-08993/tobacco-product-standard-for-characterizing-flavors-in-cigars [accessed 2022-05-24]
9. Community Guidelines. TikTok. URL: https://www.tiktok.com/community-guidelines?lang=en [accessed 2021-04-04]
10. Branded Content Policies. Facebook Business Help Center. URL: https://www.facebook.com/business/help/221149188908254 [accessed 2020-05-20]
11. Rodriguez S. TikTok usage surpassed Instagram this year among kids aged 12 to 17, Forrester survey says. CNBC. URL: https://www.cnbc.com/2021/11/18/tiktok-usage-topped-instagram-in-2021-among-kids-12-to-17-forrester-.html [accessed 2022-05-22]
12. Social networks used by teens in U.S. Statista. URL: https://www.statista.com/statistics/250172/social-network-usage-of-us-teens-and-young-adults/ [accessed 2022-06-07]

XSL•FO
RenderX

13. Vassey J, Valente T, Barker J, Stanton C, Li D, Laestadius L, et al. E-cigarette brands and social media influencers on Instagram: a social network analysis. Tob Control 2022 Feb 07:1-8 [FREE Full text] [doi: 10.1136/tobaccocontrol-2021-057053] [Medline: 35131947]
14. Vassey J, Allem JP, Barker J, Cruz TB, Pang R, Unger JB, et al. E-cigarette use and promotion by social media influencers during videogame play on Twitch. Tob Control 2021 Oct 08. [doi: 10.1136/tobaccocontrol-2021-056828] [Medline: 34625511]
15. Definition: What is an Instagram Influencer? Pixlee. URL: https://www.pixlee.com/definitions/definition-instagram-influencer [accessed 2020-01-13]
16. Wu J, Harlow AF, Wijaya D, Berman M, Benjamin EJ, Xuan Z, et al. The impact of influencers on cigar promotions: a content analysis of large cigar and Swisher Sweets videos on TikTok. Int J Environ Res Public Health 2022 Jun 09;19(12):7064 [FREE Full text] [doi: 10.3390/ijerph19127064] [Medline: 35742315]
17. Vogel EA, Guillory J, Ling PM. Sponsorship disclosures and perceptions of e-cigarette Instagram posts. Tob Regul Sci 2020 Sep 01;6(5):355-368 [FREE Full text] [doi: 10.18001/trs.6.5.5] [Medline: 33778107]
18. Childers CC, Lemon LL, Hoy MG. #Sponsored #Ad: agency perspective on influencer marketing campaigns. J Curr Issues Res Advert 2018 Nov 05;40(3):258-274. [doi: 10.1080/10641734.2018.1521113]
19. Kostygina G, Tran H, Shi Y, Kim Y, Emery S. 'Sweeter Than a Swisher': amount and themes of little cigar and cigarillo content on Twitter. Tob Control 2016 Oct;25(Suppl 1):i75-i82 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053094] [Medline: 27697951]
20. Allem J, Uppu SP, Boley Cruz T, Unger JB. Characterizing Swisher little cigar-related posts on twitter in 2018: text analysis. J Med Internet Res 2019 Jul 19;21(7):e14398 [FREE Full text] [doi: 10.2196/14398] [Medline: 31325291]
21. Smiley SL, Kim S, Mourali A, Allem J, Unger JB, Boley Cruz T. Characterizing #Backwoods on Instagram: "The Number One Selling All Natural Cigar". Int J Environ Res Public Health 2020 Jun 25;17(12):NA [FREE Full text] [doi: 10.3390/ijerph17124584] [Medline: 32630567]
22. Kim S, Mourali A, Allem J, Unger JB, Boley Cruz T, Smiley SL. Instagram posts related to Backwoods cigarillo blunts: content analysis. JMIR Public Health Surveill 2021 Feb 09;7(2):e22946 [FREE Full text] [doi: 10.2196/22946] [Medline: 33560242]
23. Vogel E, Ramo D, Rubinstein M, Delucchi K, Darrow S, Costello C, et al. Effects of social media on adolescents' willingness and intention to use e-cigarettes: an experimental investigation. Nicotine Tob Res 2021 Mar 19;23(4):694-701 [FREE Full text] [doi: 10.1093/ntr/ntaa003] [Medline: 31912147]
24. U.S. TikTok users by age 2021. Statista. URL: https://www.statista.com/statistics/1095186/tiktok-us-users-age/ [accessed 2022-06-07]
25. Is photo or video better for social media engagement? Adobe. URL: https://blog.adobe.com/en/publish/2021/10/07/is-photo-or-video-better-for-social-media-engagement [accessed 2022-10-23]
26. Sun T, Lim CCW, Chung J, Cheng B, Davidson L, Tisdale C, et al. Vaping on TikTok: a systematic thematic analysis. Tob Control 2021 Jul 26:NA. [doi: 10.1136/tobaccocontrol-2021-056619] [Medline: 34312317]
27. What is the 'For You' feed? TikTok. URL: https://www.tiktok.com/creators/creator-portal/en-us/how-tiktok-works/whats-the-for-you-page-and-how-do-i-get-there/ [accessed 2022-06-05]
28. Rutherford BN, Sun T, Lim CCW, Chung J, Cheng B, Davidson L, et al. Changes in viewer engagement and accessibility of popular vaping videos on TikTok: a 12-month prospective study. Int J Environ Res Public Health 2022 Jan 20;19(3):NA [FREE Full text] [doi: 10.3390/ijerph19031141] [Medline: 35162170]
29. Soneji S, Sargent J, Tanski S. Multiple tobacco product use among US adolescents and young adults. Tob Control 2016 Mar;25(2):174-180 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051638] [Medline: 25361744]
30. Hinds JT, Li X, Loukas A, Pasch KE, Perry CL. Flavored cigars appeal to younger, female, and racial/ethnic minority college students. Nicotine Tob Res 2018 Feb 07;20(3):347-354 [FREE Full text] [doi: 10.1093/ntr/ntx041] [Medline: 28199700]
31. Alhabash S, Mundel J, Deng T, McAlister A, Quilliam ET, Richards JI, et al. Social media alcohol advertising among underage minors: effects of models' age. Int J Advert 2020 Dec 01;40(4):552-581. [doi: 10.1080/02650487.2020.1852807]
32. FTC Releases Advertising Disclosures Guidance for Online Influencers. Federal Trade Commission. 2019. URL: https://www.ftc.gov/news-events/press-releases/2019/11/ftc-releases-advertising-disclosures-guidance-online-influencers [accessed 2020-01-11]
33. Truthorange. YouTube. URL: https://www.youtube.com/user/truthorange/featured [accessed 2022-05-25]

Abbreviations

LCC: little cigars and cigarillos

XSL•FO
RenderX

Edited by T Leung, A Mavragani; submitted 04.09.22; peer-reviewed by S Beasley, M Navarro, M Kapsetaki, H Veldandi; comments to author 27.09.22; revised version received 26.10.22; accepted 01.11.22; published 16.11.22.

Please cite as:
Vassey J, Donaldson SI, Dormanesh A, Allem JP
Themes in TikTok Videos Featuring Little Cigars and Cigarillos: Content Analysis
J Med Internet Res 2022;24(11):e42441
URL: https://www.jmir.org/2022/11/e42441
doi:10.2196/42441
PMID:36383406

©Julia Vassey, Scott I Donaldson, Allison Dormanesh, Jon-Patrick Allem. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Identifying Sleep Disorders From Search Engine Activity: Combining User-Generated Data With a Clinically Validated Questionnaire

Mairav Cohen Zion[1], PhD; Iddo Gescheit[1], MSc; Nir Levy[1], MA; Elad Yom-Tov[2], PhD

[1]Dayzz, Herzeliya, Israel

[2]Microsoft Research, Herzeliya, Israel

Corresponding Author:
Elad Yom-Tov, PhD
Microsoft Research
3 Alan Turing St
Herzeliya, 4672415
Israel
Phone: 972 779391359
Email: eladyt@microsoft.com

Abstract

Background: Sleep disorders are experienced by up to 40% of the population but their diagnosis is often delayed by the availability of specialists.

Objective: We propose the use of search engine activity in conjunction with a validated web-based sleep questionnaire to facilitate wide-scale screening of prevalent sleep disorders.

Methods: Search advertisements offering a web-based sleep disorder screening questionnaire were shown on the Bing search engine to individuals who indicated an interest in sleep disorders. People who clicked on the advertisements and completed the sleep questionnaire were identified as being at risk for 1 of 4 common sleep disorders. A machine learning algorithm was applied to previous search engine queries to predict their suspected sleep disorder, as identified by the questionnaire.

Results: A total of 397 users consented to participate in the study and completed the questionnaire. Of them, 132 had sufficient past query data for analysis. Our findings show that diurnal patterns of people with sleep disorders were shifted by 2-3 hours compared to those of the controls. Past query activity was predictive of sleep disorders, approaching an area under the receiver operating characteristic curve of 0.62-0.69, depending on the sleep disorder.

Conclusions: Targeted advertisements can be used as an initial screening tool for people with sleep disorders. However, search engine data are seemingly insufficient as a sole method for screening. Nevertheless, we believe that evaluable web-based information, easily collected and processed with little effort on part of the physician and with low burden on the individual, can assist in the diagnostic process and possibly drive people to seek sleep assessment and diagnosis earlier than they currently do.

(J Med Internet Res 2022;24(11):e41288) doi:10.2196/41288

KEYWORDS

sleep disorders; search engine queries; search advertising; internet; Bing; sleep; machine learning; questionnaire

Introduction

Studies have suggested that nearly all adults will experience sleep disturbances over the course of their lives, and that up to 40% of the population experiences chronic sleep disorders [1]. Common sleep disorders include insomnia, sleep apnea, circadian rhythm disturbances, and chronic insufficient sleep. These sleep disorders cause a major health and societal burden, contributing to a loss of personal well-being, development of chronic disease, loss of work productivity and performance, and spikes in accidents that result in needless injury and death.

Sleep disorders are diagnosed by specialized physicians, limiting access and increasing the time to diagnosis in people with sleep disorders [2]. The shortage of trained sleep experts and the lack of geographic accessibility, combined with a convoluted sleep care pathway have led to long waitlists for sleep diagnostics and intervention, with many patients dropping off along the care pathway.

XSL•FO
RenderX

Sleep disorders can be diagnosed using a variety of techniques, using different sources of subjective and objective information and with different levels of invasiveness. Some disorders, such as insomnia, can typically be initially assessed using periodic screeners by health care professionals to identify potential symptoms, while an in-depth sleep history is essential for diagnosis and treatment matching. For example, sleep apnea requires overnight sleep testing (in the laboratory or at home). Other conditions may require in-laboratory daytime sleep testing, at-home continuous sleep-wake monitoring (typically with medical-grade, wrist-worn activity monitors), physical and neurological assessments, and even blood analysis [3]. Interestingly, wearable devices for sleep monitoring have shown limited reliability [4], though some researchers have reported success in predicting sleep quality [5,6].

One approach for initial screening of sleep disorders is through web-based questionnaires. Dayzz is a digital sleep program that provides comprehensive web-based care for multiple sleep conditions. As part of the program, the Digital Sleep Questionnaire (DSQ) [7] is administered. The DSQ is a brief, clinically validated, questionnaire that, using machine learning algorithms, can provide an assessment of 4 common sleep disorders.

People's web-based activity, especially their interactions with search engines, have been shown to enable early screening for a range of medical conditions including diabetes [8], Parkinson disease [9], stroke [10], and several types of cancer [11,12]. The reason for the ability to screen using web-based interactions include the ubiquity of these services in people's daily lives and their service as a gateway to information (especially health information) for the vast majority of the population [13]. Interactions with search engines are especially useful in this regard because they are plentiful and are perceived as sufficiently anonymous, allowing people to express their truthful information needs, without fear that their sensitive or personal information will be divulged to third parties [14]. We note that social media platforms such as Twitter have also been used to characterize sleep issues [15] but not sleep disorders.

In this study, we aim to evaluate whether search engine activity can be used to screen for possible sleep disorders. We show that a machine learning model based on search engine activity can predict the risk for specific sleep disorders, although currently, the quality of the prediction is insufficient as a sole method for screening. Nevertheless, we believe that evaluable web-based information, easily collected and processed with little effort on part of the physician and with low burden on the individual, can assist in the diagnostic process and possibly drive people to seek diagnosis earlier than they currently do.

Methods

Overview

We implemented a novel experimental protocol whereby participants are recruited through internet-based advertisements after they indicated their interest in sleep disorders. People who clicked on the advertisements and consented to participate in the study were referred to a digital sleep questionnaire and were promised general feedback on their sleep patterns as an incentive to complete the questionnaire. They were then asked for permission to contribute their search engine data to the study. Finally, we linked the outcome of the sleep questionnaire with search engine data of the people who consented. These data are used to predict the questionnaire outcome.

Recruitment

We advertised on Microsoft Advertising to US-based users of Bing who queried for one of a range of sleep disorder–related terms (see Multimedia Appendix 1); those people were shown 1 of 12 advertisements, which are listed in Multimedia Appendix 2. People who clicked on the advertisements were referred to a dedicated landing page.

The web page described the experiment to potential participants. Those who consented to join were administered the DSQ [7]—a brief, clinically validated questionnaire that, using machine learning algorithms, can provide an assessment of 4 common sleep disorders: risk for obstructive sleep apnea (OSA), delayed sleep phase syndrome (DSPS), insomnia, and insufficient sleep syndrome (ISS). They were then provided general feedback on their sleep and asked if they agree to have their Bing search data linked to the results of the DSQ and analyzed.

The data of those users who consented were extracted from up to 1 year prior to the date of questionnaire completion and up to the date of the questionnaire. The data comprised an anonymized user identifier, the time and date of the query, the zip code of their location, and indicators calculated from user interactions with search results' pages, including the duration of the search session, the number of links clicked, the time they spent on each link, whether automatic spelling correction was used, and whether the user scrolled down the search page.

The campaign was gradually improved over time by providing a conversion optimization signal to the advertising campaign, such that completion of the DSQ was considered a conversion "worth" US $1 and the same combined with downloading of a dedicated sleep improvement app was set to US $10.

Analysis

We represented the activity of users through a set of aggregate attributes detailed in Textbox 1. These attributes represent several factors that could typify a sleep disorder or are risk factors for it. Specifically, we quantified diurnal activity profiles, as evident from query use; measures of interaction with the search engine (eg, time to first click), which are proxies for cognitive function; and keywords, which reflect risk factors, behaviors that pertain to sleep and its disorders, and the sleep disorders themselves. A list of the latter was developed on the basis of the literature.

Since there is no simple definition of when a day begins and ends for people with sleep disorders, we chose the hour with the lowest level of activity among participating users to be the start of a day (ie, a 24-hour period). Thus, 2 AM was set as the start of the day.

To facilitate comparison of users in our data set with the general population, we extracted a random sample of US-based Bing users who were active on the same dates as the study participants

XSL•FO
RenderX

and presented the percentage of queries at each hour of the day made by these users. We refer to these users as the "control population."

Each user was labeled in accordance with the DSQ as at risk for up to 2 of 4 sleep disorders (OSA, DSPS, insomnia, and ISS). We attempted to predict each of the sleep disorders using a binary classifier. The binary classifier used was a random forest with 100 trees.

The performance of the models was evaluated using leave-one-out cross-validation. The best attribute groups were selected using sequential forward feature selection on the training set.

Only users with at least 14 days of data and 50 queries were retained for analysis.

Textbox 1. List of attributes for predicting sleep disorders.

Query time:

- Percentage of queries at each hour of the day
- Time between consecutive queries (average, 5th and 95th percentiles)
- Average first and last hour of queries on each day
- Whether the user had noncontiguous activity during night hours
- The difference between the hourly activity rate on weekdays compared to that on weekends
- Fraction of queries prior to sunrise
- Fraction of queries after sunset

Activity attributes:

- Average click count on each results page
- Average time on links on each results page
- Average session duration
- Average scrolled distance on each results page
- Average time to first click on each results page
- Average time to last click on each results page
- Average, minimum, maximum, and median of the query text probability
- Average number of sessions per day
- Average number of queries that the user repeated
- Average number of queries which mentioned a medical symptom
- Average number of queries that mentioned a medical diagnosis
- Average number of queries that mentioned a medical drug

Terms:

- Number of times a user queried for the following specific terms (terms were used separately, but are grouped here for convenience):
  - Apnea, insomnia
  - Snoring, gasping, headache, sleepiness, waking, snoozing, dozing, awake, alert, appetite, depression, anxiety, diabetes, obesity, weight loss, memory, oxygen, breath, tired, exhausted, fatigue, no energy, nap, hypertension, high blood pressure, HBP, emotion, and mood
  - Jet lag, porn-related terms, accident, mistake, coffee, caffeine, energy drink, alcohol, alcoholic drinks, smoking, nicotine, baby-related terms, attention, concentration, pregnancy, movie, TV, credit union, loan, cannabis, CBD, THC, marijuana, gambling-related terms, online banking, inbox, CNN, and zoom
  - Sleeping pill, hypnotic, sedative, sleep aid, melatonin, and exercise

Query topics:

- The distribution of the users' query topics into 60 topics according to a proprietary classifier.

Demographics:

- User age
- User sex

XSL•FO
RenderX

Ethical Considerations

The methods were performed in accordance with relevant guidelines and regulations and approved by the institutional review board of Advarra Inc (Pro00045152).

Results

We recruited participants between November 11, 2020, and May 19, 2021. Campaign statistics are provided in Multimedia Appendix 2. A total of 582 users clicked on the advertisements and were referred to the DSQ web page. Of them, 397 consented and completed the DSQ. Among users who completed the DSQ, 305 were matched to their Bing data and 132 had at least 14 days of data and 50 queries.

The age and sex distribution, as provided by the advertising system, is shown in Figure 1. Two measures are provided: the percentage of advertisements clicked on among all those shown (referred to as the clickthrough rate [CTR]), and the percentage of questionnaires completed among all those who clicked on the advertisements (denoted by conversion rate). As shown in Figure 1, younger females were more likely to respond to the advertisements by clicking on them (higher CTR) and to complete the DSQ (higher conversion rate). Younger people were overall more likely to complete the DSQ (that is, to "convert").

According to the DSQ, 58 participants were at risk for OSA, 52 for DSPS, 89 for Insomnia, and 46 ISS. The distribution of ages and sex among sleep disorders is shown in Figure 2. When comparing among the conditions, males accounted for a larger proportion of individuals at risk for ISS and those who are at risk of OSA, while those at risk for DSPS accounted for the largest proportion of females. DSPS was most prevalent among young participants; risk for OSA, in older participants.

Figure 1. Clickthrough rate (left) and conversion rate (right) by age and sex. Blue bars denote males and orange bars denote females.

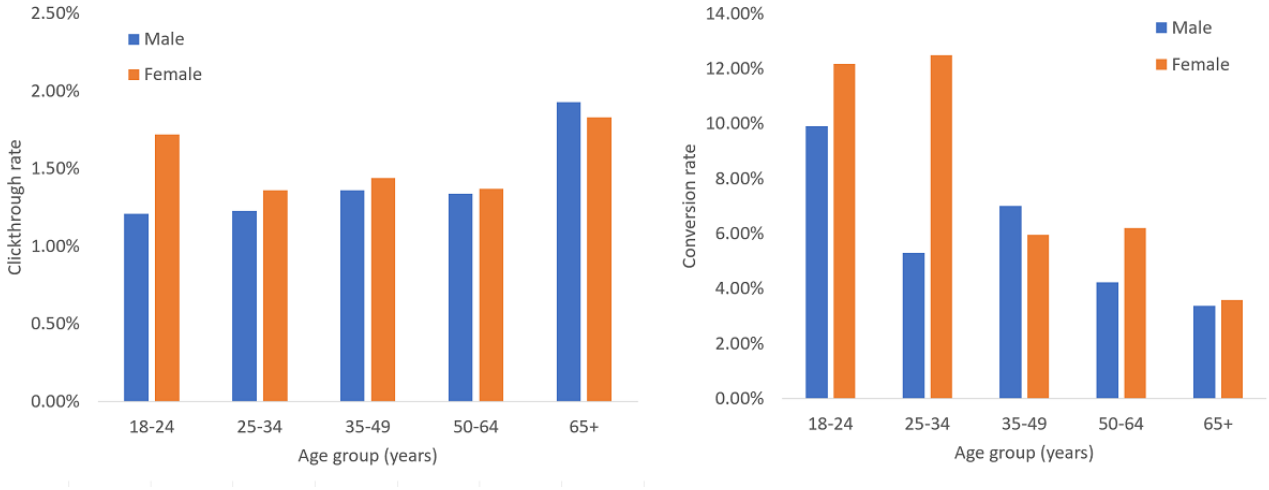
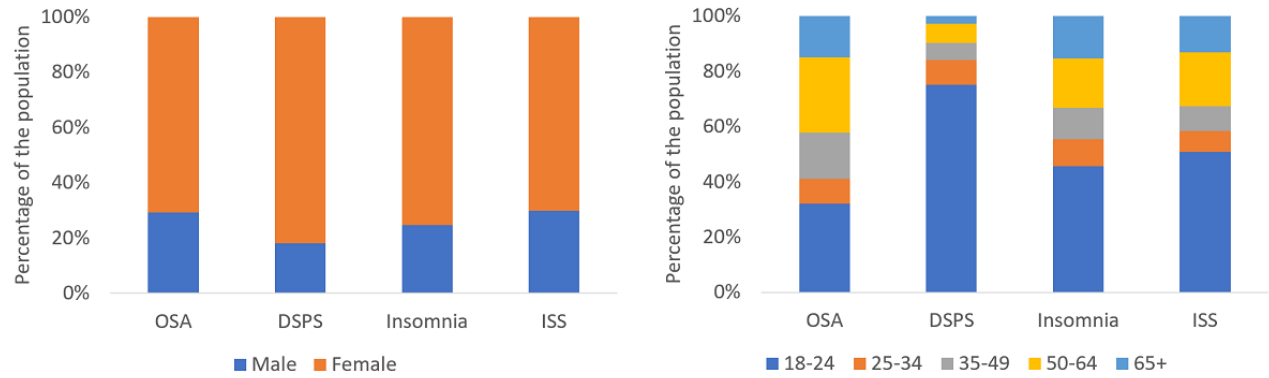


Figure 2. Distribution of ages (left) and sex (right) among sleep disorders. DSPS: delayed sleep phase syndrome; ISS: insufficient sleep syndrome; OSA: obstructive sleep apnea.



We estimated the comorbidity similarity between sleep disorders by the Hamming distance applied to the matrix of participants and their conditions. Figure 3 shows the comorbidity similarity among conditions and the activity profile of users over the day. As shown in Figure 3, people with DSPS were most likely to also have ISS, and those at risk for OSA were most likely to also have insomnia. The daily activity profile of users is similar among sleep disorders (though most similar among users who were at the risk of OSA and those with ISS), but very different from that of the control population. Specifically, the control population begin their activity earlier in the morning hours and end their activity earlier at night, compared to those with sleep disorders. The best correlation between the activity of the control population and the activity of all users with sleep disorders was approached when the activity of the former is shifted by 2 hours for the risk of OSA and insomnia, and by 3 hours for DSPS and ISS. The improvement in correlation due to these time shifts was significantly different ($P<.01$ with Bonferroni correction).

XSL•FO
RenderX

Table 1 shows the performance of the sleep disorder classifiers as evaluated by their area under the curve, together with attribute classes most commonly selected for classification and the number of days of data that approached the best performance. For the latter, we experimented with using different values between 15 and 120 days. Figure 4 shows the receiver operating characteristic curve for each sleep disorder.

We also evaluated which terms would best predict each of the sleep disorders if used without any other attribute classes. This was achieved by finding the terms that increase in prediction error if the values of that term are permuted for the test observations, using the appropriate MATLAB function. The results are summarized in Textbox 2.

Figure 3. Comorbidity similarity among sleep disorders according to the users who shared them (left) and the daily activity profile of users (right). The distance among conditions was computed using Hamming distance of users. DSPS: delayed sleep phase syndrome; ISS: insufficient sleep syndrome; OSA: obstructive sleep apnea.

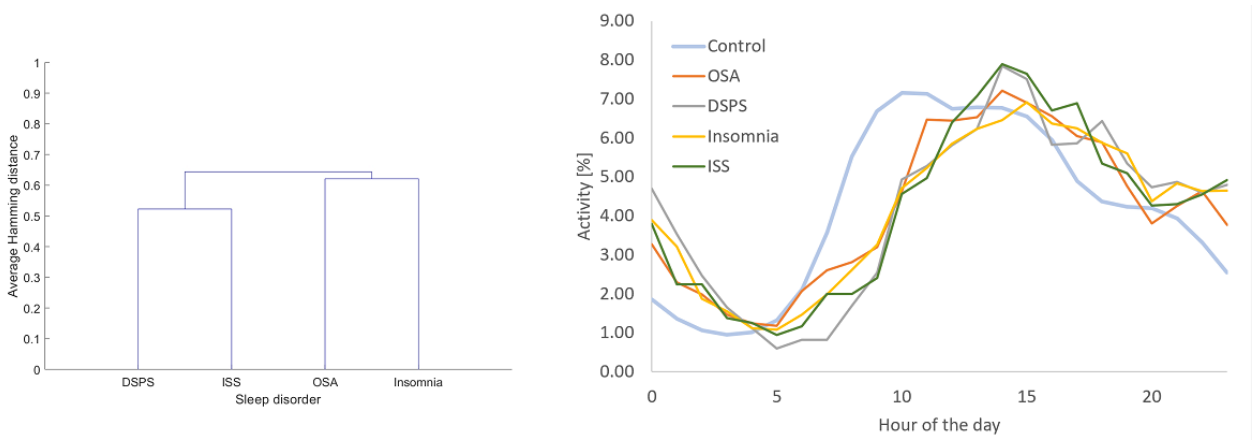
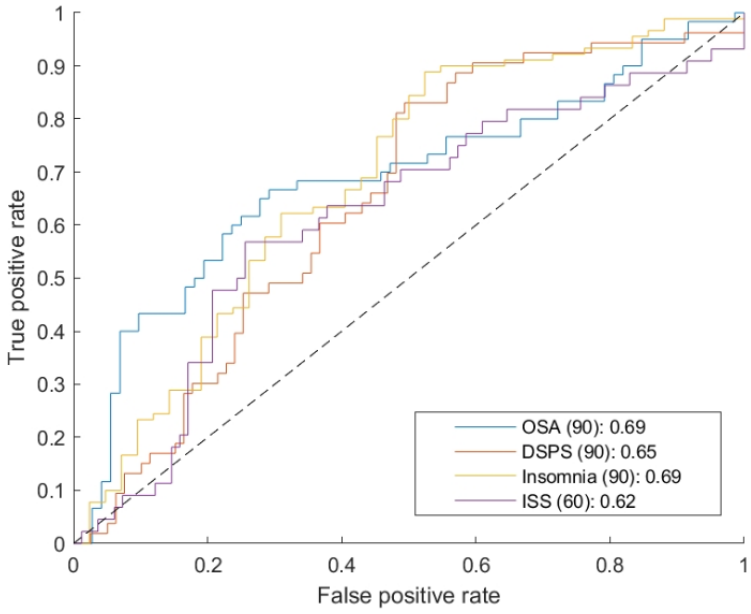


Table 1. Classifier area under the curve, the best number of days selected for use by the classifier from prior to the Digital Sleep Questionnaire and until the day of the questionnaire, and the most commonly selected attributes.

Sleep disorder classifier	Area under the curve	Number of days	Selected attribute classes
Risk for obstructive sleep apnea	0.69	90	Query time, activity attributes, and demographics
Delayed sleep phase syndrome	0.65	90	Query time and demographics
Insomnia	0.69	90	Activity attributes, query topics, and demographics
Insufficient sleep syndrome	0.62	60	Query time, activity attributes, and specific terms

Figure 4. Receiver operating characteristic curve for the 4 sleep disorders. The numbers in the legend indicate the number of days during which data were obtained for each condition and the area under the curve. DSPS: delayed sleep phase syndrome; ISS: insufficient sleep syndrome; OSA: obstructive sleep apnea.



XSL•FO
RenderX

Textbox 2. Most predictive terms of each sleep disorder.

Risk for obstructive sleep apnea:

- Coffee
- Tired, exhausted, fatigued
- Waking
- Hypertension
- Awake

Delayed sleep phase syndrome:

- Headache
- Awake
- Porn-related keywords
- Smoking-related keywords
- Pregnancy-related keywords

Insomnia:

- Coffee
- Online banking, inbox, CNN, zoom
- Sleepiness
- Waking
- Awake

Insufficient sleep syndrome:

- Exercise
- Depression
- Memory
- Anxiety
- Pregnancy-related keywords

Discussion

Sleep disorders have a significant personal and societal cost. However, barriers to wide-scale screening and diagnosis prevent timely treatment for many people with such disorders. There is a need for more accessible and affordable tools that enable remote screening and diagnosis of sleep disorders, especially those which can be performed passively without any effort from the user. Here we test whether interactions with search engines could be used to identify persons at risk for sleep disorders, as a first step in navigating them to further evaluation and intervention.

Our campaign was used to the greatest extent by younger females. In general, younger people were more likely to complete the web-based screening questionnaire. This is in agreement with past research, which showed that younger people are more likely to interact with web-based medical questionnaires [11,16]. However, this also underscores a potential bias in screening using an advertisements-based questionnaire.

The comorbidity among certain sleep disorders found in our sample is in line with that reported in previous studies, showing that risk for OSA and insomnia often coexist [17], and that unmanaged DSPS can often lead to chronic insufficient sleep [18]. Moreover, we found that adults with sleep concerns and potentially increased unwanted or unregulated wakefulness at night tend to browse differently than those without sleep concerns, with search patterns that started and ended later in the day. Perhaps unsurprisingly, we also found that those with natural delays in their biological clock, a distinguishing characteristic of DSPS, were even more shifted in these internet usage patterns than those at risk for other sleep disorders. These findings suggest that when developing algorithms to identify persons at the risk of sleep disorders—in addition to specific sleep—and related health and behavior search content that we found to support differing sleep disorder risk, metadata may offer more fine-grained activity patterns, increasing the predictive performance of the models. Interestingly, independent of the other predictor attribute classes, the search terms that participants used were indicative of their underlying risk for certain sleep disorders; in many cases, the search terms they chose were diagnostic symptoms, functional deficits, or common

XSL•FO
RenderX

comorbidities of the specific sleep disorder that they were at the risk of; for example, hypertension and sleepiness-related terms in the case of risk for OSA [19], and mental health and cognitive deficits in the case of chronic insufficient sleep [20]. Our findings are promising; however, future research should focus on additional content, query parameters, and time periods that can improve and refine the predictive ability of search engine data to identify individuals at the risk of sleep disorders.

The area under the curve for all 4 conditions ranged between 0.62 and 0.69, indicating that search engine data are predictive of sleep disorders, with the prediction being of medium accuracy. We note that sleep disorders are known to be a difficult problem to assess, as evident, for example, in the low agreement among experts at different laboratories on the correct interpretation of diagnostic tests, such as polysomnography, in people with sleep disorders [21], and modest agreement rates between telemedicine and in-person diagnosis [22]. Moreover, our data show that the last 60 to 90 days of search data reach the highest prediction quality. This may be because people responded more readily to the campaign advertisements when their symptoms were more acute.

Our study has several limitations. First, our recruitment occurred among people who were interested or concerned about their sleep. Thus, our results focus on the ability to inform people who are already concerned about possible sleep problems and not the ones who are not part of the general population. Among people who saw our advertisements, younger females were more likely to respond to the advertisements. Among those who clicked on the advertisements, younger people were more likely to complete the DSQ. Thus, our prediction model is biased toward younger people. Future work will attempt to distinguish between healthy populations and those with sleep disorders as well as obtain a more balanced sample of people with sleep disorders. The former will allow a wider application of the model to additional populations, while the latter will ensure that the prediction is applicable to people of all ages.

Conflicts of Interest

IG, MCZ, and NL are employees of Dayzz. EYT is an employee of Microsoft, which is the owner of Bing.

Multimedia Appendix 1
List of campaign keywords.
[DOCX File , 13 KB - jmir_v24i11e41288_app1.docx]

Multimedia Appendix 2
Campaign statistics.
[DOCX File , 80 KB - jmir_v24i11e41288_app2.docx]

References

1. Institute of Medicine, Board of Health Sciences Policy, Committee on Sleep Medicine and Research. Sleep disorders and sleep deprivation: an unmet public health problem. Washington, DC: National Academies Press; 2006.
2. Miller CM. Lack of training in sleep and sleep disorders. Virtual Mentor 2008 Sep 01;10(9):560-563 [FREE Full text] [doi: 10.1001/virtualmentor.2008.10.9.medu1-0809] [Medline: 23211107]
3. Bloom HG, Ahmed I, Alessi CA, Ancoli-Israel S, Buysse DJ, Kryger MH, et al. Evidence-based recommendations for the assessment and management of sleep disorders in older persons. J Am Geriatr Soc 2009 May;57(5):761-789 [FREE Full text] [doi: 10.1111/j.1532-5415.2009.02220.x] [Medline: 19484833]
4. 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]
5. Sathyanarayana A, Joty S, Fernandez-Luque L, Ofli F, Srivastava J, Elmagarmid A, et al. Sleep quality prediction from wearable data using deep learning. JMIR Mhealth Uhealth 2016 Nov 04;4(4):e125. [doi: 10.2196/mhealth.6562]
6. Fagherazzi G, El Fatouhi D, Bellicha A, El Gareh A, Affret A, Dow C, et al. An international study on the determinants of poor sleep amongst 15,000 users of connected devices. J Med Internet Res 2017 Oct 23;19(10):e363 [FREE Full text] [doi: 10.2196/jmir.7930] [Medline: 29061551]
7. Schwartz AR, Cohen-Zion M, Pham LV, Gal A, Sowho M, Sgambati FP, et al. Brief digital sleep questionnaire powered by machine learning prediction models identifies common sleep disorders. Sleep Med 2020 Jul;71:66-76. [doi: 10.1016/j.sleep.2020.03.005] [Medline: 32502852]
8. Hochberg I, Daoud D, Shehadeh N, Yom-Tov E. Can internet search engine queries be used to diagnose diabetes? Analysis of archival search data. Acta Diabetol 2019 Oct 15;56(10):1149-1154. [doi: 10.1007/s00592-019-01350-5] [Medline: 31093762]
9. Youngmann B, Allerhand L, Paltiel O, Yom-Tov E, Arkadir D. A machine learning algorithm successfully screens for Parkinson's in web users. Ann Clin Transl Neurol 2019 Dec 12;6(12):2503-2509 [FREE Full text] [doi: 10.1002/acn3.50945] [Medline: 31714022]

XSL•FO
RenderX

10. Shaklai S, Gilad-Bachrach R, Yom-Tov E, Stern N. Detecting impending stroke from cognitive traits evident in internet searches: analysis of archival data. J Med Internet Res 2021 May 28;23(5):e27084 [FREE Full text] [doi: 10.2196/27084] [Medline: 34047699]
11. Yom-Tov E. Screening for cancer using a learning internet advertising system. ACM Trans Comput Healthcare 2020 Apr 30;1(2):1-13. [doi: 10.1145/3373720]
12. Soldaini L, Yom-Tov E. Inferring individual attributes from search engine queries and auxiliary information. 2017 Presented at: WWW '17: 26th International World Wide Web Conference; April 3-7, 2017; Perth, WA, Australia. [doi: 10.1145/3038912.3052629]
13. Yom-Tov E. Crowdsourced health: how what you do on the internet will improve medicine. Cambridge, MA: MIT Press; 2016.
14. Pelleg D, Yom-Tov E, Maarek Y. Can you believe an anonymous contributor? On truthfulness in Yahoo! answers. 2012 Presented at: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing; September 3-5, 2012; Amsterdam. [doi: 10.1109/socialcom-passat.2012.13]
15. McIver DJ, Hawkins JB, Chunara R, Chatterjee AK, Bhandari A, Fitzgerald TP, et al. Characterizing sleep issues using Twitter. J Med Internet Res 2015 Jun 08;17(6):e140 [FREE Full text] [doi: 10.2196/jmir.4476] [Medline: 26054530]
16. Yom-Tov E. Active syndromic surveillance of COVID-19 in Israel. Sci Rep 2021 Dec 27;11(1):24449 [FREE Full text] [doi: 10.1038/s41598-021-03977-3] [Medline: 34961786]
17. Smith S, Sullivan K, Hopkins W, Douglas J. Frequency of insomnia report in patients with obstructive sleep apnoea hypopnea syndrome (OSAHS). Sleep Med 2004 Sep;5(5):449-456. [doi: 10.1016/j.sleep.2004.03.005] [Medline: 15341889]
18. Nesbitt AD. Delayed sleep-wake phase disorder. J Thorac Dis 2018 Jan;10(Suppl 1):S103-S111 [FREE Full text] [doi: 10.21037/jtd.2018.01.11] [Medline: 29445534]
19. Calhoun DA. Obstructive sleep apnea and hypertension. Curr Hypertens Rep 2010 Jun 20;12(3):189-195. [doi: 10.1007/s11906-010-0112-8] [Medline: 20424946]
20. Zhai L, Zhang H, Zhang D. Sleep duration and depression among adults: a meta-analysis of prospective studies. Depress Anxiety 2015 Sep 05;32(9):664-670. [doi: 10.1002/da.22386] [Medline: 26047492]
21. Collop NA. Scoring variability between polysomnography technologists in different sleep laboratories. Sleep Medicine 2002 Jan;3(1):43-47. [doi: 10.1016/s1389-9457(01)00115-0]
22. Yurcheshen ME, Pigeon W, Marcus CZ, Marcus JA, Messing S, Nguyen K, et al. Interrater reliability between in-person and telemedicine evaluations in obstructive sleep apnea. J Clin Sleep Med 2021 Jul 01;17(7):1435-1440 [FREE Full text] [doi: 10.5664/jcsm.9220] [Medline: 33687321]

Abbreviations

CTR: clickthrough rate
DSPS: delayed sleep phase syndrome
DSQ: Digital Sleep Questionnaire
ISS: insufficient sleep syndrome
OSA: obstructive sleep apnea

Edited by G Eysenbach; submitted 21.07.22; peer-reviewed by S Moga, J Roland; comments to author 21.10.22; revised version received 25.10.22; accepted 15.11.22; published 23.11.22.

Please cite as:
Cohen Zion M, Gescheit I, Levy N, Yom-Tov E
Identifying Sleep Disorders From Search Engine Activity: Combining User-Generated Data With a Clinically Validated Questionnaire
J Med Internet Res 2022;24(11):e41288
URL: https://www.jmir.org/2022/11/e41288
doi:10.2196/41288
PMID:36416870

©Mairav Cohen Zion, Iddo Gescheit, Nir Levy, Elad Yom-Tov. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Countering Antivax Misinformation via Social Media: Message-Testing Randomized Experiment for Human Papillomavirus Vaccination Uptake

Sunny Jung Kim[1,2], MA, MS, PhD; Jenna E Schiffelbein[3], MPH, CHES; Inger Imset[3,4], BA, NBC-HWC; Ardis L Olson[3,4,5], MD

[1]Health Behavior and Policy, School of Medicine, Virginia Commonwealth University, Richmond, VA, United States

[2]Massey Cancer Center, Virginia Commonwealth University, Richmond, VA, United States

[3]Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Hanover, NH, United States

[4]Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Hanover, NH, United States

[5]Department of Pediatrics, Geisel School of Medicine at Dartmouth, Hanover, NH, United States

Corresponding Author:
Sunny Jung Kim, MA, MS, PhD
Health Behavior and Policy
School of Medicine
Virginia Commonwealth University
830 E Main Street 4th Fl
Richmond, VA, 23298
United States
Phone: 1 8046284688
Email: sun.jung.kim@vcuhealth.org

Abstract

Background: Suboptimal adolescent human papillomavirus (HPV) vaccination rates have been attributed to parental perceptions of the HPV vaccine. The internet has been cited as a setting where misinformation and controversy about HPV vaccination have been amplified.

Objective: We aimed to test message effectiveness in changing parents' attitudes and behavioral intentions toward HPV vaccination.

Methods: We conducted a web-based message-testing experiment with 6 control messages and 25 experimental messages and 5 from each of the 5 salient themes about HPV vaccination (theme 1: safety, side effects, risk, and ingredient concerns and long-term or major adverse events; theme 2: distrust of the health care system; theme 3: HPV vaccine effectiveness concerns; theme 4: connection to sexual activity; and theme 5: misinformation about HPV or HPV vaccine). Themes were identified from previous web-based focus group research with parents, and specific messages were developed by the study team using content from credible scientific sources. Through an iterative process of message development, the messages were crafted to be appropriate for presentation on a social media platform. Among the 1713 participants recruited via social media and crowdsourcing sites, 1043 eligible parents completed a pretest survey questionnaire. Participants were then randomly assigned to 1 of the 31 messages and asked to complete a posttest survey questionnaire that assessed attitudes toward the vaccine and perceived effectiveness of the viewed message. A subgroup of participants (189/995, 19%) with unvaccinated children aged 9 to 14 years was also assessed for their behavioral intention to vaccinate their children against HPV.

Results: Parents in the experimental group had increased positive attitudes toward HPV vaccination compared with those in the control group (t_{969}=3.03, *P*=.003), which was associated with increased intention to vaccinate among parents of unvaccinated children aged 9 to 14 years (*r*=1.14, *P*=.05). At the thematic level, we identified 4 themes (themes 2-5) that were relatively effective in increasing behavioral intentions by positively influencing attitudes toward the HPV vaccine (χ^2_5=5.97, *P*=.31, root mean square error of approximation [RMSEA]=0.014, comparative fit index [CFI]=0.91, standardized root mean square residual [SRMR]=0.031). On the message level, messages that provided scientific evidence from government-related sources (eg, the Centers for Disease Control and Prevention) and corrected misinformation (eg, "vaccines like the HPV vaccine are simply a way

XSL•FO
RenderX

for pharmaceutical companies to make money. That isn't true") were effective in forming positive perceptions toward the HPV vaccination messages.

Conclusions: Evidence-based messages directly countering misinformation and promoting HPV vaccination in social media environments can positively influence parents' attitudes and behavioral intentions to vaccinate their children against HPV.

(J Med Internet Res 2022;24(11):e37559) doi:10.2196/37559

KEYWORDS

misinformation; vaccine hesitancy; vaccine communication; social media; human papillomavirus; HPV; HPV vaccine

Introduction

Human Papillomavirus Vaccination in the United States

Human papillomavirus (HPV) is a group of >200 viruses that can infect all genders [1]. HPV infection can cause 6 types of cancer, including cancers of the penis, vulva, cervix, vagina, anus, and oropharynx [1]. The HPV vaccine Gardasil 9 protects against infection with 7 high-risk types of HPV that cause cancer and 2 types that cause genital warts [2]. Within 6 years after the HPV vaccination became available, studies found that the HPV vaccine had lowered the rate of cancer-causing HPV infections by 64% and the risk of cervical precancer by 47% [3]. According to a report by the Centers for Disease Control and Prevention (CDC) published in 2018, a total of 92% of HPV-associated cancers can be prevented by the HPV vaccine [4]. Despite evidence that the HPV vaccine is highly effective in preventing cancers caused by HPV infections, the percentage of adolescents starting the vaccine series in the United States is lower compared with the percentage receiving other routinely recommended vaccines [5]. Only 51% of teens in the United States received the full HPV vaccine series in 2018, which is far lower than that in other industrialized countries, such as the United Kingdom and the Healthy People 2020 goal of 80% [6-8].

Vaccine Hesitancy and Misinformation on Social Media

Since the HPV vaccine was approved in 2006, the prevalence of misinformation on social media as well as the lack of educational campaigns regarding HPV vaccines have contributed to HPV vaccine hesitancy and refusal [9-11]. Within the first 10 years since the approval of the vaccine, nearly 40% of Facebook posts about the HPV vaccine contained texts, links, and images that amplified the perceived risks of the HPV vaccine [12]. Studies have reported that parents' knowledge and HPV vaccine hesitancy are influenced by various factors and media sources, including repeated exposure to misinformation on social media [2,13,14]. One study [15] found that myths and safety concerns about the HPV vaccine were widespread on Twitter, and Twitter users were more likely to be exposed to tweets reporting safety concerns than tweets supporting the evidence of HPV vaccine effectiveness and safety. When looking at aggregated Twitter data, researchers found that vaccination rates were lower in the United States, where tweets on negative representations of vaccines, misinformation, and conspiracies were most prevalent [16]. Parents who declined to vaccinate their children against HPV were shown to mistrust health care systems and have negative attitudes toward HPV vaccination [11]. These psychological states, resulting from repeated exposure to misinformation about HPV vaccination, contribute to parents' resistance and hesitancy to follow the HPV vaccine recommendation [17]. Effectively addressing misinformation is critical for resolving vaccine hesitancy and refusal [16,18].

Health Communication on Social Media for HPV Vaccine Uptake

Health communication and education programs play a vital role in countering misinformation regarding HPV vaccines. According to the CDC, strategic mass media communication is one of the best approaches to addressing a specific public health issue [19,20]. Web-based platforms offer a cost-effective outlet to share, convey, and disseminate health information without geographic constraints and can be leveraged to engage hard-to-reach populations and monitor health behaviors in real time [2]. Despite the potential public health impact and benefits of public health messages that can be promoted on social media, a substantial knowledge gap exists in how to design, examine, and identify effective messages for social media campaigns that are persuasive in promoting HPV vaccination uptake. Given the effect of misinformation on eliciting vaccine hesitancy, it is critical to develop and evaluate accurate and persuasive social media campaign messages.

In this study, we compared different message themes related to HPV vaccine hesitancy and refusal and examined the themes and messages that were more effective in promoting HPV vaccine uptake. We hypothesized that exposure to HPV vaccine uptake messages will influence parents' behavioral intention to vaccinate their children against HPV by improving their attitudes toward the vaccine.

Methods

Participant Recruitment and Eligibility

We recruited participants through paid advertisements on Facebook and Amazon's Mechanical Turk (MTurk). MTurk is a web-based labor market with more than 500,000 anonymous workers worldwide, where requesters distribute tasks and registered workers complete them. MTurk has been widely used for various research purposes such as surveys, cognitive tasks, and web-based experiments [21,22]. Prospective participants who viewed our recruitment messages on MTurk or Facebook were redirected to a secure survey platform for consent and eligibility screening. In our screening survey, we assessed participant characteristics such as ZIP code, sex, age, and gender of children. To be eligible, prospective participants needed to be residents of the United States, have at least 1 child, and

XSL•FO
RenderX

provide their ZIP code and state of residence. Eligible participants who provided consent were then directed to the message-testing experiment.

Message-Testing Experiment

Once directed to the message-testing experiment, participants completed a pretest survey (see Multimedia Appendix 1 for survey instruments) that assessed baseline attitudes and knowledge about HPV and HPV vaccination as well as their behavioral intention to vaccinate their children against HPV. After completing the pretest survey questions, participants were randomly assigned to 1 of 31 messages (25 experimental messages and 6 control messages), viewing 1 message per participant. All the stimulus messages were presented in a simulated social media environment (see Multimedia Appendix 2 for messages). To experimentally control for high elaboration likelihood (eg, processing information carefully rather than heuristically), participants were asked to view the assigned message carefully and report what they read and viewed [23]. After exposure to the assigned message, participants were asked to complete the posttest surveys. When the participants completed the posttest survey, they received educational resources on HPV vaccination.

Message Development

We conducted a series of virtual focus groups with parents to identify the key barriers and reasons for HPV vaccine hesitancy and refusal (results reported elsewhere) before this study. Through focus group discussions, we identified 5 salient themes to address in our messages on HPV vaccination.

1. Safety, side effects, risk, and ingredient concerns and long-term or major adverse events
2. Distrust of the health care system
3. HPV vaccine effectiveness concerns
4. Connection to sexual activity
5. Misinformation about HPV or HPV vaccine

We developed 5 experimental messages for each theme and 6 control messages on e-cigarettes. The experimental messages were grounded in scientific information from published articles and reputable health organizations (eg, CDC, National Cancer Institute, and American Cancer Society). The experimental messages were designed to help the audience understand HPV and HPV vaccination and included a call to action to vaccinate their children against HPV.

Message Framing

Message-framing tactics were applied to change parents' attitudes and behavioral intentions toward HPV vaccination. To avoid confounding effects, we controlled for elements that were not part of the experimental manipulation designs. First, we controlled the length of the messages to no more than 130 words and the message source across all 31 messages to be identical and neutral (ie, the same Facebook account appeared as the poster of the simulated Facebook messages). Second, we controlled the message structure: each message started with introductory statements, included tailored statements specific to the message theme, provided a photo or link to a video that was congruent with the message, and ended with the same call to action. We countered common myths and misinformation and added the correct information. This message structure is consistent across all stimulus messages.

The research team conducted multiple iterative review processes with content experts in the fields of adolescent health, cancer prevention, and health communication before finalizing the messages (Multimedia Appendix 2). Through a web-based pre-post randomized message-testing experiment, we examined the persuasion effects of these messages on changing parents' attitudes toward the vaccine and their behavioral intention to vaccinate their children against HPV.

Measures

Pretest Survey

Eligible participants were asked to answer pretest questions assessing their baseline knowledge of and attitudes toward HPV and HPV vaccination, response efficacy beliefs, and behavioral intention to vaccinate their children against HPV. For response efficacy, we developed items that measured the degree to which participants thought the HPV vaccine was effective in preventing cervical cancer and mouth or throat cancer on a 5-point Likert scale (1=not at all effective; 5=extremely effective) with an additional option indicating "I don't know." To measure attitudes toward HPV vaccination, we used 9 items from a study by Kim and Niederdeppe [24] and modified them to assess parents' attitudes toward the HPV vaccine on the 7-point semantic differential scale, including measures of *bad-good*, *harmful-beneficial*, *useless-useful*, and *unsafe-safe* (see Multimedia Appendix 1 for the full list of survey items).

To assess the vaccination status of their children against HPV, we first asked, "Do you have any sons (or any daughters)?" Among parents who indicated having sons or daughters, we asked "How old are your sons (or daughters)? Please select all that apply." Multiple choice options were available to indicate the ages of sons or daughters ("less than 5 years old," "5 to 8 years old," "9 to 14 years old," "15 to 18 years old," and "older than 18 years old)." For those who reported having sons or daughters aged between 9 and 14 years, we asked, "Has your 9-14 year-old son (or daughter) been vaccinated for HPV? If you have more than one son (or daughter) 9-14 years old, please answer about the son (or daughter) who had the most recent birthday." Four response options were available, including "Yes, s/he has received two or more HPV vaccine shots," "Yes, s/he has received one HPV vaccine shot," "No, s/he has not received any HPV vaccine shots," and "I don't know." For parents who indicated not receiving any HPV vaccine shots or "I don't know," we displayed the following behavioral intention item on a 5-point Likert scale, "Thinking about the same 9-14 years old son (or daughter), how likely is it that s/he will receive the HPV vaccine in the next 12 months?" The response options were 1=very likely and 5=very unlikely. Greater values indicate a lower intention to not vaccinate their unvaccinated children against HPV.

Posttest Survey

During the posttest survey, in addition to remeasuring the items in the pretest survey, such as attitudes toward HPV vaccination, participants were guided to answer a series of additional posttest

XSL•FO
RenderX

items, including manipulation check questions to ensure participants had viewed the messages, as well as message perceptions. To understand message perceptions, we adapted 5 items from the message sensation value scale [25] and 4 items from the perceived message effectiveness scale [26] and modified their wording to fit into the study context. Message sensation value, the degree to which message features elicit affective and arousal responses, was assessed on a 7-point Likert scale, and perceived message effectiveness was measured on a 5-point Likert scale. These 2 questionnaires were reported separately.

An identical behavioral intention question was displayed for a subgroup of parents who indicated at pretest having unvaccinated children aged 9 to 14 years (see Multimedia Appendix 1 for survey measures).

Ethical Considerations

All study procedures were reviewed and approved by the Committee for the Protection of Human Subjects (the institutional review board) at Dartmouth College, and the institutional review board reliance has been approved by Virginia Commonwealth University and Dartmouth College (HM20014090 and MOD00009013, respectively). The study team obtained a waiver for documenting participant signatures during the consenting process; an information sheet about the study was provided to prospective participants, and they were asked to click *next* to provide their consent and continue to participate in the study. After a trained researcher verified worker IDs entered into the survey, participants recruited from MTurk who completed the experiment were compensated at US $1.80. Given the length of our experiment, this compensation rate on MTurk was acceptable, as the average compensation rate was 10 cents per minute. Worker IDs were deleted and not linked to any of the study data before analysis. Participants recruited from paid advertisements on Facebook had a chance to enter a drawing to receive one of 50 e-gift cards each worth US $10. To protect the confidentiality of participants who were recruited via Facebook paid advertisements, the study team did not collect participant names and only collected email addresses for the purposes of compensation administration; those email addresses were collected, stored, and managed securely separately from the study data.

Data Analysis

We first conducted descriptive statistics on demographic variables with the sample and by comparing experimental and control groups. These subgroup comparisons were conducted for pretest variables, including prior knowledge and response efficacy beliefs measures. Mean-based composite scores were generated to assess attitudes toward HPV vaccination on the pretest (Cronbach α=.96; mean 4.98, SD 1.50) and posttest survey items (Cronbach α=.96; mean 5.11, SD 1.47). The attitude change scores were calculated by subtracting the pretest composite values from the posttest composite values for each participant (mean difference 0.12, SD 0.92). Greater values indicate more positive changes in attitudes toward HPV vaccination. Five posttest items assessing the sensation value of the viewed messages on a 7-point Likert scale were averaged to form a composite score (Cronbach α=.92; mean 4.57, SD 1.50). Greater values indicate a more positive sensation toward the messages. A total of 4 posttest items measuring perceived message effectiveness on a 5-point Likert scale were averaged to form a composite score (Cronbach α=.80; mean 3.29, SD 0.99). Greater values indicate a greater perceived effectiveness of the message.

For thematic-level analyses, a path-modeling approach was used to model persuasion pathways from message exposure to attitudinal changes toward HPV vaccination, which indirectly influenced behavioral intentions to vaccinate. For the eligibility of thematic-level analyses, which included behavioral intention measures for HPV vaccination in the next 12 months measured at the posttest survey, we only included a subset of data from 189 parents with children aged between 9 and 14 years who had not received any HPV vaccine shot. We used 5 dummy variables to generate exogenous constructs in the path model. The control condition was used as the reference group. Path coefficients from dummy-coded variables to the attitudinal change construct were the magnitude of the associations compared with those from the reference group (control group).

For individual message-level analyses, we used the ANOVA with a Bonferroni correction within each theme. Composite scores of measures assessing attitudes toward HPV vaccination, sensation values of the viewed message, and perceived message effectiveness were evaluated in the ANOVA framework to statistically identify relatively more effective messages within each of the 5 thematic blocks.

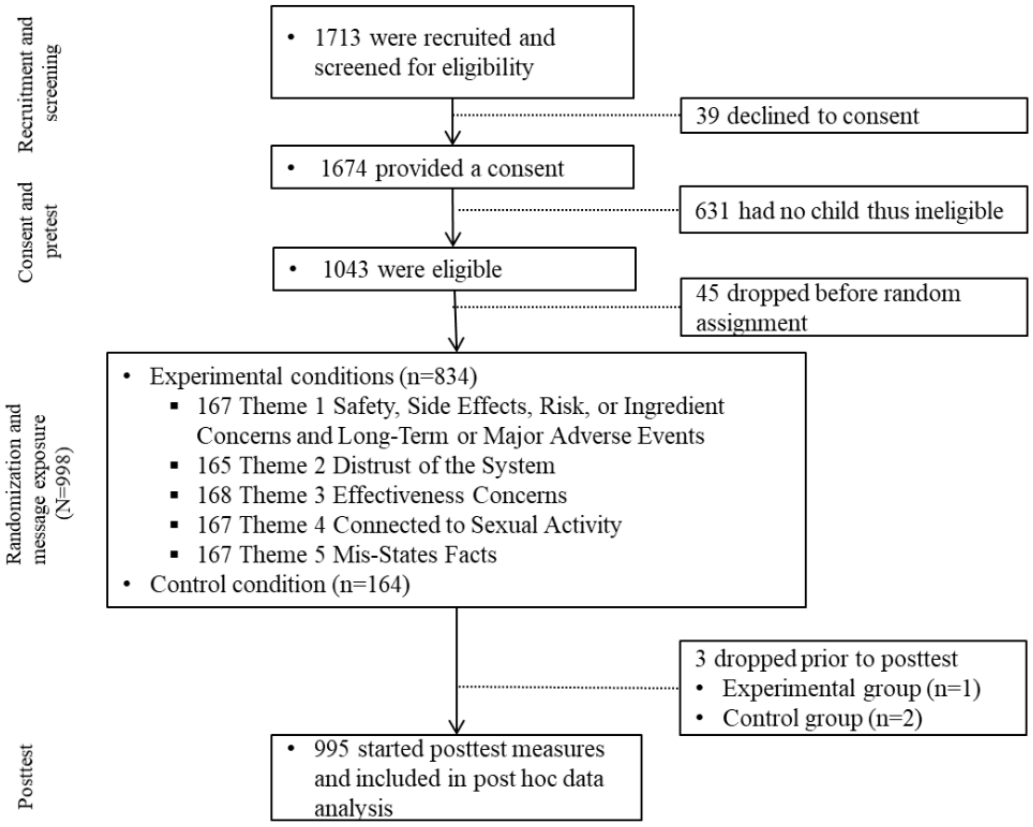
Results

Participants

A total of 1674 respondents provided consent and participated in the eligibility screening survey, and 1043 (62.31%) were eligible to participate in the study. Among the 1043 participants, 45 (4.31%) left the study before being randomized to a message, and an additional 3 participants left during the posttest questionnaire, leading to a final sample size of 998 for pretest data analysis and 995 for the posttest data analysis (Figure 1). The average age was 35.39 (SD 10.58) years; 616 (61.7%) were female; 202 (37.5%) had an annual income of US ≤$50,000 (see Table 1 for demographic characteristics of our participants).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram showing the flow of participants through pre-post message-testing experiment.



XSL•FO
RenderX

Table 1. Baseline characteristics by the assigned group (N=998).

Characteristics	Experimental group (n=834)	Control group (n=164)	Total (N=998)
Age (years), mean (SD)	37.24 (10.5)	34.77 (10.5)	35.39 (10.6)
Female, n (%)	520 (62.4)	96 (58.5)	616 (61.7)
Income			
<US $50,000, n (%)[a]	172 (38.6)	30 (32.3)	202 (37.5)
Education			
No college, n (%)	58 (13.1)	16 (17.2)	74 (13.8)
Daughter (yes), n (%)	653 (78.3)	126 (76.8)	779 (78.1)
Son (yes), n (%)	585 (70.1)	125 (76.2)	710 (71.1)
Rural, n (%)	180 (22.4)	36 (23.2)	216 (22.6)
Previous knowledge at pretest			
Heard of HPV[b]? (Yes), n (%)[c]	730 (88.1)	142 (87.1)	872 (87.9)
Heard of the HPV vaccine? (Yes), n (%)[d]	691 (83.3)	136 (83.4)	827 (83.3)
Perceived importance[e] of HPV vaccines, mean (SD)	2.44 (1.3)	2.34 (1.4)	2.42 (1.3)
Response efficacy[f] at pretest of HPV vaccine			
In preventing cervical cancer? (Do not know), n (%)	118 (14.1)	22 (13.4)	140 (14)
In preventing mouth or throat cancer? (Do not know), n (%)	234 (28.1)	53 (32.3)	287 (28.8)
In preventing cervical cancer? mean (SD)	2.53 (1.6)	2.48 (1.5)	2.52 (1.6)
In preventing mouth or throat cancer? mean (SD)	2.26 (1.8)	2.08 (1.8)	2.23 (1.8)

[a]Education and income were assessed in the later phase of the experiment after posttest measures; thus, the sample size for these socioeconomic measures was smaller (n=539 and n=536 for reporting income and education, respectively).

[b]HPV: human papillomavirus.

[c]N=992 for this row because 6 people did not answer this question.

[d]N=993 for this row because 5 people did not answer this question.

[e]The perceived importance of vaccinating their children against HPV was measured on a 5-point Likert scale, where 1=extremely important and 5=not important.

[f]Response efficacy belief at baseline=greater values indicate stronger belief in response efficacy.

Outcomes on Pretest Measures

Nearly 88% (872/992, 87.3%) of the participants reported having heard of HPV before the study, and 83.3% (827/993) reported that they had heard of the HPV vaccine. At the pretest survey, 14% (140/998) of participants indicated that they did not know about the efficacy of the HPV vaccine in the prevention of cervical cancer, and 28.8% (287/998) of participants did not know of the efficacy of the HPV vaccine in preventing mouth or throat cancer (Table 1).

From Message Exposure to Behavioral Intention to Vaccinate

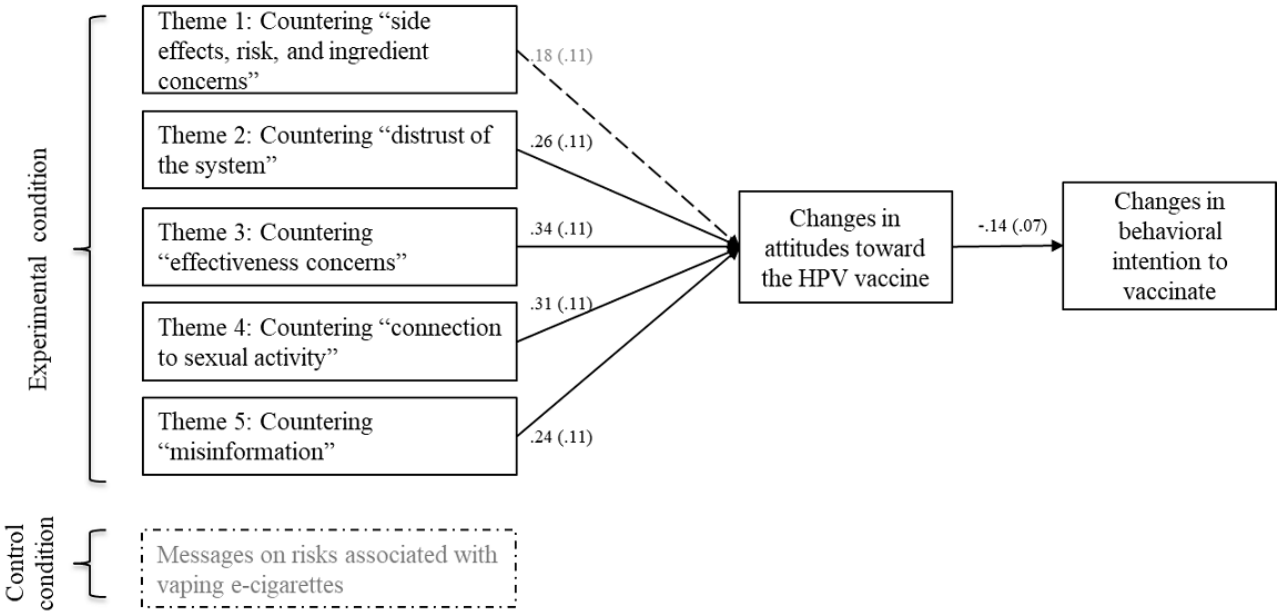
At a group level (experimental group vs control group), experimental messages about the HPV vaccine significantly increased positive attitudes toward HPV vaccination compared with the control messages about e-cigarettes (t_{969}=3.03, *P*=.003). Parents' behavioral intention to vaccinate their children against HPV was significantly associated with the increase in positive attitudes toward HPV vaccination (B=1.14, *P*=.05).

Thematic-Level Outcomes

At the thematic level, according to the results from path modeling, messages countering 4 themes—theme 2: distrust of the health care system, theme 3: HPV vaccine effectiveness concerns, theme 4: connection to sexual activity, and theme 5: misinformation about HPV or HPV vaccine—were more likely to increase behavioral intention to vaccinate, in part because of the increased positive attitudes toward the vaccine (χ^2_5=5.97, *P*=.31, root mean square error of approximation [RMSEA]=0.014, comparative fit index [CFI]=0.91, standardized root mean square residual [SRMR]=0.031). Figure 2 denotes thematic-level path analysis results from message exposure to changes in behavioral intention among parents with unvaccinated children aged 9 to 14 years (n=189). Greater values of the changes in behavioral intention in Figure 2 indicate lower behavioral intention to vaccinate. A dotted line in Figure 2 indicates a nonsignificant path. Straight lines with standardized path coefficients are at the 0.05 level of significance. Greater values to the attitudes construct indicate more positive attitudinal changes toward the HPV vaccine after viewing the message.

XSL•FO
RenderX

Figure 2. Thematic-level path analysis. HPV: human papillomavirus.



Message-Level Outcomes

Within each theme, the influence of each message on changing vaccine attitudes, sensation values of the message viewed, and perceived message effectiveness were examined. For example, at the individual message level, compared with message IDs b, d, and e in theme 1, the message ID c in theme 1, which directly countered safety concerns on the HPV vaccine with an educational video that addressed HPV vaccine effectiveness, significantly increased the sensation value (eg, positive attitudes) toward the message effectiveness (mean 5.38, SD 1.07, $F_{4,160}$=6.59, P<.01) and perceived message effectiveness (mean 3.79, SD 0.83, $F_{4,160}$=4.06, P=.004; Table 2). Within theme 2, the message ID b (mean 5.07, SD 1.31) countering a myth about a pharmaceutical money-making scheme (eg, "vaccines like the HPV vaccine are simply a way for pharmaceutical companies to make money. That isn't true") was significantly more effective in increasing positive attitudes toward the message than the message ID c (mean 3.90, SD 1.76) that countered a myth about a doctor money-making scheme (eg, "We know that some people worry that vaccines like the HPV vaccine are just helping doctors make money"; $F_{4,158}$=3.01, P=.02). Within theme 3 countering HPV vaccine effectiveness concerns, the message ID c (mean 3.65, SD 0.92) reporting statistical evidence of the HPV vaccine's effectiveness in preventing cancer (eg, "...among women who were vaccinated in Finland 15 years ago, none of them got HPV cancers" and "...we have seen a 71% decrease in HPV infections that cause most HPV cancers and genital warts among teen girls vaccinated in the U.S.") was perceived more effective than the message ID a (mean 2.91, SD 1.08) that stated the importance of HPV vaccination to effectively treat HPV infections (eg, "The HPV vaccine does not treat HPV infections that already exist, though, and there is no treatment for HPV infections. That's why it's so important for boys and girls to get vaccinated before they can ever get exposed to HPV infections"), $F_{4,161}$=3.32, P=.01. Within theme 4, although the ANOVA test reported that the mean values of the 5 messages differed ($F_{4,160}$=3.01, P=.02), the significance did not emerge when a Bonferroni correction was applied. For theme 5, which focused on countering misinformation about HPV or HPV vaccination, the message ID e (mean 3.79, SD 0.72) correcting misinformation about HPV vaccine recommendations (eg, "Why both boys and girls? Because everyone can get HPV cancers caused by HPV" and "Why two doses? For full protection against approximately 93% of HPV cancers, more than one vaccine dose is needed") was more persuasive than the message ID d (mean 3.06, SD 1.01) addressing the misinformation that the HPV vaccine can cause ovarian failure and fertility issues ($F_{4,159}$=3.07, P=.02).

XSL•FO
RenderX

Table 2. Results from ANOVA tests at the message level with a Bonferroni correction. Italicized values indicate a thematic block with a significant difference less than 0.05 before a post hoc test as shown in *F* values in ANOVA. The significance of italicization level is noted next to the *F* value within the block.

Themes and message IDs	Changes in attitudes toward HPV vaccination, mean (SD)	Perceived sensation value toward messages, mean (SD)	Perceived message effectiveness, mean (SD)
Theme 1 (n=162-165)		$F_{4,160}$=6.59, *P*<.001	$F_{4,160}$=4.06, *P*=.004
a	−0.10 (0.84)	*4.69* [e] *(1.42)*	*3.41 (0.98)*
b	−0.09 (0.61)	*4.29* [c] *(1.44)*	*3.19 (1.02)*
c	0.03 (0.89)	*5.38* [b,e] *(1.07)*	*3.79* [e] *(0.83)*
d	0.19 (0.62)	*4.18* [c] *(1.68)*	*3.35 (1.01)*
e	0.40 (0.92)	*3.66* [a,c] *(1.50)*	*2.85* [c] *(1.05)*
Theme 2 (n=163)		$F_{4,158}$=3.01, *P*=.02	Not significant
a	0.27 (0.73)	*4.64 (1.52)*	3.38 (1.07)
b	0.31 (1.17)	*5.07* [c] *(1.31)*	3.45 (0.88)
c	−0.003 (0.71)	*3.90* [b] *(1.76)*	3.12 (1.12)
d	0.16 (0.63)	*4.28 (1.15)*	2.99 (0.96)
e	0.06 (0.83)	*4.37 (1.38)*	3.19 (0.99)
Theme 3 (n=165-166)		Not significant	$F_{4,161}$=3.32, *P*=.01
a	0.12 (0.81)	4.27 (1.48)	*2.91* [c,d] *(1.08)*
b	0.24 (0.87)	4.93 (1.34)	*3.44 (0.83)*
c	0.34 (1.06)	4.92 (1.46)	*3.65* [a] *(0.92)*
d	0.33 (1.14)	4.91 (1.39)	*3.55* [a] *(0.70)*
e	0.10 (0.82)	4.78 (1.44)	*3.30 (0.99)*
Theme 4 (n=165)		$F_{4,160}$=3.01, *P*=.02	Not significant
a	0.24 (1.23)	*4.33 (1.35)*	3.48 (0.95)
b	0.17 (0.68)	*5.09 (1.67)*	3.54 (1.12)
c	0.44 (0.95)	*5.09 (1.44)*	3.79 (0.95)
d	0.03 (0.71)	*4.09 (1.65)*	3.16 (1.1)
e	0.14 (0.75)	*4.61 (1.24)*	3.24 (0.7)
Theme 5 (n=163-164)		Not significant	$F_{4,159}$=3.07, *P*=.02
a	0.004 (1.06)	4.14 (1.54)	*3.29 (0.90)*
b	0.37 (1.67)	4.59 (1.62)	*3.36 (1.10)*
c	−0.03 (1.02)	4.48 (1.57)	*3.15 (0.93)*
d	0.04 (0.85)	4.12 (1.36)	*3.06* [e] *(1.01)*
e	0.29 (0.68)	4.98 (1.47)	*3.79* [d] *(0.72)*

[a,b,c,d,e]Message IDs with a significant difference after Bonferroni correction were reported. Superscript letters indicate a significant difference (*P*<.05) against the message ID in the row within each theme. Theme 1: safety, side effects, risk, and ingredient concerns and long-term or major adverse events. Theme 2: distrust of the health care system. Theme 3: HPV vaccine effectiveness concerns. Theme 4: connection to sexual activity. Theme 5: misinformation about HPV or HPV vaccine. Greater values indicate more positive changes in attitudes toward HPV vaccination, greater sensation and positive attitude toward the message viewed, and greater perceived effectiveness of the message viewed.

XSL•FO
RenderX

Discussion

Principal Findings

Findings from path modeling confirmed that HPV messages can yield positive directions in changing attitudes and behavioral intentions toward the HPV vaccine. Our path model indicated that most of the themes were effective in significantly changing parents' attitudes toward HPV vaccination, which, in turn, was strongly associated with behavioral intention to vaccinate their children against HPV. Those effective communication themes were targeting the distrust of the health care system by correcting misinformation (eg, "doctors/pharmaceutical companies make money-scheme"), addressing HPV vaccine effectiveness concerns, educating on perceived connection to sexual activity, and correcting common myths about HPV vaccination. Given that parental views are often altered because of misinformation that is easily accessible on social media or web-based communities, our study demonstrates that evidence-based messages that directly correct misinformation may be effective in enhancing public knowledge and attitudes about HPV vaccination.

As shown in the ANOVA results, certain messages within each thematic block were more effective than other messages in generating positive perceptions of the message viewed. For example, messages that provided numeric evidence from credible sources (eg, "As you will hear in this video from the Minnesota Department of Health...") were perceived more persuasive than messages that did not refer to a credible source (eg, "The vaccine's safety has continued to be studied in the 12 years..."). It should also be noted that there was no message that generated adverse effects or boomerang effects of persuasion, such as stimulating the opposite stance [27,28].

We acknowledge that attitudes and behavioral intentions toward HPV vaccination are likely to differ according to the level of issue involvement [29], the HPV vaccination status in this case. Thus, in our path model, we included only a subgroup of parents with unvaccinated children aged 9 to 14 years (n=189). Thus, findings based on the path model should be specific to the parents of unvaccinated children in this age group. However, it should be noted that the analyses of constructs related to message evaluations (ie, message sensation and perceived message effectiveness) were based on the full sample of parents. Exposure to media and campaign messages is conceptualized as distal predictors of behavioral changes according to the Integrative Model of Behavioral Prediction [30]. Evaluations of messages from parents who had already vaccinated their children could be different from those of parents who have not yet vaccinated their children. Although one's issue involvement status (ie, whether vaccinated) could influence a parent's attitudes and behavior intentions toward HPV vaccination, in this work regarding perceptions about the messages, we do not have either predefined theoretical justification or statistical power to gauge whether message sensation and perceived message effectiveness will differ by children's vaccination status. Future studies should examine whether children's HPV vaccination status will influence parents' perceived message sensation and perceived message effectiveness.

The prevalence of myths and misinformation about HPV and HPV vaccination on social media influences parents' decisions about HPV vaccination [31,32]. Our study empirically examined the persuasion effects of messages that were strategically designed to counter misinformation about HPV vaccines while promoting HPV vaccination in social media environments. As evidenced in the path model, we demonstrated that systematic designs of communication themes and message components not only influence overall parental attitudes and evaluations of the messages but also can positively change behavioral intention toward HPV vaccination for unvaccinated children. Accumulating evidence suggests that digital technologies, including social media, are ubiquitous in promoting public health, even among often hard-to-reach populations such as rural residents [33]. Our study indicates that social technologies can be leveraged to deliver strategic public health communications to people who may have limited health care resources to intervene in their behaviors and improve their attitudes toward the HPV vaccine [9,22].

Parents' hesitancy toward vaccinating their children against HPV was, in part, because of psychological barriers, such as skepticism about HPV vaccine effectiveness and distrust toward the health care system, which was influenced by misinformation about HPV vaccination. On the basis of the salient themes that emerged as parents' reasons not to follow HPV vaccine recommendations, most of the themes we used in our message designs focused on addressing the lack of knowledge and misunderstandings about HPV vaccine recommendations. For example, we developed messages that directly countered the misinformation that the HPV vaccine is lucrative for health care systems and pharmaceutical industries, which had led some parents not to follow the HPV vaccine recommendations. We also developed multiple messages that directly countered misinformation regarding the side effects and safety issues associated with the HPV vaccine. This type of misinformation has been prevalent on social media, which has been a contributing factor in vaccine refusal and hesitancy among parents [34,35].

It should be noted that our parent sample was slightly younger, more educated, and poorer than the nationwide samples of parents reported by the Pew Research Center's Internet and American Life Project and the US Census Bureau Social and Economic Supplement Surveys [36,37]. For example, among those who reported their income, 37.5% (202/539) of our participants had annual incomes lower than US $50,000, whereas in the 2018 Annual Social and Economic Supplement survey, 32.1% of households with children ≤18 years had an annual income lower than US $50,000 in 2017 [37]. However, given that our study was an experiment examining internal validity, random sampling was not the primary concern in designing the study. Rather, our primary goals centered on ensuring random assignment, assessing the status of the theoretical explanations, and testing whether the findings will replicate in other settings. Our primary goals of the experimental study were to "apply the theory beyond the research setting" and "the degree to which the specific sample represents the population of interest was of less importance" as emphasized in Highhouse [38].

XSL•FO
RenderX

In fact, multiple replication studies support the utility of web-based samples such as MTurk for experimental studies. Through a series of 15 replication experiments comparing results from MTurk convenience samples to probability samples, Coppock [39] confirmed that estimates of causal effects obtained on MTurk samples were similar to those obtained on probability-based national samples. Mullinix et al [40] replicated 20 experiments and found a high level of concordance between estimates obtained from MTurk-based samples and national probability samples. Along with the supporting evidence confirming the utility of web-based convenience samples in experimental studies, we carefully implemented several methods to improve our ability to generalize causal inferences across different settings and contexts. First, to make our findings relevant and generalizable, we generated stimulus materials that reflect social media posts that people would encounter in their real life and conducted the experiment in a real-world setting. These stimuli were designed to simulate the way people view, read, and process social media posts. Second, we conducted formative research to generate the most salient campaign themes and used multiple instances of a stimulus category (HPV vaccine-related posts on social media) to ensure that the experimental stimuli were representative of a predefined population of stimuli (ie, social media posts related to HPV vaccine and misinformation) while controlling for possible third variables (eg, by having the length of the messages consistent across stimuli and using the same source profile across stimulus materials).

Our study methods were designed to enhance our understanding of the theory-based causal process (eg, how parents process social media messages countering misinformation and promoting HPV uptake) and ecological validity (ie, "the realism of the experimental methods, materials, and settings" [38]. Given the relatively small size of the subgroup sample in the path model, we have implications of the path model findings only for parents with unvaccinated children aged 9 to 14 years who use the web, not generalizing across subpopulations of parents with specific background factors.

It should be noted that neither external nor internal validity can be captured through a single experiment. In addition, as Lin et al [41] argue, there may be a trade-off between internal and external validity. We posit that the external validity of our findings should be examined via replication. Moreover, when a number of studies on this topic become available, conducting a meta-analysis of the causal process will help examine whether the message effects found in our study remain reliable across the results of many studies. As per concerns related to generalizability, future research should further examine whether the message effect is moderated by interactions with one or more demographic factors (parents of children in different age groups). A strong replication pattern across samples will improve the credence of the findings.

Limitations

Our findings should be interpreted in light of the limitations imposed by the study designs and settings. First, the experiment was based on a pre-post design that measured message effects without a time lag or long-term follow-up; that is, although our message-testing experiment used random assignment to a stimulus message to build internal validity, perceived message effectiveness and sensation values (eg, attitude toward the messages) measured after message exposure could be driven by immediate recency effect and priming effect that can diminish over time [42]. To measure the long-term effects of message exposure, a separate condition that entails a pre-post design with a long-term follow-up should be implemented to rule out the possible immediate recency and priming effects of message exposure. Future research should also examine the impact of social media campaign messages on vaccination behavior during follow-up assessments.

Second, for our path analyses, we generated dummy-coded exogenous variables to examine persuasion pathways from the conditions (eg, experiment condition=1 and control condition=0) and also from the 5 thematic-level variables, as shown in Figure 2. However, we did not compute dummy-coded variables for the message-level constructs in the path model. This finding was partially due to the sample size. If dummy coding were generated at the message level, the model would entail 25 exogenous variables, leading to overfitting problems and unspecified model fits. To resolve the issue, we instead conducted composite score–based tests using ANOVA with a Bonferroni correction. This approach helped specify the exact message units that have shown effects on persuasion outcomes. In this study, although we used a pre-post randomization experiment as the most suitable design for between-subject tests across conditions and thematic-level analyses, we suggest that future research should consider implementing a within-subject discrete choice experiment [9] to test prespecified message attributes, such as types of message sources and pictorial vs moving images.

Finally, although most of our analyses were based on a sample of over 900 participants, the final path model was based on a subset of the sample (n=189) with children aged between 9 and 14 years, who did not receive any HPV vaccines at the moment, to reflect the age range for the HPV vaccination most recommended by the CDC; that is, we did not ask a hypothetical question to all participants (eg, "*If* you had a child aged between 9 and 14, would you consider HPV vaccination in the next 30 days?"). Instead, we asked about the true behavioral intention for HPV vaccination only among parents with children aged between 9 and 14 years who had not yet vaccinated their children against HPV (189/995, 19%). Future experimental research may replicate the path analysis and further test potential moderators or mediators discussed in the literature. Identifying moderators and mediators involved in the persuasion pathway for HPV vaccination uptake will help to identify effective communication tailoring strategies.

Conclusions and Public Health Implications

Our study highlights that strategic communication efforts in social media environments can offer the opportunity to change parental attitudes and behavioral intentions toward HPV vaccination. Given how ubiquitous social media platforms are today, promoting evidence-based messages—such as ours—on social media may play an important role in promoting accurate health information and enhancing knowledge and attitudes about

XSL•FO
RenderX

HPV and HPV vaccination for cancer prevention. We found that when harnessing social media platforms for public health communications, directly countering dominant misinformation themes and providing accurate science-based information can be particularly effective in promoting HPV vaccination uptake.

Acknowledgments

This research was supported by National Cancer Institute P30 Supplement Grant (P30 CA023108-38S4). The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pretest and posttest survey items.
[PDF File (Adobe PDF File), 133 KB - jmir_v24i11e37559_app1.pdf]

Multimedia Appendix 2
Stimulus messages for the experimental group.
[PDF File (Adobe PDF File), 1888 KB - jmir_v24i11e37559_app2.pdf]

References

1. Forman D, de Martel C, Lacey CJ, Soerjomataram I, Lortet-Tieulent J, Bruni L, et al. Global burden of human papillomavirus and related diseases. Vaccine 2012 Nov 20;30 Suppl 5:F12-F23. [doi: 10.1016/j.vaccine.2012.07.055] [Medline: 23199955]
2. Petrosky E, Bocchini JA, Hariri S, Chesson H, Curtis CR, Saraiya M, Centers for Disease Control and Prevention (CDC). Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on immunization practices. MMWR Morb Mortal Wkly Rep 2015 Mar 27;64(11):300-304 [FREE Full text] [Medline: 25811679]
3. Markowitz LE, Liu G, Hariri S, Steinau M, Dunne EF, Unger ER. Prevalence of HPV after introduction of the vaccination program in the United States. Pediatrics 2016 Mar;137(3):e20151968. [doi: 10.1542/peds.2015-1968] [Medline: 26908697]
4. An Estimated 92% of Cancers Caused by HPV Could be Prevented by Vaccine. Centers for Disease Control and Prevention. 2019 Aug 22. URL: https://www.cdc.gov/media/releases/2019/p0822-cancer-prevented-vaccine.html [accessed 2022-10-02]
5. Senkomago V, Henley SJ, Thomas CC, Mix JM, Markowitz LE, Saraiya M. Human papillomavirus-attributable cancers - United States, 2012-2016. MMWR Morb Mortal Wkly Rep 2019 Aug 23;68(33):724-728 [FREE Full text] [doi: 10.15585/mmwr.mm6833a3] [Medline: 31437140]
6. United States Department of Health and Human Services. How Many Cancers Are Linked with HPV Each Year? Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/cancer/hpv/statistics/cases.htm [accessed 2022-10-02]
7. Walker TY, Elam-Evans LD, Yankey D, Markowitz LE, Williams CL, Fredua B, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13-17 years - United States, 2018. MMWR Morb Mortal Wkly Rep 2019 Aug 23;68(33):718-723 [FREE Full text] [doi: 10.15585/mmwr.mm6833a2] [Medline: 31437143]
8. 2020 Topics and Objectives. Healthy People 2020. 2016. URL: https://www.healthypeople.gov/2020/topics-objectives/topic/immunization-and-infectious-diseases/objectives [accessed 2020-01-27]
9. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
10. Margolis MA, Brewer NT, Shah PD, Calo WA, Gilkey MB. Stories about HPV vaccine in social media, traditional media, and conversations. Prev Med 2019 Jan;118:251-256. [doi: 10.1016/j.ypmed.2018.11.005] [Medline: 30414396]
11. Massey PM, Kearney MD, Hauer MK, Selvan P, Koku E, Leader AE. Dimensions of misinformation about the HPV vaccine on Instagram: content and network analysis of social media characteristics. J Med Internet Res 2020 Dec 03;22(12):e21451 [FREE Full text] [doi: 10.2196/21451] [Medline: 33270038]
12. Luisi ML. From bad to worse II: risk amplification of the HPV vaccine on Facebook. Vaccine 2021 Jan 08;39(2):303-308. [doi: 10.1016/j.vaccine.2020.11.065] [Medline: 33303232]
13. Grimes RM, Benjamins LJ, Williams KL. Counseling about the HPV vaccine: desexualize, educate, and advocate. J Pediatr Adolesc Gynecol 2013 Aug;26(4):243-248. [doi: 10.1016/j.jpag.2013.04.002] [Medline: 24049807]
14. Cipriano JJ, Scoloveno R, Kelly A. Increasing parental knowledge related to the human papillomavirus (HPV) vaccine. J Pediatr Health Care 2018;32(1):29-35. [doi: 10.1016/j.pedhc.2017.06.006] [Medline: 28822674]
15. Neuman WR, Guggenheim L. The evolution of media effects theory: a six-stage model of cumulative research. Commun Theory 2011 May;21(2):169-196. [doi: 10.1111/j.1468-2885.2011.01381.x]

XSL•FO
RenderX

16. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016 Aug 29;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
17. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
18. Oliver JE, Wood T. Medical conspiracy theories and health behaviors in the United States. JAMA Intern Med 2014 May;174(5):817-818. [doi: 10.1001/jamainternmed.2014.190] [Medline: 24638266]
19. Brewer NT, Chapman GB, Rothman AJ, Leask J, Kempe A. Increasing vaccination: putting psychological science into action. Psychol Sci Public Interest 2017 Dec;18(3):149-207. [doi: 10.1177/1529100618760521] [Medline: 29611455]
20. Zimet GD, Rosberger Z, Fisher WA, Perez S, Stupiansky NW. Beliefs, behaviors and HPV vaccine: correcting the myths and the misinformation. Prev Med 2013 Nov;57(5):414-418 [FREE Full text] [doi: 10.1016/j.ypmed.2013.05.013] [Medline: 23732252]
21. Best practices for comprehensive tobacco control programs—2014. Centers for Disease Control and Prevention. Atlanta, GA, USA: U.S. Department of Health and Human Services, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014. URL: https://www.cdc.gov/tobacco/stateandcommunity/best_practices/pdfs/2014/comprehensive.pdf [accessed 2020-01-27]
22. Massey PM, Budenz A, Leader A, Fisher K, Klassen AC, Yom-Tov E. What drives health professionals to tweet about #HPVvaccine? Identifying strategies for effective communication. Prev Chronic Dis 2018 Feb 22;15:E26 [FREE Full text] [doi: 10.5888/pcd15.170320] [Medline: 29470166]
23. Berinsky AJ, Huber GA, Lenz GS. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. Polit Anal 2017 Jan 04;20(3):351-368. [doi: 10.1093/pan/mpr057]
24. Kim SJ, Niederdeppe J. Emotional expressions in antismoking television advertisements: consequences of anger and sadness framing on pathways to persuasion. J Health Commun 2014;19(6):692-709. [doi: 10.1080/10810730.2013.837550] [Medline: 24417482]
25. Palmgreen P, Stephenson MT, Everett MW, Baseheart JR, Francies R. Perceived message sensation value (PMSV) and the dimensions and validation of a PMSV scale. Health Commun 2002;14(4):403-428. [doi: 10.1207/S15327027HC1404_1] [Medline: 12375769]
26. Davis KC, Nonnemaker JM, Farrelly MC, Niederdeppe J. Exploring differences in smokers' perceptions of the effectiveness of cessation media messages. Tob Control 2011 Jan;20(1):26-33. [doi: 10.1136/tc.2009.035568] [Medline: 20852323]
27. Kim SJ, Hancock JT. How advertorials deactivate advertising schema: MTurk-based experiments to examine persuasion tactics and outcomes in health advertisements. Commun Res 2016 Apr 25;44(7):1019-1045. [doi: 10.1177/0093650216644017]
28. Eisenberg IW, Bissett PG, Canning JR, Dallery J, Enkavi AZ, Whitfield-Gabrieli S, et al. Applying novel technologies and methods to inform the ontology of self-regulation. Behav Res Ther 2018 Feb;101:46-57 [FREE Full text] [doi: 10.1016/j.brat.2017.09.014] [Medline: 29066077]
29. Lee EJ, Kim YW. Effects of infographics on news elaboration, acquisition, and evaluation: prior knowledge and issue involvement as moderators. New Media Soc 2016 Jul 09;18(8):1579-1598. [doi: 10.1177/1461444814567982]
30. Fishbein M, Yzer MC. Using theory to design effective health behavior interventions. Commun Theory 2003 May;13(2):164-183. [doi: 10.1111/j.1468-2885.2003.tb00287.x]
31. Neimeyer GJ, Guy J, Metzler A. Changing attitudes regarding the treatment of disordered eating: an application of the Elaboration Likelihood Model. J Soc Clin Psychol 1989 Mar;8(1):70-86. [doi: 10.1521/jscp.1989.8.1.70]
32. Stephenson MT, Benoit WL, Tschida DA. Testing the mediating role of cognitive responses in the elaboration likelihood model. Commun Stud 2001 Dec;52(4):324-337. [doi: 10.1080/10510970109388567]
33. Durkin S, Wakefield M, Spittal M. Looking for boomerang effects: a pre-post experimental study of the effects of exposure of youth to television advertising for nicotine replacement therapy and Zyban. Addict Behav 2006 Dec;31(12):2158-2168. [doi: 10.1016/j.addbeh.2006.02.020] [Medline: 16567056]
34. Campbell AN, Turrigiano E, Moore M, Miele GM, Rieckmann T, Hu MC, et al. Acceptability of a Web-based community reinforcement approach for substance use disorders with treatment-seeking American Indians/Alaska Natives. Community Ment Health J 2015 May;51(4):393-403 [FREE Full text] [doi: 10.1007/s10597-014-9764-1] [Medline: 25022913]
35. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
36. Zickuhr K, Rainie L, Purcell K. Parents, Children, Libraries, and Reading. Part 1: A Profile of Parents. Pew Research Center. 2013 May 1. URL: https://www.pewresearch.org/internet/2013/05/01/part-1-a-profile-of-parents/ [accessed 2022-09-30]
37. Current Population Survey 2018 Annual Social and Economic Supplement. United States Census Bureau. 2018. URL: https://www.census.gov/data/datasets/time-series/demo/cps/cps-asec.html [accessed 2022-09-30]
38. Highhouse S. Designing experiments that generalize. Organ Res Methods 2007 Jul 23;12(3):554-566. [doi: 10.1177/1094428107300396]

XSL•FO
RenderX

39. Coppock A. Generalizing from survey experiments conducted on Mechanical Turk: a replication approach. Political Sci Res Methods 2019;7(3):613-628. [doi: 10.1017/psrm.2018.10]
40. Mullinix KJ, Leeper TJ, Druckman JN, Freese J. The generalizability of survey experiments. J Exp Polit Sci 2015;2(2):109-138. [doi: 10.1017/xps.2015.19]
41. Lin H, Werner KM, Inzlicht M. Promises and perils of experimentation: the mutual-internal-validity problem. Perspect Psychol Sci 2021 Jul;16(4):854-863. [doi: 10.1177/1745691620974773] [Medline: 33593177]
42. Kim SJ, Marsch LA, Brunette MF, Dallery J. Harnessing Facebook for smoking reduction and cessation interventions: Facebook user engagement and social support predict smoking reduction. J Med Internet Res 2017 May 23;19(5):e168 [FREE Full text] [doi: 10.2196/jmir.6681] [Medline: 28536096]

Abbreviations

CDC: Centers for Disease Control and Prevention
CFI: comparative fit index
HPV: human papillomavirus
MTurk: Mechanical Turk
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual

Edited by A Mavragani; submitted 25.02.22; peer-reviewed by J Park, A Baldwin; comments to author 23.08.22; revised version received 02.10.22; accepted 19.10.22; published 24.11.22.

Please cite as:
Kim SJ, Schiffelbein JE, Imset I, Olson AL
Countering Antivax Misinformation via Social Media: Message-Testing Randomized Experiment for Human Papillomavirus Vaccination Uptake
J Med Internet Res 2022;24(11):e37559
URL: https://www.jmir.org/2022/11/e37559
doi:10.2196/37559
PMID:36422887

©Sunny Jung Kim, Jenna E Schiffelbein, Inger Imset, Ardis L Olson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Examining Chinese Users' Feedback Comments on HIV Self-testing Kits From e-Commerce Platforms: Thematic and Content Analysis

Miao Liu[1*], PhD; Yi Zhu[1*], PhD; Haiqing Gao[1]; Jialing Li[1], BA

School of Journalism and Communication, Beijing Normal University, Beijing, China

*these authors contributed equally

Corresponding Author:
Yi Zhu, PhD
School of Journalism and Communication
Beijing Normal University
Room 504B
Haidian District
Beijing, 100875
China
Phone: 86 18518009255
Email: zhuyi5@bnu.edu.cn

Abstract

Background: HIV self-testing is preferred by many Chinese people for its convenience and confidentiality. However, most studies on HIV self-testing (HIVST) uptake in China overfocused on men who have sex with men and overrelied on obtrusive methods such as surveys and interviews to collect data.

Objective: We aimed to explore Chinese HIVST kit users' authentic experience via their feedback comments posted on e-commerce platforms using an unobtrusive approach.

Methods: In total, 21,018 feedback comments about buying and using HIVST kits posted on Chinese e-commerce platforms (Tmall and Pinduoduo) were collected. An inductive thematic analysis based on a random sample of 367 comments yielded several thematic features. These thematic features were developed into coding categories for a quantitative content analysis of another random sample of 1857 comments.

Results: Four themes were identified in the first study, including the expression of positive and negative emotions after and before getting the test, respectively, calling for living a clean and moral life in the future, comments on the sellers and HIVST kits, and the reasons for buying HIVST kits. The results from the second study suggested that there were significant associations between different platforms and several thematic features. Nearly 50% of the comments were related to the product itself and the disclosures of HIV-negative test results. More than 25% of the comments showed users' feelings of gratefulness after receiving negative test results such as "thank heavens for sparing my life."

Conclusions: The results suggested that Chinese users relied on HIVST kits to reduce and prevent HIV infection, while they also considered HIV infection a punishment related to moral violation such as being sexually promiscuous. The traditional Chinese health belief that health is influenced by one's morality still persists among some Chinese users. Many users also lacked appropriate knowledge about HIV transmission and self-testing kits.

(J Med Internet Res 2022;24(11):e38398) doi:10.2196/38398

KEYWORDS

HIV self-testing; Chinese; feedback comments; e-commerce platforms

XSL•FO
RenderX

Introduction

Background

In recent years, there has been an upsurge of HIV infection rates in China, particularly among young people [1]. Early detection of HIV is of vital importance to AIDS intervention and treatment. Among different HIV testing options, HIV self-testing (HIVST), also known as home testing, is more preferred by young people for its convenience, confidentiality, control over disclosing the status to others [2], and social acceptability [3]. HIVST is a process whereby individuals collect their specimens such as oral fluid and blood to perform and interpret HIV rapid diagnostic tests themselves in private [4,5]. This procedure was recommended by the World Health Organization as an option to expand HIV testing coverage [6]. A recent study has shown good consistency between HIVST kit results and HIV antibody detection results [7]. There is a potential of using HIVST to reach the population at high risk such as sexually active young adults who preferred HIVST kits to traditional clinical tests for the accessibility and speed of results of HIVST kits [8].

Many pilot studies showed that HIVST received high acceptability for populations at risk of HIV infection in multiple countries such as Kenya [9], Canada [10], and the United States [11] because it is easy to use and could address stigma-related barriers to testing [9,12,13]. However, HIVST may result in risks in some resource-limited settings and bring about concerns with regard to policies, ethics, privacy issues, users' safety, and public health issues such as the lack of linkage to care or failure to detect accurate HIV infection [4-6]. Although HIVST has its drawbacks, it expands HIV testing to a wider population, provides preliminary HIV knowledge [6], and contributes to informed sexual decision-making [4,12].

To lower the increasing HIV infection rate in China, the State Council of China issued Thirteenth Five-Year Plan (2017-22) for HIV and AIDS prevention and control in which the government encourages the expansion of HIVST by exploring [14] the strategies to promote HIVST through selling self-testing kits in pharmacies and sellers on the web. The Chinese Center for Disease Control and Prevention and Chinese STD and AIDS Association have also been promoting HIVST by providing HIVST-related information through different channels [15]. In 2019, the Chinese Center for Disease Control and Prevention published the "Manual of AIDS Self-testing Guidance" [16] to provide at-risk populations with informational support.

There are robust HIVST studies examining high-risk populations such as men who have sex with men (MSM) in China [7,17-19]. Han et al [17] found that self-testing among MSM was positively related to being married, having ≥6 male anal sex partners in the past 3 months, and having HIV tested within 12 months. Another study found that older age and marriage could be barriers for HIVST uptake, and higher educational levels and more HIV and AIDS knowledge are positive predictors of HIVST uptake among MSM [7]. In addition, administering HIVST among MSM was also preferred by students and those who have a junior college degree or higher and had anal intercourse with men in the past 6 months [20].

Objectives

Overall, current HIVST studies in China have been overemphasizing MSM while ignoring a more general population who use HIVST. However, the largest group that uses HIVST kits in China might not be MSM. A recent study examined the profiles of those who purchased HIVST kits on the web and found ≥57% of the web-based buyers were heterosexual [21]. In addition, it is found that web-based buyers of HIVST kits tended to have higher education and upper-middle income and to be unmarried young men [21]. It should be noted that HIVST studies in China have overwhelmingly adopted obtrusive methods such as survey [7,17-19,21] and interview [12,22] in which those who took HIVST were approached by researchers or they were aware of their status as study participants. Because of HIV stigma in Chinese society [23], participants who completed the survey or got interviewed may not disclose truthful information about their self-testing experiences. In addition, due to the limitations of survey and interview methods, those studies discussed earlier have seldomly examined emotional responses and other sociopsychological factors for those who took HIVST. Taken together, we aimed to adopt an unobtrusive approach [24] to explore HIVST kit users' authentic experiences of administering self-testing based on the web-based comments posted under HIVST kits sold on e-commerce platforms. This approach of examining user-generated health-related information is referred to as infodemiology, which helps to understand user knowledge, attitudes, behaviors, and information consumption [25-27]. Infodemiological approaches examined health communication and information distributed on the internet such as characteristics of health websites or consumers' health informatics [25,28]. Infodemiological approaches have been used widely in health areas [29] to inform public health and policy and obtain user-generated opinions. Users' feedback comments on health-related devices were used as textual data in infodemiological studies on Amazon customer reviews of direct-to-consumer hearing devices [28] and penile clamps [30]. However, there are few infodemiological studies on feedback comments for using HIVST kits in China. It is important to explore how user-generated comments on Chinese e-commerce websites reflect HIVST users' emotional reactions, the reasons for buying HIVST kits, and concerns with their health status. More importantly, those comments are generated by a more general population in China instead of only MSM. As a result, this infodemiological study expanded on the current HIVST studies in China and provides a broader picture of Chinese HIVST kit users by examining their authentic user experience based on their feedback comments.

Methods

Overview

This research included 2 studies examining HIVST kit feedback comments on e-commerce platforms. The first study was an inductive thematic analysis to identify patterns from qualitative data to generate unanticipated insights [31]. Owing to the limited number of studies on the user experiences of HIVST kits, there is little evidence that can be used to derive specific research questions, so inductive thematic analysis was adopted to provide

XSL•FO
RenderX

a preliminary picture of HIVST kit users in China. Themes identified from the comments on HIVST kits in the first study were discussed and then served as the coding categories for the second study, a quantitative content analysis. Quantitative content analysis is frequently used to analyze media messages to explore the patterns in message contents [32]. In the second study, we aimed to explore more specific questions raised from the first study based on a larger sample of comments to understand HIVST kit users' profiles and their responses after purchasing HIVST kits.

Data Selection and Collection

Feedback comments of HIVST kits from Pinduoduo and Tmall were selected for analyzing the HIVST users' experiences. Tmall and Pinduoduo were chosen because both platforms have a huge number of users in China, but they have a different target market. Tmall is an e-commerce platform for top-quality branded products that guarantees all goods on the platform are authentic [33] and had 500 million users by January 2021 [34]. Pinduoduo is a rapidly growing e-commerce platform and had 867 million active buyers in 2021 [35]. Most Pinduoduo users are price-sensitive customers who live in lower-tier cities in China [35].

Data were collected in early December 2021. We searched the keyword "hiv test" in the platforms and identified the top 10–selling HIVST kits at Tmall mobile app and top 15–selling HIVST kits on the Pinduoduo mobile app based on their sales volume within the past month of data collection. The rationale to choose the top-selling products is that, usually, best-selling products generate most feedback comments. There were relatively fewer comments for each HIVST kit posted on Pinduoduo than Tmall, so more HIVST kits sold on Pinduoduo were selected to obtain enough comments. The HIVST kits sold on Tmall or Pinduoduo are either blood based or saliva based. All the HIVST kits discussed in this paper have passed China's National Medical Device Registration Approval.

Feedback comments are defined as information including texts, signs, and emoticons posted publicly on Tmall and Pinduoduo's comment section by HIVST kit users in this study. There were several comments such as "This user thinks this product is really good and gives a five-star rating" or "This user does not leave any comment," which were automatically generated by the platforms while data were retrieved. These comments were excluded from the data corpus because they yielded little implication for our study and were not made by real commenters. Duplicated comments for a particular kit from the same commenter posted during the same period were also removed from this analysis because they might be fake reviews. The final data corpus pool included 21,018 comments (Pinduoduo: n=4989; Tmall: n=16,029).

Ethical Considerations

All the comments retrieved from Pinduoduo and Tmall were posted publicly on the comment sections and were accessible to everyone visiting these 2 platforms. All personally identifiable information (such as commenters' usernames in these platforms and names of locations mentioned in comments) included in the comments was removed from the analysis to protect commenters' anonymity and privacy.

Results

Study 1: Thematic Analysis

Overview

A sample of 192 comments were randomly selected from the overall data corpus for thematic analysis. We followed the procedure of thematic analysis as suggested by Braun and Clarke [31] to generate initial codes, sort those codes into themes, review themes, and define themes. We then read another sample of 175 comments from the overall data to ensure no new themes appear [36].

Four major themes were identified from 367 comments posted on Pinduoduo (n=81) and Tmall (n=286), including the expression of positive and negative emotions, calling for living a clean and moral life, comments on the sellers and HIVST kits, and the reasons for buying HIVST kits. These themes are not mutually exclusive, so a single comment may include all 4 themes.

Expression of Positive and Negative Emotions

Under this theme, HIVST kit users expressed their positive emotions specifically focusing on gratefulness and relief after getting their negative test results, their negative emotions before getting the test, and their hope for others to get negative test results and stay healthy. In most cases, commenters tended to express their gratefulness to the heavens such as "Thank heavens and I will be definitely more cautious about my sexual activities," while several commenters express their gratefulness to many targets. For example, 1 user commented as follows:

> *At last, I will thank that scumbag guy for not killing me. Thank heavens for sparing my life. Thank heavens for not killing me. Thank [the brand of HIVST kits].*

In addition, comments expressing positive emotions such as relief were observed frequently:

> *Thank heavens for not killing me. Now I finally feel relief.*

Some commenters expressed negative emotions before the test, such as fear and guilt, and positive emotions (gratefulness and relief) after the test, in a single comment, for example:

> *I will never forget the feeling at the day while waiting for shipping [of the HIVST kits] that I would rather be dead. Once recalling this, I was trembling all the time. Luckily, the result turned to be good. Take 20 minutes [to conduct HIVST] to get peace for the eternal life. Thank heavens. If you are afraid of [getting HIV]. Don't panic. Don's trust information on the Internet which is exaggerated. Please get [HIV] test.*
>
> *I am looking at the reagent paper and thinking about my life for a while. The result is negative and I feel I get a rebirth inside. Recalling that experience in a few months ago, it might be the most regretful thing I did in my whole life. I would never think that I was*

XSL•FO
RenderX

associated with "AIDS" a few months ago. AIDS had deeply influenced my life in the past few months, I was bothered by different kinds of fear and worries inside. [AIDS] just surrounded me entirely and almost suffocated me.

I have been scared for a few months. It's like I had been living in a hell in the past few months. After the test today, I finally feel relief. I will live a clean and moral life in the future and not feel worried any longer.

In front of [sexual] temptation, you just could not make any reasonable decision. Previously I committed such a sin and I was so frightened. I am too young and have no children. If I really got shot [to get AIDS], I deserved it. Luckily, thanks for heaven's mercy to spare my life. Since this event, I feel so grateful and I will be a good person.

In addition to gratefulness and relief, HIVST kit users also expressed their hope for both themselves and other HIVST users to receive negative results. Two users commented as follows:

Thank heavens for not killing me. I hope the test result is accurate. I hope everyone who purchases this [HIVST kits] gets a negative test result!!!

If you found yourself getting infected earlier, you can get yourself treated earlier. I hope all of you just have a false alarm! Hope all of you [have] negative [test results]!

Calling for Living a Clean and Moral Life in the Future

This theme features calling for living a clean and moral life (*Jieshenzihao* in original Chinese comments) in the future with greater self-control for users themselves and other HIVST kit users. Commenters made promises to resist promiscuous lifestyles and made a determination to live a self-disciplined life. The following are a few examples:

I got a box of oral fluid test kits and a box of blood test kits...both test results are negative. I hope the test results are accurate. I will live a clean and moral life so diseases [HIV] will not attack me. I posted this comment here as a reminder to myself! And I shared this (comment) with everyone!

AIDS-phobia is so scary. I was so scared while waiting for the test result. I was too young to know how harmful it is when I had these risky acts. I could not fall into sleep these days [because I was] imagining how I should deal with my family members if I really got AIDS. Thank heavens for not killing me. If I was going out for some promiscuous acts in the future, may thunders strike me down.

Other HIVST kit users shared their experiences as a warning to raise HIV awareness and to call for others to resist promiscuous lifestyles:

I am warning everyone here. Never live promiscuously. You must live a clean and moral life and be careful about your health. I hope everyone gets a bright future, and I will start to live my new life.

After this experience, I have learned the importance of living a clean and moral life and will no longer fool around my own life...I hope everyone can live a clean and moral life and don't fool around with your own life so easily.

Comments on the Sellers and HIVST Kits

Under this theme, HIVST kit users provided either positive or negative evaluations for web-based sellers and HIVST kits. As for their web-based shopping experiences with sellers on these e-commerce platforms, most comments were concerned with privacy and convenience about purchasing HIVST kits on the web, which is consistent with the previous findings on the advantages of using HIVST for the sake of privacy and convenience [18,22,37]. For example, some commenters made a comparison between using HIVST kits and visiting the clinic for a test and showed their preference for the former:

I feel that a heavy stone was finally removed from my heart. This reagent paper was amazing to protect me from awkwardness for visiting a clinic.

The price is really reasonable and much cheaper than visiting a clinic. The [shipping] speed is really fast. I bought it today and it will arrive tomorrow. Shipping service is really good.

Cheap and useful. Accuracy is as high as testing at a hospital.

An interesting observation from these comments is that few HIVST kit users complained about the cost of buying kits. Unlike previous findings based on MSM samples in which high cost is one of the top barriers for using HIVST [18,22], some users in this study even considered HIVST kits sold on e-commerce platforms to be very cheap. One explanation is that web-based buyers of HIVST kits tended to have upper-middle income. This may result from the demographic differences between the general population examined in this study and MSM samples in previous studies.

There are comments addressing how patient the customer service representatives were. Some buyers posted positive comments about sellers for fast delivery and discreet shipping to protect their privacy. Because of HIV stigma in China [38], those who purchased HIVST kits on the web were worried about their privacy, so discreet shipping plays an important role for those buyers to assess sellers; for example:

Very nice. The customer service representative is very patient. I finally feel relief after this test. One line [appeared on the reagent paper as a negative result] and there is not any [other] problem.

Shipping speed is fast. Operation is easy and [test result is] accurate. The package was delivered in discreet shipping.

This time of torture is finally over. I will definitely love my life after this. [AIDS] is such terrible. The reagent paper is good. The package [of HIVST kits]

XSL•FO
RenderX

was handled in discreet shipping. This seller is considerate and the speed of delivery is fast.

As for the user experience of administering HIVST, there are several positive comments on the accuracy of the test and the speed of getting the result. The following are several examples:

Reagent paper has good quality. The test result is very accurate. Administering this test at home is so convenient.

Very good to check health status. I have been worried for several days and finally I feel relief to work. It's so frightening. Reagent paper is nice and operation is convenient. I conducted [this test] really quick.

The quality is guaranteed. Safe and effective. The package from this seller is complete. The most important is to get the result fast and conveniently. [I] don't have to worry about this every day!

Previously I used HIVST kits of other brands, but I still felt unsure [about test results]. Then I found this [brand name of HIVST kits] and knew many hospitals and disease control centers all used this [brand]. I also knew their HIVST kits got certified from World Health Organization PQ, which makes me trust the accuracy of test results. Thank [this brand] for producing such excellent kits.

Accurate, formal, safe, and convenient, but the lancet to collect blood samples is too short to use. It took several times to push blood out. Maybe it's my own problem.

Interestingly, few commenters mentioned visiting a clinical center for conventional voluntary counseling and testing (VCT) to ensure their test results are accurate after obtaining negative results from HIVST. There are also few comments questioning the accuracy of the test result, which is not consistent with the previous findings that MSM in China were worried about the accuracy of HIVST [18,22]. One study found that MSM in China recognized HIVST as a supplement rather than a replacement for facility-based testing [12]. It remained a question as to whether the general HIVST kit users in China recognize HIVST as a replacement for facility-based testing methods without questioning the testing accuracy. Similar to the aforementioned quotation where a commenter took the testing accuracy of HIVST kit for granted because of its certificate from the World Health Organization Prequalification, most HIVST kit users did not express their concerns with operation errors and misinterpretation yielding negative test results, not to mention seeking VCT from clinic centers to ensure their test results.

There are also a few negative comments on sellers and user experiences. Some comments complained about the shipping speed, customer service, and the quality of tools inside HIVST kits. For example, one user complained about the shipping speed and finally gave up on waiting for the HIVST kits ordered on the web:

Shipping is a little bit slow. I ordered [HIVST kits] on Monday night…and received it on Thursday afternoon. I could not wait for that and went to a clinic (to test), which cost me 40 Chinese Yuan and 2 hours to get the result.

Another buyer posted a negative comment on a customer service representative who denied the problem caused by the quality of kits:

Thumb down (for this product). [Reagent paper] did not absorb the collected blood sample. [The customer service representative] said this was due to my operation error. I bought HIVST kits from many [other] sellers and I bought kits from this seller several times. It is the first time to have this problem…Please do not buy [HIVST kits] from this seller.

Several buyers commented on the lancet used to collect blood samples, such as biosafety of using such a lancet and how hurtful it would be. Zhang et al [22] reported similar observations in which MSM were worried that needles or lancets in HIVST kits were used by other people carrying HIV. A similar result is observed in this study:

One of the lancets looks like broken. Don't know whether or not it was used by someone else. I was so scared.

Reasons for Buying HIVST Kits

Commenters also provided their reasons about why they decided to purchase and use HIVST kits. The reasons included risky sexual activities, nonsexual transmission, pure AIDS phobia, and preventive purposes. Most of the reasons are related to commenters' own or their sexual (ex-) partners' risky sexual activities, which made them purchase HIVST kits, for example:

Thanks! Heavens gave me a chance for a new life. I met a lady…and after that I got neck and spine pain with slight fever latterly. I saw AIDS prevention campaigns in my community and I found I got similar symptoms. I felt so scared. My brothers and sisters, please live a clean and moral life! I would not behave like that after this lesson. At last, thanks for this HIVST kit to bring me a new life.

It is necessary to test. Nowadays young people do not have only one partner. This disease has a long incubation period so it is necessary to test.

Don't date young adults anymore in the future. It's scary. Thank heavens for sparing my life. I will behave myself and I must use a condom. Safety comes first, my sisters.

I was once got drunk and had sex with someone. I was so scared but don't have the courage [to get test]. Thank heavens for sparing my life. I will not drink alone in the future. I will live a clean and moral life.

There are also reasons focusing on using HIVST because of fear of nonsexual HIV transmission via blood and oral fluids; for example:

I was having a meal with a friend and was kissed suddenly [by this friend]. I felt so scared and decided to get a [HIV] test. I knew he went out for prostitution before. I went to a disease control center for

XSL•FO
RenderX

> *counseling, and the doctor there was stubborn and did not give me any clear suggestions. Then I went to a hospital that specializes in infectious diseases for counseling. The nurse there told me HIV could not be transmitted via kissing, even tongue kissing, and suggested that I seek psychological therapies if I was so fearful.*
>
> *While I was giving an injection to someone who might be a drug addict, [this person's blood sample or oral fluids (not specified in the comment)] dropped on my fingernail. On the second day, it was found out that this [patient] got AIDS. Although there was no open wound on my skin, I was scared to death. I am a pregnant woman and I am so angry. I hope all AIDS patients should not hide their status while visiting a clinic. This is really dangerous for health practitioners. It's harmful for both others and patients themselves.*
>
> *Negative. Thank heavens for sparing me. I went to a small beauty salon to remove pimples a few years ago. I was so frightened since the needle used in this beauty salon might not get disinfected. Listen to me, if you want to do some cosmetic surgeries like ear piercing or eyebrow tattoo, you definitely should go to a formal beauty salon. Otherwise, you might have to worry about getting AIDS.*

Clearly, these commenters know of nonsexual HIV transmission on some level, and they may have legitimate concerns to worry about getting HIV infection via medical tools or blood. However, their self-perceived likelihood of getting infected might be overestimated because of their misunderstanding of HIV transmission. They knew the ways of transmission but did not have a clear picture of the likelihood of getting HIV infection for each type of transmission and other HIV information such as how long HIV will be infectious outside the human body, which finally caused their panic and phobia and the buying of HIVST kits. The contagiousness of HIV was exaggerated, which is consistent with the previous study [39].

There are also a few cases in which HIVST kit users tested themselves because of AIDS phobia elicited by the popular science of HIV or they chose to test for preventive purposes, for instance:

> *I was told getting tooth extraction can even cause AIDS from popular science [articles or videos (not specified in the comment)]. I will get a premarital examination really soon. I was scared so I bought HIVST kits to test myself. I was so nervous before testing. My head was buzzing. I wondered what I should do if I really got [infected by HIV]? I put so much effort into dating a good boyfriend right now and I was afraid that [he] would leave me [because of AIDS]. Later, I realized that I just thought too much… Thank heavens for sparing my life. Although I did not fool around, I am neither homosexual nor a drug addict. However, it is still risky [to get HIV] from wounds if you do not use any bandage. As someone on the Internet said: AIDS-phobia makes me live a discreet life. Same as me.*

Interestingly, this commenter did not engage in any risky behaviors and did not contact anyone who might carry HIV. She chose to take HIVST only because she was frightened by the popular science of HIV. She wanted to ensure her HIV status to prevent any possible embarrassment and loss before her marriage. This commenter even suggested that AIDS phobia helped her live a discreet life.

The aforementioned examples also demonstrated the stereotypes about people who live with HIV or AIDS, such as portraying people who live with HIV or AIDS as homosexual or drug addict. It raises a question about the effectiveness and the side effect of the popular science of HIV and people who live with HIV or AIDS, which may lead to further HIV-related panic, fear, and stigma.

In addition to these 4 main themes, there are some minor themes that did not appear frequently from these comments. For instance, there are a few comments concerning the misinformation and rumors about HIV or AIDS that they searched on the internet. One commenter said: "I will suggest to everyone that they should not check [a Chinese-based search engine] because [information searched on this site] frightened me a lot." Another commenter was so scared by the misinformation on the internet about HIV and was not able to get a test immediately:

> *I found AIDS symptoms on the Internet matched with mine perfectly. I had been panicking for almost a whole year. I was too scared to get a test. I could only get comfort by telling myself that the chance [of getting HIV infection] is so slight via searching information on the Internet. However, it was said that the more you know, the more you fear. If you search [HIV] on [a Chinese-based search engine], you find yourself getting an incurable disease.*

There is also a comment concerning the effect of using HIVST on mental health: "Just a suggestion, make sure you are psychologically prepared before purchasing [HIVST kits]." This is consistent with one of the drawbacks of taking HIVST that HIVST kit users receive little professional and appropriate psychological support and counseling in a nonclinical setting [4,13,22].

As a result, 4 main themes were identified from HIVST kit users' comments. The first theme is concerned with the expression of positive and negative emotions. Many commenters under this theme thanked the heavens for sparing their life and showed their relief after receiving negative test results. Other commenters under this theme expressed hope for everyone else to receive negative test results and stay healthy. In addition, this theme is also characterized by commenters' panic and fear of HIV infection before receiving negative test results and commenters felt extremely uncertain and unsafe about their future life interrupted by potential HIV infection. The second theme is concerned with living a clean and moral life in the future. Commenters under this theme not only expressed their determination to live a careful life but also offered sincere suggestions for others not to be sexually promiscuous around

XSL•FO
RenderX

in the future. The third theme focuses on the evaluative comments on either sellers for their services, shipping speed, discreet shipping, conveniences, and other advantages over visiting clinics or HIVST kits for their quality. The fourth theme focuses on commenters' reasons about why they chose to purchase HIVST kits. Most of the reasons reflected commenters' fear of HIV infection by sexual transmission. Other reasons focused on nonsexual HIV transmission via blood or unclean medical tools. The implications for these findings are explained in the Discussion section after the results of the content analysis.

The first thematic analysis study explored some features of HIVST kit users' comments posted on e-commerce platforms. However, owing to the inductive nature of the thematic analysis, this study only provides a basic picture of users' concerns with a limited data sample. More systematic observations are needed to provide quantitative data about the frequency of each thematic feature based on a larger sample of comments. In addition, the sample yielding these 4 main themes was based on the mixed comments from 2 different e-commerce platforms, so the possible differences in these themes between platforms were not observed in this study. We will address these aforementioned questions in the second study of the quantitative content analysis based on 1857 comments on HIVST kits.

Study 2: Quantitative Content Analysis

Overview

The second study provides quantitative data about frequencies of different thematic features of comments based on a larger sample of comments. In addition, we aimed to explore any potential differences regarding HIVST kit users' comments between different e-commerce platforms. We conducted a quantitative content analysis of 1857 HIVST kits users' comments posted on Tmall (n=975) and Pinduoduo (n=882). These 1857 comments were randomly selected from the entire data pool including 21,018 comments retrieved in December 2021. Regarding the random selection process for the content analysis, a Microsoft Excel add-in named Fangfanggezi [40] that enables random sampling within Excel was used. Quantitative content analysis was adopted for this study because it is widely accepted for studying buyers' feedback comments [41,42] to understand patterns of a large-sized sample of messages.

Codebook Development

The codebook was developed based on the main themes identified in the first study. Different coding categories were elaborated upon based on thematic features under each theme including gratefulness, positive emotions (eg, relief and happiness) other than gratefulness, negative emotions, hope for negative test results and good health, calling for living a clean and moral life in the future, comments for sellers, comments for HIVST kits, and the reasons for buying HIVST kits. In addition to these coding categories developed from thematic features mentioned in the first study, there is one newly added coding category, users' disclosures of their self-testing results, which emerged inductively from coders' training processes. A single coding unit could be coded for multiple thematic features mentioned in the codebook. Details of the coding categories are presented in Table 1.

XSL•FO
RenderX

Table 1. Coding categories for quantitative content analysis.[a]

Category (thematic features)	Description of each level within the category[b]	Examples[c,d]
Gratefulness	• 1=commenters expressed gratefulness after receiving negative test results	• "Thank heavens for sparing my life." (Coded as 1) • "Thanks. Heavens give me this opportunity. I will live a sexually abstinent life in the future. Being negative forever." (Coded as 1)
Positive emotions other than gratefulness	• 1=commenters' positive emotions (eg, relief and happiness) except from gratefulness	• "I used it but I am not sure about the accuracy of the test result. Anyway, I feel much better now." (Coded as 1) • "Feel relief." (Coded as 1)
Negative emotions	• 1=commenters' negative emotions such as fear and guilt	• "I heard about AIDS every and it scared me to death. In the future, I must live a sexually abstinent life." (Coded as 1) • "Finally, I don't need to have my heart in my mouth any longer. The food today is so delicious." (Coded as 1)
Hope for negative test results and good health	• 1=commenters expressed their hope for everyone receiving negative test results and staying healthy	• "Wish everyone [getting] negative [test results]..." (Coded as 1) • "Hope being negative forever. Hope everyone to stay healthy!" (Coded as 1)
Calling for living a clean and moral life in the future[e]	• 1=commenters called themselves to live a clean and moral life • 2=commenters called others to live a clean and moral life • 3=commenters called both themselves and others to live a clean and moral life • 4=commenters encouraged others to get HIV tests but did not mention living a clean and moral life • 5=keep on fooling around	• "Thanks. [Hope] negative always. Keep living a sexually abstinent life. [Live] healthy and happily." (Coded as 1) • "Dear strangers. Promise me. Love yourself." (Coded as 2) • "Live a sexually abstinent life, bros." (Coded as 2) • "Thank heavens for giving me an opportunity of the rebirth. I must live a sexually abstinent life in the future. And hope other bros to live a sexually abstinent life." (Coded as 3) • "I can continue to fool around..." (Coded as 5)
Comments for sellers	• 1=positive evaluation for sellers (regarding convenience, shipping speed, discreet shipping, price, customer service, etc) • 2=Negative evaluation for sellers	• "It is the discreet shipping. Shipping speed is fast..." (Coded as 1) • "First time I bought two kits...now I finally got off from this fear. Thank the seller for their patient explanations." (Coded as 1) • "Customer service representative No. XXX is very irresponsible!!!...Good attitude before the purchase but attitude changed after the purchase!!!...When the customer had a problem which needed her explanation, she was not patient! Being distracted!!!..." (Coded as 2)
Comments for kits	• 1=positive evaluation for kits (regarding testing accuracy, biosafety, and easiness of operation and interpretation, etc) • 2=Negative evaluation for kits	• "Safe and convenient. Good attitudes for customer service." (Coded as 1) • "The reagent paper is very good and convenient. Thank the boss." (Coded as 1) • "The lancet is broken and could not be used. Please don't buy it everyone." (Coded as 2) • "Useless. I doubt if this thing can test (the disease) at all." (Coded as 2) • "Garbage. The reagent paper is useless..." (Coded as 2)
Reasons	• 1=commenters' uptake of HIVST[f] because of pure phobia • 2=commenters' uptake of HIVST because of their or their (ex)partners' risky sexual activities • 3=commenters' uptake of HIVST because of nonsexual transmission • 4=commenters' uptake of HIVST because of feeling bored or curious • 5=commenters' uptake of HIVST for preventive purposes	• "I must be crazy. I did nothing. Once the teacher mentioned this in class, I felt panic. I bought the kit and it's nothing wrong. I just want to feel relief. In the future, I shall no longer look at these websites about hiv..." (Coded as 1) • "I had two tests and both were negative. In the future, I must stay away from the scumbag guy. Thank the seller." (Coded as 2) • "I have tattoo. It was said the tattoo needles are not clean. I was scared so I bought the kit to test as soon as I can...It is best that you don't get tattoo or teeth extraction if there isn't any problem. It saves your time for worrying about yourself." (Coded as 3) • "Bought this for my curiosity." (Coded as 4) • "It is necessary for adults to take the test regularly. Very good, easy to use, and convenient." (Coded as 5)

XSL•FO
RenderX

Category (thematic features)	Description of each level within the category[b]	Examples[c,d]
Disclosures of self-testing results	• 1=there is enough evidence to infer that the user received a negative test result (eg, Thank heavens for sparing my life!) • 2=there is enough evidence to infer that the user received a positive test result • 3=commenters felt unsure about results	• "Being worried for a few days. I received [the kit] and get self-testing. There is not any problem at all. I can feel relief now." (Coded as 1) • "Positive results twice..." (Coded as 2) • "I am not sure whether the test result is accurate. The operation is convenient." (Coded as 3)

[a]All coding categories are not mutually exclusive. Krippendorff α values for all categories ranged from .79 to 1.00.

[b]0=absence of a particular category.

[c]We corrected typos and grammar errors.

[d]Examples were selected from 1857 comments for the content analysis.

[e]There are several newly added levels within a category such as levels 3, 4, and 5 in the category of calling for action. They were included during the training process as suggested by 3 coders.

[f]HIVST: HIV self-testing.

Coder Training and Procedures of Analysis

Three coders were trained using the codebook to code 50 comments to test the codebook's clarity and ensure their understanding of it. Then, each of the 3 coders were provided the same set of 50 new comments to code independently. After they completed coding, intercoder reliabilities between them were calculated, and the coders were brought together to discuss and reconcile their discrepancies in terms of coding. Their feedback was also used to modify the coding categories in the codebook. This procedure of independent coding was repeated thrice. One new coding category of disclosing HIVST results mentioned earlier was included in the original codebook during these 3 trials. The training finally yielded accepted intercoder reliability (Krippendorff α values ranged from .79 to 1.00 for all 9 coding categories). Then, each of the coders was provided a different set of about 600 new comments to code independently, yielding 1857 comments in total for the quantitative content analysis.

Frequencies of Thematic Features in Total

The frequencies of different thematic features in total and in each platform are displayed in Table 2. The top-5 predominant thematic features in total (N=1857) included comments for HIVST kits (906/1857, 48.79%), disclosures of self-testing results (869/1857, 46.79%), comments for sellers (593/1857, 31.93%), gratefulness (493/1857, 26.55%), and calling for living a clean and moral life (341/1857, 18.36%). It is not surprising to find that the predominant thematic features observed are related to evaluations for HIVST kits and sellers since the data were based on feedback comments on e-commerce platforms. Surprisingly, there are many comments disclosing their self-testing results. HIVST kit users may consider e-commerce platforms as an anonymous stage to express their concerns and discuss their test results without worrying about HIV stigma. However, only a few commenters disclosed their positive test results (6/1857, 0.32%). Most commenters reported negative test results (833/1857, 44.86%). It is questionable whether those receiving positive test results may feel more reluctant to disclose their status in their feedback comments because of HIV stigma. On the other hand, even though most commenters received negative test results, their acts of disclosing the test results and (more importantly) the fact that they conducted HIVST implied their HIV awareness, self-efficacy, and courage to counter HIV stigma instead of regarding the taking of HIV tests as a controversial and stigmatized topic. In addition, regarding the comments for sellers and kits, there were more positive evaluations than negative evaluations, which suggested that, at least for HIVST kit users, most kits sold on the web met their needs.

XSL•FO
RenderX

Table 2. Frequencies for thematic features in total and each platform (N=1857; Tmall: n=975; Pinduoduo: n=882).

Thematic features	Total, n (%)	Tmall, n (%)	Pinduoduo, n (%)
Gratefulness[a]	493 (26.5)	206 (21.1)	287 (32.5)
Positive emotions other than gratefulness	184 (9.9)	97 (9.9)	87 (9.9)
Negative emotions	251 (13.5)	134 (13.7)	117 (13.3)
Hope for negative test results and good health	71 (3.8)	38 (3.9)	33 (3.7)
Calling for living a clean and moral life in the future[b]	341 (18.3)	139 (14.2)	202 (22.9)
Calling themselves to live a clean and moral life	310 (16.7)	124 (12.7)	186 (21.1)
Calling others to live a clean and moral life	19 (1)	11 (1.1)	8 (0.9)
Calling themselves and others to live a clean and moral life	10 (0.5)	2 (0.2)	8 (0.9)
Keep on fooling around in the future	2 (0.1)	2 (0.2)	0 (0)
Comments for sellers[a]	593 (31.9)	291 (29.9)	302 (34.3)
Positive evaluation	562 (30.2)	266 (27.3)	296 (33.6)
Negative evaluation	31 (1.7)	25 (2.6)	6 (0.7)
Comments for HIVST[c] kits[a]	906 (48.8)	470 (48.2)	436 (49.4)
Positive evaluation	856 (46.1)	430 (44.1)	426 (48.3)
Negative evaluation	50 (2.7)	40 (4.1)	10 (1.1)
Reasons	290 (15.7)	156 (16.1)	134 (15.1)
Pure phobia	133 (7.2)	78 (8)	55 (6.2)
Risky sexual activities	85 (4.6)	30 (3.1)	55 (6.2)
Nonsexual transmission	29 (1.6)	21 (2.2)	8 (0.9)
Feeling bored or curious	9 (0.5)	6 (0.6)	3 (0.3)
Preventive purpose	34 (1.8)	21 (2.2)	13 (1.5)
Disclosures of self-testing results	869 (46.8)	417 (42.8)	452 (51.2)
Negative	833 (44.9)	393 (40.3)	440 (49.9)
Positive	6 (0.3)	4 (0.4)	2 (0.2)
Unsure about results	30 (1.6)	20 (2.1)	10 (1.1)

[a]A significant association between this thematic feature and platforms (P<.001).

[b]There is level 4 under the category of "calling for living a clean and moral life" in which commenters encouraged others to get HIV tests but did not mention living a sexually abstinent life (Table 1). This level was added during the training process based on the training data, but there were not any comments in the main study which can be coded for this level. Hence, it was removed from Table 2.

[c]HIVST: HIV self-testing.

Differences Between Tmall and Pinduoduo

The relationship between each one of the thematic features and e-commerce platforms were examined by using Chi-square tests to explore possible platform differences. Some significant results were observed. The association between the thematic feature of gratefulness and platforms was significant (N=1857, χ^2_1=30.9, P<.001). The association between comments for sellers and platforms was significant (N=1857, χ^2_2=17.2, P<.001). The association between comments for HIVST kits and platforms was also significant (N=1857, χ^2_2=17.1, P<.001). Table 2 presents details on the platform differences.

Other Findings

Chi-square tests were also conducted to examine the relationships between different types of thematic features, and some significant relationships were observed. Gratefulness was associated with multiple other themes: positive emotions other than gratefulness (N=1857, χ^2_1=90.7, P<.001); negative emotions (N=1857, χ^2_1=67.3, P<.001); comments for seller (N=1857, χ^2_2=53.8, P<.001); comments for HIVST kits (N=1857, χ^2_1=202.6, P<.001). In addition, the theme of negative emotions was associated with positive emotions other than gratefulness (N=1857, χ^2_1=25.3, P<.001); the theme of negative emotions was also associated with comments for HIVST kits (N=1857, χ^2_1=76.7, P<.001).

XSL•FO
RenderX

The content analysis of the comments on HIVST kits from the 2 popular e-commerce platforms in China yielded significant differences in thematic features across the platforms. Some significant relationships between the themes themselves were also observed. The frequencies for thematic features also provided a clear picture of predominant features observed in these comments. The implications of these findings are explained in the Discussion section.

Discussion

Principal Findings

On the basis of the infodemiological approach, both studies yielded important findings on Chinese HIVST kit users' concerns and health awareness across 2 e-commerce platforms. To the best of our knowledge, this is the first study examining the feedback comments of web-based HIVST kit users. The qualitative thematic analysis identified 4 themes featuring users' emotional responses before and after self-testing, behavioral intentions (eg, living a clean and moral life), evaluations for sellers and kits, and reasons for purchase. The quantitative content analysis provided the frequencies of different thematic features, explored platform differences of these features, and examined the relationships between thematic features. The implications of both studies are discussed in subsequent sections.

Thematic Analysis

On the basis of the results of the first thematic analysis, many commenters expressed their gratefulness and relief for receiving negative results, but few positive-result comments were observed. This theme features HIVST kits users thanking the heavens for sparing their life and giving them rebirth. For these commenters, getting an HIV infection depends on the heavens instead of themselves. This result also implies HIV fatalism, which is the belief that HIV acquisition and mortality is beyond one's own control [43]. Such a fatalistic attitude might become an obstacle for health promotion and HIV prevention campaigns since HIV fatalism is also associated with risky sexual activities [44,45].

Interestingly, few comments addressed the importance of using condoms before engaging in risky sexual activities, but many comments emphasized the importance of living a clean and moral life in the future (theme 2). This calling for living a better life suggests HIVST kit users, although being influenced by HIV fatalism on some level, still believe that they can control their health in the future by living a moral life. This contradicting phenomenon might result from Chinese moral beliefs based on Confucianism and Buddhism [46] that misbehaviors bring about punishments from the heavens and correcting misbehaviors bring about the removal of punishments from the heavens. Confucianists believe that one's destiny is determined by their moral efforts and that negative outcomes are the consequences of moral failures; Buddhists, on the other hand, believe that good acts will earn positive outcomes and bad acts will lead to negative repercussions [46]. Chinese virtues emphasizing sexual purity tend to equalize promiscuous lifestyles, infidelity, and prostitution as a sin and moral failure (see the study by Zhai [47] for details on Chinese virtues about sexual purity). Getting an HIV infection via risky sexual activities is the sign of punishment from the heavens for committing a sex-related sin. Therefore, those who expressed thanks to the heavens may consider themselves sinful or as failing morally for their debauched and promiscuous lifestyles, so their HIV status depends on the heavens. If they promise to the heavens to live a clean and moral life in the future, then the punishment of an HIV infection is preventable or at least controllable due to their moral efforts. Future studies should explore how to emphasize self-efficacy to prevent HIV instead of equalizing HIV or AIDS as a sin or punishment due to moral failures.

In addition, few commenters mentioned seeking VCT to ensure their test results are correct, and there were few doubts about the accuracy of testing results. It is plausible that commenters felt no need to confirm since they had received negative results. However, there is also a risk of misoperation or misinterpretation of results, not to mention the poor quality of uncertified kits or kits damaged during shipping that may also result in accurate results. Although previous studies suggested that the consistency between HIVST results and antibody detection results was 90.5% [7], and most participants could perform the self-test correctly and obtain accurate test results at home [48], there is always a risk of a false negative. Many HIVST users did not receive appropriate education on the limitations of HIVST kits. The relief HIVST kit users expressed in the first theme might just be seeming relief without confirmation from formal VCT centers. Some comments also showed a misunderstanding of HIV transmission. The results suggested that more HIV and AIDS education campaigns are needed.

Content Analysis

The results of content analysis showed that a large number of commenters expressed their gratefulness suggesting these commenters considered that whether they get HIV is not fully controllable but dependent on some spiritual factors outside of their control. Therefore, there is a need to understand the underlying reason why these people like to take their chances rather than refrain from risky activities. The examination of the reasons for buying HIVST kits revealed that the motivations behind HIVST self-testing are mostly pure AIDS phobia or related to risky sexual activities. Many commenters indicated that they suspected they might get infected after watching an AIDS-related video on a short video app. On the one hand, such media-induced AIDS phobia is an effective way to increase HIVST uptake in the general population. On the other hand, media-induced AIDS phobia could cause anxiety and worry, leading to long-term emotional distress. In addition, the percentage of pure phobia reasons is higher than risky sexual activities reasons. One possible explanation is that many Chinese buyers feel uncomfortable sharing their sexual experiences on the web because people do not usually talk about sex publicly in traditional Chinese culture [47]. It is reasonable to expect that there were many more people who took HIVST because of their previous engagement in risky sexual activities as there is a high percentage of comments calling themselves to live an abstinent life. Apart from pure phobia and risky sexual activities, there are other reasons such as preventive purposes and fear of nonsexual transmission.

XSL•FO
RenderX

Nearly 50% of the comments were related to the product itself and 32% were related to the seller. It suggested that most HIVST kit users were concerned with the quality of the product and services that the seller provided. Interestingly, it was observed that many users commented that "the test result is very accurate," although they might not have the result from the hospital to compare with. Such a comment reflected the users' sincere hope for a negative result. There were more positive comments (n=562) than negative comments (n=31) for sellers. Similarly, for HIVST kits, more positive comments (n=856) were observed in the data than negative comments (n=50). It is plausible that Chinese users in our data sample felt reluctant to report a negative user experience or that some positive feedback reviews were fabricated. Some firms pay people to write a fake review using fake identities on the web [49], so the numbers and frequencies of positive comments on sellers and kits might be overly estimated.

There were some significant differences between Tmall and Pinduoduo data. For example, the percentage of expressing gratefulness in comments is higher for Pinduoduo than Tmall. One explanation is that there are user differences. The user portrait of Pinduoduo is notably different from Tmall as a large portion of Pinduoduo users are older adults looking for affordable products in small cities [50], and most Tmall users are between the ages of 30 and 39 years who come from first-tier cities [51]. In addition, some thematic features were significantly related to other thematic features. For example, gratefulness was related to comments on sellers, comments on HIVST kits, positive emotions other than gratefulness, and negative emotions. Similarly, negative emotions were related to positive emotions other than gratefulness and comments on HIVST kits. This may suggest that expressing feelings of gratefulness and negative emotions might be an indicator for expressing other types of emotions and feedbacks. The findings might also imply that the HIVST kit users' emotional status was related to their evaluation on products. Gratefulness felt by these users, as a positive emotion, was related to users' evaluation on both sellers and products. Negative emotions, on the other hand, were related to the evaluation on products rather than sellers. Future studies may consider conducting sentiment analysis to understand the contextual meaning of feedback comments of HIVST kits. In addition, future studies might adopt the survey method to measure HIVST kit users' gratefulness and other emotions, their evaluation of products and sellers, their intention to live a clean and moral life in the future, and their acts of disclosing self-testing results in feedback comments on platforms.

Implications and Future Directions

Theoretically, this study explored Chinese cultural influences on attitudes toward HIV prevention and acquisition. Although comments such as "thank heavens" implied HIV fatalism that HIV acquisition is out of one's own control. Many commenters believed they should live a discreet and virtuous life in the future since the heavens had decided to spare their life. This implied a traditional Chinese health belief in which health is influenced by one's morality [46]. This belief may result in health-related stigma leading to fear of moral contamination and of losing face for people who carry such health stigma [23,52]. People who live with HIV or AIDS are stigmatized as immoral and indecent [39]. Some HIVST kit users in this study carried the same moral judgment toward people who live with HIV or AIDS and HIV self-stigma when suspecting themselves of getting infected by HIV. However, by cultivating morality via calling for living a clean and moral life in the future, HIVST kit users believe that they still have control over their health to prevent HIV. Public health scholars, practitioners, and policy makers should explore how this traditional Chinese health belief relating morality to health still influences contemporary Chinese users' responses to HIV infection and prevention. More specifically, does this health belief elicit HIV stigma or fatalism in China? Or does this health belief help to strengthen people's agency to prevent HIV? Future studies should consider analyzing HIVST kit users' feedback comments regarding their self-stigmatization, concerns with face loss, feelings of shame, and fear of moral contamination threatening to themselves and their family members.

Methodologically, relying on authentic user-generated information from a more general population and the multimethod approach, this paper heralded a new direction for HIVST research in China. Previous HIVST studies in China mostly focused on MSM populations and heavily relied on obtrusive methods such as surveys [18,53] and interviews [12,22]. By analyzing user-generated information from 2 e-commerce platforms, this paper avoided possible bias caused by obtrusive methods such as social desirability bias. In addition, by adopting the multimethod approach and the infodemiological perspective, this study provided a detailed picture of HIVST kit users' authentic concerns, which were rarely addressed in previous HIVST studies in China. It is recommended that future studies should consider adopting machine learning techniques or the Linguistic Inquiry and Word Count-22 to analyze more feedback comments of HIVST kits on e-commerce websites. Techniques detecting fake reviews should be incorporated in a future study to obtain a more accurate estimation of positive evaluations on sellers and HIVST kits. There were few commenters disclosing a positive test result in the data, future studies should consider how to use unobtrusive methods to obtain the information about their responses. For example, HIVST kit users who got a positive test result might seek help from customer service representatives to ensure their results are accurate. Future studies may also consider interviewing the customer service representatives of HIVST kits about their communication with HIVST kit users to know the latter's inquiries and concerns. In addition, little is known about how these HIVST kit users read the manual included in the kit package and how they conducted testing based on feedback comments. These might be addressed by either interviewing customer service representatives or HIVST kit users themselves about the difficulties the users encountered during self-testing.

Practically, this study raised a few concerns with the uptake of HIVST in China. Both studies indicated that only a few Chinese users questioned the quality of HIVST kits based on the feedback comments of top-selling HIVST kits sold on popular e-commerce platforms. In addition, very few commenters mentioned visiting hospitals to confirm their test results. There are still possible errors due to misoperation and misinterpretation while taking HIVST. It is crucial to incorporate counseling and

XSL•FO
RenderX

confirmation services into HIVST business to avoid potential false reassurances caused by HIVST. In addition, the results of this study suggested that HIVST kit users were overwhelmed by misinformation on the web about HIV transmission, infection, and symptoms, which led to some users' AIDS phobia. Future public health campaign designers should consider including brochures or flyers with credible HIV-related knowledge inside HIVST kit packages.

Limitations

This study has its limitations. First, the results might be subject to sampling bias as the buyers whose HIVST results were positive felt reluctant to share their results on the web. Only 0.32% (6/1857) of the commenters shared their positive results in the content analysis. In the same vein, the results of the reasons category may not represent the whole picture as only a small percent of commenters shared the reason why they took HIVST. Second, the comments analyzed in this study was based on the most recent comments on December 2021. It is hard to know whether an extended data collection period would yield different results. The comments analyzed in this study were based on the top-selling products from 2 e-commerce websites and may not be representative of the whole HIVST products sold in China. Such limitations may raise some concerns with the external validity of the results. Third, although platform differences were observed, there is a lack of users' demographic information for further analysis due to methodological limitations of analyzing users' feedback comments. Fourth, it is difficult to distinguish the comments based on users' authentic experiences from intentionally produced fake reviews posted by sellers or competitive sellers [49]. Many duplicated comments generated by the same commenters posted during the same period were observed in retrieved Pinduoduo data. These comments were doubted for being fake reviews and were removed from the entire data corpus. Although we took some procedures to remove potentially fake reviews in the data, there is no guarantee that all kept comments were authentic. Future studies should consider adopting fake-review detection techniques such as developing algorithms based on big data from the social platforms or adopting sentiment analysis of written comments [54] to remove fabricated feedback comments. After removal of the fake comments, future studies may consider adopting purposive sampling methods to only examine the relevant comments related to HIVST kit users' certain experiences such as the emotional responses after getting negative test results or their complaints on kits. Fifth, several HIVST kits were sold as part of kit packages including different types of self-testing kits for detecting various diseases. Some packages include the HIVST kit and the kit to test hepatitis B or syphilis (which also requires a blood sample to conduct self-testing similar to HIVST). Because the transmission of hepatitis B and syphilis is similar to HIV transmission, it is difficult to ensure that all the comments were targeted at HIVST kits alone. On the other hand, among 1857 comments analyzed in this study, only 3 mentioned hepatitis B and only 8 mentioned syphilis. Hence, it is assumed that most comments were still concerned with HIV and AIDS while future studies should consider how to distinguish comments specifically targeting HIVST from other comments. Sixth, all comments were retrieved from either the Tmall mobile app or the Pinduoduo mobile app. While Pinduoduo is a mobile-only platform, Tmall has both mobile and website platforms, and comments from both Tmall platforms were mixed together. Future studies should explore how the mobile and website difference influences the content of feedback comments. Finally, when users enter feedback comments on Pinduoduo, some pop-up words are suggested by this platform that are displayed next to the comment section, such as "good quality and low price." If users click these pop-up words, the words will be automatically entered as a part of feedback comments. It is hard to evaluate the authenticity of certain comments that have the same content because of this platform feature, and future studies may consider how to detect these comments.

Acknowledgments

This study was partially supported by Beijing Normal University's Young Scholars' Fund.

Conflicts of Interest

None declared.

References

1. Li G, Jiang Y, Zhang L. HIV upsurge in China's students. Science 2019 May 24;364(6442):711. [doi: 10.1126/science.aay0799] [Medline: 31123112]
2. Ritchwood TD, Selin A, Pettifor A, Lippman SA, Gilmore H, Kimaru L, et al. HIV self-testing: South African young adults' recommendations for ease of use, test kit contents, accessibility, and supportive resources. BMC Public Health 2019 Jan 29;19(1):123 [FREE Full text] [doi: 10.1186/s12889-019-6402-4] [Medline: 30696422]
3. Koris AL, Stewart KA, Ritchwood TD, Mususa D, Ncube G, Ferrand RA, et al. Youth-friendly HIV self-testing: acceptability of campus-based oral HIV self-testing among young adult students in Zimbabwe. PLoS One 2021 Jun 29;16(6):e0253745 [FREE Full text] [doi: 10.1371/journal.pone.0253745] [Medline: 34185815]
4. Johnson C, Baggaley R, Forsythe S, van Rooyen H, Ford N, Napierala Mavedzenge S, et al. Realizing the potential for HIV self-testing. AIDS Behav 2014 Jul;18 Suppl 4:S391-S395. [doi: 10.1007/s10461-014-0832-x] [Medline: 24986599]
5. Wong HT, Tam HY, Chan DP, Lee SS. Usage and acceptability of HIV self-testing in men who have sex with men in Hong Kong. AIDS Behav 2015 Mar;19(3):505-515. [doi: 10.1007/s10461-014-0881-1] [Medline: 25145608]

XSL•FO
RenderX

6. Wong V, Johnson C, Cowan E, Rosenthal M, Peeling R, Miralles M, et al. HIV self-testing in resource-limited settings: regulatory and policy considerations. AIDS Behav 2014 Jul;18 Suppl 4:S415-S421. [doi: 10.1007/s10461-014-0825-9] [Medline: 24957979]
7. Liu Y, Wu G, Lu R, Ou R, Hu L, Yin Y, et al. Facilitators and barriers associated with uptake of HIV self-testing among men who have sex with men in Chongqing, China: a cross-sectional survey. Int J Environ Res Public Health 2020 Mar 03;17(5):1634 [FREE Full text] [doi: 10.3390/ijerph17051634] [Medline: 32138263]
8. Brown 3rd W, Carballo-Diéguez A, John RM, Schnall R. Information, motivation, and behavioral skills of high-risk young adults to use the HIV self-test. AIDS Behav 2016 Sep;20(9):2000-2009 [FREE Full text] [doi: 10.1007/s10461-016-1309-x] [Medline: 26885813]
9. Mugo PM, Micheni M, Shangala J, Hussein MH, Graham SM, Rinke de Wit TF, et al. Uptake and acceptability of oral HIV self-testing among community pharmacy clients in Kenya: a feasibility study. PLoS One 2017 Jan 26;12(1):e0170868 [FREE Full text] [doi: 10.1371/journal.pone.0170868] [Medline: 28125699]
10. 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]
11. Carballo-Diéguez A, Frasca T, Balan I, Ibitoye M, Dolezal C. Use of a rapid HIV home test prevents HIV exposure in a high risk sample of men who have sex with men. AIDS Behav 2012 Oct;16(7):1753-1760 [FREE Full text] [doi: 10.1007/s10461-012-0274-2] [Medline: 22893194]
12. Liu F, Qin Y, Meng S, Zhang W, Tang W, Han L, et al. HIV self-testing among men who have sex with men in China: a qualitative implementation research study. J Virus Erad 2019 Nov 04;5(4):220-224 [FREE Full text] [Medline: 31754445]
13. Martinez O, Carballo-Diéguez A, Ibitoye M, Frasca T, Brown W, Balan I. Anticipated and actual reactions to receiving HIV positive results through self-testing among gay and bisexual men. AIDS Behav 2014 Dec;18(12):2485-2495 [FREE Full text] [doi: 10.1007/s10461-014-0790-3] [Medline: 24858480]
14. Thirteenth Five-Year Plan (2017-22) for HIV prevention and control. The State Council of the People's Republic of China. 2017. URL: http://www.gov.cn/zhengce/content/2017-02/05/content_51655htm [accessed 2022-03-11]
15. The 24th World AIDS conference China satellite conference was successfully held. Fuzhou Centers for Disease Control and Prevention. 2022. URL: https://cdc.fuzhou.gov.cn/zz/crbkz/azb/202208/t20220804_4410916.htm [accessed 2022-10-16]
16. The manual of AIDS self-testing guidance. Chinese Centers for Disease Control and Prevention. 2019. URL: https://ncaids.chinacdc.cn/xxgx/jszl/202011/t20201123_222904.htm [accessed 2021-12-16]
17. Han L, Bien CH, Wei C, Muessig KE, Yang M, Liu F, et al. HIV self-testing among online MSM in China: implications for expanding HIV testing among key populations. J Acquir Immune Defic Syndr 2014 Oct 01;67(2):216-221 [FREE Full text] [doi: 10.1097/QAI.0000000000000278] [Medline: 24991972]
18. Zhong F, Tang W, Cheng W, Lin P, Wu Q, Cai Y, et al. Acceptability and feasibility of a social entrepreneurship testing model to promote HIV self-testing and linkage to care among men who have sex with men. HIV Med 2017 May;18(5):376-382 [FREE Full text] [doi: 10.1111/hiv.12437] [Medline: 27601301]
19. Yan H, Yang H, Raymond HF, Li J, Shi LE, Huan X, et al. Experiences and correlates of HIV self-testing among men who have sex with men in Jiangsu province, China. AIDS Behav 2015 Mar;19(3):485-491 [FREE Full text] [doi: 10.1007/s10461-014-0968-8] [Medline: 25480598]
20. Huang Z, Dai W, Zhou Y, Li X, Lin K, Jiang H. Preference and related factors of HIV testing modes among men who have sex with men. Chinese J AIDS STD 2020;26(4):373-380.
21. Lu Y, He X, Ma J, Yao J, Xin W, Liu P, et al. Online purchase of HIV rapid detection reagents for self-testing population characteristics. Chinese J AIDS STD 2019;25(6).
22. Zhang C, Li X, Heilemann MV, Chen X, Wang H, Koniak-Griffin D. Facilitators and barriers of HIV self-testing among Chinese men who have sex with men: a qualitative study. J Assoc Nurses AIDS Care 2021;32(5):599-609 [FREE Full text] [doi: 10.1097/JNC.0000000000000218] [Medline: 33165181]
23. Yang LH, Kleinman A. 'Face' and the embodiment of stigma in China: the cases of schizophrenia and AIDS. Soc Sci Med 2008 Aug;67(3):398-408 [FREE Full text] [doi: 10.1016/j.socscimed.2008.03.011] [Medline: 18420325]
24. Blackstone A. Principles of Sociological Inquiry: Qualitative and Quantitative Methods. Version 1.0. Washington, DC, USA: Saylor Foundation; 2014.
25. Eysenbach G. Infodemiology: the epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
26. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
27. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
28. Manchaiah V, Amlani AM, Bricker CM, Whitfield CT, Ratinaud P. Benefits and shortcomings of direct-to-consumer hearing devices: analysis of large secondary data generated from Amazon customer reviews. J Speech Lang Hear Res 2019 May 21;62(5):1506-1516. [doi: 10.1044/2018_JSLHR-H-18-0370] [Medline: 31034320]

XSL•FO
RenderX

29. Zeraatkar K, Ahmadi M. Trends of infodemiology studies: a scoping review. Health Info Libr J 2018 Jun;35(2):91-120 [FREE Full text] [doi: 10.1111/hir.12216] [Medline: 29729073]
30. Lee A, Mmonu NA, Thomas H, Rios N, Enriquez A, Breyer BN. Qualitative analysis of Amazon customer reviews of penile clamps for male urinary incontinence. Neurourol Urodyn 2021 Jan;40(1):384-390. [doi: 10.1002/nau.24572] [Medline: 33165983]
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. Riffe D, Lacy S, Fico F. Analyzing Media Messages: Using Quantitative Content Analysis in Research. 2nd edition. London, UK: Routledge; 2005.
33. Walk-Morris T. What is it about Tmall? Retail Dive. 2019 Sep 19. URL: https://www.retaildive.com/news/what-is-it-about-tmall/562907/ [accessed 2022-08-28]
34. 22 amazing Tmall statistics. DMR. 2017. URL: https://expandedramblings.com/index.php/tmall-statistics/ [accessed 2022-03-17]
35. CIW Team. Pinduoduo's active buyers totaled 867 million in Q3 2021. China Internet Watch. 2021. URL: https://www.chinainternetwatch.com/31102/pinduoduo-quarterly/ [accessed 2022-03-17]
36. Shank GD. Qualitative Research: A Personal Skills Approach. 2nd edition. New York, NY, USA: Pearson Merrill Prentice Hall; 2005.
37. Frasca T, Balan I, Ibitoye M, Valladares J, Dolezal C, Carballo-Diéguez A. Attitude and behavior changes among gay and bisexual men after use of rapid home HIV tests to screen sexual partners. AIDS Behav 2014 May;18(5):950-957 [FREE Full text] [doi: 10.1007/s10461-013-0630-x] [Medline: 24077975]
38. Zhang YX, Golin CE, Bu J, Emrick CB, Nan Z, Li MQ. Coping strategies for HIV-related stigma in Liuzhou, China. AIDS Behav 2014 Feb;18 Suppl 2:S212-S220 [FREE Full text] [doi: 10.1007/s10461-013-0662-2] [Medline: 24337724]
39. Zhou YR. Morality, discrimination, and silence: understanding HIV stigma in the sociocultural context of China. In: Liamputtong P, editor. Stigma, Discrimination and Living with HIV/AIDS: A Cross-Cultural Perspective. Dordrecht, The Netherlands: Springer; 2013:117-132.
40. FFCell. URL: http://www.ffcell.com/ [accessed 2022-11-22]
41. Qu Z, Zhang H, Li H. Determinants of online merchant rating: content analysis of consumer comments about Yahoo merchants. Decis Support Syst 2008 Dec;46(1):440-449. [doi: 10.1016/j.dss.2008.08.004]
42. Pavlou PA, Dimoka A. The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation. Inf Syst Res 2006 Dec;17(4):392-414. [doi: 10.1287/isre.1060.0106]
43. Sileo KM, Bogart LM, Wagner GJ, Musoke W, Naigino R, Mukasa B, et al. HIV fatalism and engagement in transactional sex among Ugandan fisherfolk living with HIV. SAHARA J 2019 Dec;16(1):1-9 [FREE Full text] [doi: 10.1080/17290376.2019.1572533] [Medline: 30727838]
44. Hess RF, McKinney D. Fatalism and HIV/AIDS beliefs in rural Mali, West Africa. J Nurs Scholarsh 2007;39(2):113-118. [doi: 10.1111/j.1547-5069.2007.00155.x] [Medline: 17535310]
45. Akande A. Black South African adolescents' attitudes towards AIDS precautions. School Psychol Int 1997 Nov;18(4):325-341. [doi: 10.1177/0143034397184003]
46. Lam C, Tsang HW, Corrigan P, Lee YT, Angell B, Shi K, et al. Chinese lay theory and mental illness stigma: implications for research and practices. J Rehabil 2010 Jan;76(1):35-40.
47. Zhai X. Perspectives on Chinese "Face": Psychological Motives and Social Representations. Beijing, China: Peking University; 2011.
48. Ibitoye M, Frasca T, Giguere R, Carballo-Diéguez A. Home testing past, present and future: lessons learned and implications for HIV home tests. AIDS Behav 2014 May;18(5):933-949 [FREE Full text] [doi: 10.1007/s10461-013-0668-9] [Medline: 24281697]
49. Malbon J. Taking fake online consumer reviews seriously. J Consum Policy 2013 Jan 23;36(2):139-157. [doi: 10.1007/s10603-012-9216-7]
50. Graziani T. Pinduoduo expands to Tier 1 cities. WalktheChat. 2019 Aug 26. URL: https://walkthechat.com/pinduoduo-expands-to-tier-1-cities/ [accessed 2022-03-01]
51. Product Analysis Report-Tmall. Jianshu. 2018. URL: https://www.jianshu.com/p/2a90e0684193 [accessed 2022-03-25]
52. Yang LH, Kleinman A, Link BG, Phelan JC, Lee S, Good B. Culture and stigma: adding moral experience to stigma theory. Soc Sci Med 2007 Apr;64(7):1524-1535. [doi: 10.1016/j.socscimed.2006.11.013] [Medline: 17188411]
53. Marley G, Kang D, Wilson EC, Huang T, Qian Y, Li X, et al. Introducing rapid oral-fluid HIV testing among high risk populations in Shandong, China: feasibility and challenges. BMC Public Health 2014 May 03;14:422 [FREE Full text] [doi: 10.1186/1471-2458-14-422] [Medline: 24884431]
54. Reyes-Menendez A, Saura JR, Filipe F. The importance of behavioral data to identify online fake reviews for tourism businesses: a systematic review. PeerJ Comput Sci 2019 Sep 23;5:e219 [FREE Full text] [doi: 10.7717/peerj-cs.219] [Medline: 33816872]

XSL•FO
RenderX

Abbreviations

HIVST: HIV self-testing
MSM: men who have sex with men
VCT: voluntary counseling and testing

Edited by G Eysenbach; submitted 31.03.22; peer-reviewed by Y Zhang, J Huang; comments to author 18.07.22; revised version received 29.08.22; accepted 20.10.22; published 30.11.22.

Please cite as:
Liu M, Zhu Y, Gao H, Li J
Examining Chinese Users' Feedback Comments on HIV Self-testing Kits From e-Commerce Platforms: Thematic and Content Analysis
J Med Internet Res 2022;24(11):e38398
URL: https://www.jmir.org/2022/11/e38398
doi:10.2196/38398
PMID:36449327

©Miao Liu, Yi Zhu, Haiqing Gao, Jialing Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Ice Flavor–Related Discussions on Twitter: Content Analysis

Artur Galimov[1], MD, PhD; Julia Vassey[1], MPH, MSc; Ellen Galstyan[1], BA; Jennifer B Unger[1], PhD; Matthew G Kirkpatrick[1], PhD; Jon-Patrick Allem[1], PhD

Department of Population and Public Health Sciences, Keck School of Medicine of University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Jon-Patrick Allem, PhD
Department of Population and Public Health Sciences
Keck School of Medicine of University of Southern California
1845 N Soto St
Los Angeles, CA, 90089-9239
United States
Phone: 1 858 603 0812
Email: allem@usc.edu

Abstract

Background: The US Food and Drug Administration (FDA) recently restricted characterizing flavors in tobacco products. As a result, ice hybrid–flavored e-cigarettes, which combine a cooling flavor with fruit or other flavors (eg, banana ice), emerged on the market. Like menthol, ice-flavored e-cigarettes produce a cooling sensory experience. It is unclear if ice hybrid–flavored e-cigarettes should be considered characterizing flavors or menthol, limiting regulatory action. Monitoring the public's conversations about ice-flavored e-cigarettes on Twitter may help inform the tobacco control community about these products and contribute to the US FDA policy targets in the future.

Objective: This study documented the themes pertaining to vaping and ice flavor–related conversations on Twitter. Our goal was to identify key conversation trends and ascertain users' recent experiences with ice-flavored e-cigarette products.

Methods: Posts containing vaping-related (eg, "vape," "ecig," "e-juice," or "e-cigarette") and ice-related (ie, "Ice," "Cool," "Frost," and "Arctic") terms were collected from Twitter's streaming application programming interface from January 1 to July 21, 2021. After removing retweets, a random sample of posts (N=2001) was selected, with 590 posts included in the content analysis. Themes were developed through an inductive approach. Theme co-occurrence was also examined.

Results: Many of the 590 posts were marked as (or consisted of) marketing material (n=306, 51.9%), contained positive personal testimonials (n=180, 30.5%), and mentioned disposable pods (n=117, 19.8%). Other themes had relatively low prevalence in the sample: neutral personal testimonials (n=45, 7.6%), cannabidiol products (n=41, 7%), negative personal testimonials (n=41, 7%), "official" flavor description (n=37, 6.3%), ice-flavored JUUL (n=19, 3.2%), information seeking (n=14, 2.4%), and comparison to combustible tobacco (n=10, 1.7%). The most common co-occurring themes in a single tweet were related to marketing and disposable pods (n=73, 12.4%).

Conclusions: Our findings offer insight into the public's experience with and understanding of ice-flavored e-cigarette products. Ice-flavored e-cigarette products are actively marketed on Twitter, and the messages about them are positive. Public health education campaigns on the harms of flavored e-cigarettes may help to reduce positive social norms about ice-flavored products. Future studies should evaluate the relationship between exposure to personal testimonials of ice-flavored vaping products and curiosity, harm perceptions, and experimentation with these products among priority populations.

(J Med Internet Res 2022;24(11):e41785) doi:10.2196/41785

KEYWORDS

electronic cigarettes; Twitter; social media; ice flavors; tobacco policy; public health; infodemiology; FDA; tobacco; smoking; vaping; e-cigarette; public

Introduction

The use of e-cigarettes, or vaping, among adolescents and young adults is a public health concern [1,2]. One factor contributing to e-cigarette use among adolescents and young adults is product diversity and the availability of appealing flavors [3,4]. Use of flavored e-cigarettes like fruit, sweet, mint, and menthol are the most widely used flavors among adolescents and young adults

XSL•FO
RenderX

[4-6]. In February 2020, to counteract e-cigarette use among adolescents and young adults, the US Food and Drug Administration (FDA) restricted the manufacture, distribution, marketing, and sale of prefilled cartridge-based e-cigarettes (eg, JUUL) in flavors other than menthol and tobacco [7]. Additionally in July 2020, the US FDA issued warning letters to 10 prominent e-cigarette manufactures, including Cool Clouds Distribution, the parent company of the popular brand Puff Bar, requesting that they remove their flavored disposable e-cigarette products because they lacked the required premarket authorization [8,9]. In the context of these regulations, ice hybrid–flavored products were introduced into the market [3,7]. Ice hybrid–flavored products combine a cooling flavor with fruit, dessert, or other characterizing flavor (eg, blueberry ice or melon ice), which may allow e-cigarette companies to circumvent regulatory action [10,11]. Recent evidence suggests that the use of ice-flavored e-cigarettes among young adults may be common and positively associated with combustible tobacco use [10,12].

Ice hybrid–flavored products may contain menthol. Menthol is a flavor additive that produces pleasant cooling sensations and analgesic effects in the throat and mouth, which reduces the harshness of combustible smoke and nicotine's irritating effects on the airways [11,12]. Research from human and animal studies indicates that menthol increases the intensity of how cigarettes are smoked (eg, deeper inhalation) and facilitates smoking initiation and nicotine intake [11,13,14]. In addition, studies have demonstrated that menthol additives in e-liquids mask the harshness and irritation associated with vaping nicotine and increases the liking of vaping products [12,15,16]. Many ice-flavored e-cigarette products contain synthetic coolants (eg, menthone, eucalyptol, peppermint oil, and WS-3) that produce comparable or stronger menthollike cooling and counterirritant effects [11,12,14]. In fact, even low levels of these cooling additives, which may not have been classified as a characterizing flavor, could increase appeal, promote deeper aerosol inhalation, and encourage more frequent e-cigarette use [11,12,15]. Thus, ice flavors may not fit into existing flavor profile categories, such as characterizing flavors (eg, fruit) or menthol, which may circumvent current regulatory efforts [10].

One approach that may be helpful in informing regulatory efforts is the tracking and analyzing of real-world conversations and depictions of novel tobacco products on Twitter [17]. Monitoring tobacco-related content on Twitter can provide valuable insights into the public's beliefs, attitudes, and experiences surrounding new tobacco products [18-20]. Engagement (eg, seeing, posting, liking, and sharing) with tobacco-related content and posting about tobacco products on Twitter, Facebook, and other social media is associated with an increased risk of tobacco use [21,22]. While Twitter users have been shown to be vulnerable to the effects of positive messages about vaping, to the best of our knowledge, prior research has not assessed ice flavor discussions with Twitter data [23].

This study used Twitter data to document the themes in posts pertaining to ice flavors. Examining and monitoring the public's conversations about ice-flavored e-cigarettes may help inform the tobacco control community about these products and uncover trends, knowledge, and attitudes toward e-cigarette flavors. Findings from this study may contribute to the US FDA policy targets and tobacco control campaigns in the future.

Methods

Overview

Twitter posts containing both vaping-related (eg, "vape," "ecig," "e-juice," "e-cigarette," or "JUUL"; see Multimedia Appendix 1 for the complete list of keywords) and "Ice," "Cool," "Frost," or "Arctic" terms were collected from January 1 to July 21, 2021, from Twitter's streaming application programming interface. A total of 976,347 posts containing these terms were identified. Similar to previous studies, after excluding all retweets, we sampled out a random subset of 2001 posts for a content analysis [18,20]. Two coders (AG and EG) worked together with the last author (JPA) to become familiar with the data, created a codebook, and coded tweets into themes using an inductive approach. The unit of analysis was the text of the tweet. The purpose of the approach was to summarize the raw text-based data into a summary format and report the underlying themes evident in the data. Saturation was determined to be reached with 10 themes.

Vaping-related posts that contained the terms "Ice," "Cool," "Frost," or "Arctic" but were determined to be unrelated to our research objectives were identified through manual content analysis and removed. For instance, irrelevant posts contained the following phrases: "cool vape trick" and "vaping is cool." After excluding irrelevant and non-English posts, we were left with 590 (29.5%) tweets that were included in the content analysis.

The codebook consists of the following themes: *personal testimonials* (which were further divided into the following five subcategories: *positive sentiment*, *negativesentiment*, *neutral sentiment*, *information seeking*, and *comparison to combustible tobacco*), *marketing*, *official (promotional) flavor description*, *disposable pod devices*, *JUUL*, *cannabidiol* (*CBD*) *products*, and o*ther*. A tweet could be classified into multiple themes. To establish interrater reliability, the same subsample of posts (590 of the original 2001 posts) were double coded, with percent agreement ranging from 99.2% to 100% and Cohen κ ranging from 0.90 to 1.00. JPA served as an arbitrator resolving discrepancies between the coders.

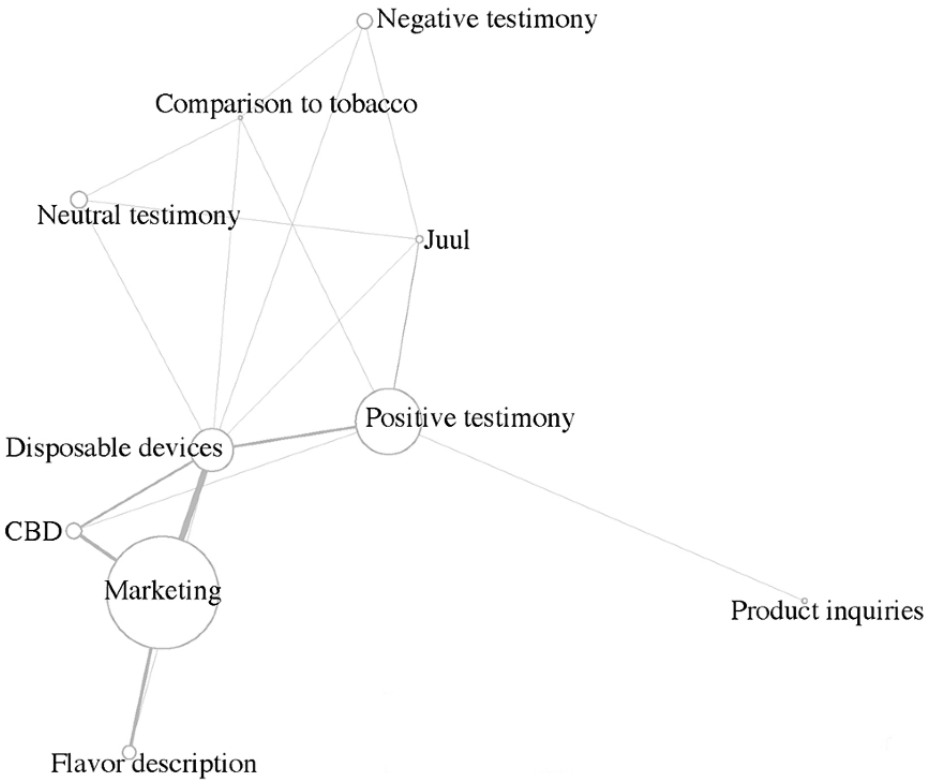
Descriptive analyses were conducted first to show the prevalence of each theme. Further, pairwise co-occurrence analyses were used to evaluate connections across themes, since tweets often included more than one theme. R base package (version 4.0.2; R Foundation for Statistical Computing) was used to construct the co-occurrence matrix, and the igraph package (version 1.2.6) was used to visualize the results. Since the theme related to *personal testimonials* consisted of *positive*, *negative*, and *neutral testimonies*, Figure 1 does not include the overarching *personal testimonials* theme, and only includes its components: *positive*, *negative*, and *neutral testimonies*. The *other* category was also excluded since it had few co-occurrences with other themes.

All Twitter posts in this data set were publicly available and anonymized, and all analyses adhered to the terms and conditions, terms of use, and privacy policies of Twitter. To

XSL•FO
RenderX

further protect privacy, posts exemplifying themes were paraphrased; no tweets are reported verbatim.

Figure 1. Co-occurrences of themes in corpus tweets. The size of the circles represents the frequency of the theme occurrences. The proximity of the circles and the size of the lines represent the frequency of the theme pairwise co-occurrences. CBD: cannabidiol.



Ethical Approval

The protocol was approved by the university's institutional review board (protocol HS-18-00697).

Results

Descriptive characteristics of the 590 tweets from this corpus and example paraphrased posts are presented in Table 1. Overall, the most prominent themes were *marketing* (n=306, 51.9%) of ice-flavored products and *personal testimonials* (n=277, 47%). Mentions of using ice-flavored *disposable pod devices* (n=117, 19.8%) was also a common theme. Other themes had relatively low prevalence in the sample: "ice-flavored" *CBD products* (n=41, 7%), *official flavor description* (n=37, 6.3%), and ice-flavored *JUUL* (n=19, 3.2%). *Positive sentiment* (n=180, 30.5%) was the most prevalent theme among personal testimonials, while the other testimonial themes were less common: *neutral sentiment* (n=45, 7.6%), *negative sentiment* (n=41, 7%), *information seeking* (n=14, 2.4%), and *comparison to combustible tobacco* (n=10, 1.7%).

The most commonly co-occurring themes in a single tweet were related to the most prominent themes in the data set: *marketing*, *positive sentiment*, and *disposable pod devices*. The marketing of ice-flavored disposable pod devices was mentioned in 73 (12.4%) of the 590 tweets; the marketing of ice-flavored CBD products was mentioned in 40 (6.8%) of the tweets; the marketing of ice-flavored products and official flavor descriptions was mentioned in 37 (6.3%) of the tweets. Positive testimonies about disposable devices were observed in 31 (5.3%) of the tweets, and conversations about disposable devices containing CBD were observed in 26 (4.4%) of the tweets. Pairwise co-occurrences of the remaining themes were found in 2% of the tweets (see Figure 1, where the thickness of the lines represents the frequency of pairwise co-occurrences and the size of the circles represents the frequency of theme occurrences).

XSL•FO
RenderX

Table 1. Definition for each theme, descriptive statistics, and selected paraphrased twitter posts.

Theme	Posts (n=590), n (%)	Definition	Paraphrased post
Positive sentiment testimonials	180 (30.5)	Posts containing a positive experience with (eg, tastes, smells great) or favorable opinion about (eg, love, like, my favorite) "ice-flavored" tobacco products, including e-liquids	"Lychee Ice is the best e-cigarette flavor"
Negative sentiment testimonials	41 (7.0)	Posts containing a negative experience with (eg, tastes, smells horrible, bad) or negative opinion about (eg, hate, dislike) "ice-flavored" tobacco products, including e-liquids	"Recently bought a banana ice vape and it tastes horrible..."
Neutral sentiment testimonials	45 (7.6)	Posts containing a neutral opinion about "ice-flavored" tobacco products, including e-liquids (eg, if no valence is determinable or deemed neutral)	"This tangerine-apple ice e-cigarette tastes like 7-up"
Information seeking	14 (2.4)	Posts where a consumer or potential consumer asks for information on or opinions about "ice-flavored" tobacco products, including e-liquids	"Do you guys know if a mango ice e-liquid tastes good?"
Comparison to combustible tobacco	10 (1.7)	Posts containing mentions of or comparisons between "ice-flavored" products, including e-liquids and combustible tobacco products	"...I am not planning to switch from my menthol cigarettes to vaping, but I will definitely buy one of these lychee ice flavored disposables..."
Marketing	306 (51.9)	Posts promoting/selling/marketing an "ice-flavored" e-liquids or devices	"Guava ice e-liquid is now available at https://xxx..."
Disposable pod devices	117 (19.8)	Mentions of "ice-flavored" disposable pod devices (ie, Puff Bar)	"Banana Ice Puff Bar is worth every penny..."
JUUL	19 (3.2)	Mentions of "ice-flavored" JUUL	"@xxx Ice flavored JUUL is the best JUUL"
CBD[a] products	41 (7.0)	Mentions of "ice-flavored" CBD vaping products	"New CBD Blackcurrant Ice Disposable vapes are now in stock at https://xxx..."
Official flavor description	37 (6.3)	Posts containing an official (ie, "media," promotional, marketing) description of "ice-flavored" e-liquids or products	"...The sweet juice of Fuji apples combined with a hint of nectarines & strawberry to create a well rounded fruity vape experience with a cool refreshing finish..."
Other	6 (1.0)	Any other posts that contain ice flavor–related themes and do not fit into any category listed in the codebook	"Use of ice-flavored e-cigarettes may be associated with nicotine dependence. See https://xxx..."

[a]CBD: cannabidiol.

Discussion

Principal Findings

This study provides a summary of public Twitter posts collected over the course of a 7-month period, which includes mentions of both vaping-related and ice flavor–related terms. Posts in our corpus were related to ice-flavored e-cigarette product marketing, personal testimonials, and ice-flavored disposable pod devices. Theme co-occurrence in a single post was examined.

Marketing was a common theme in this study, while marketing and disposable pod devices represented a common theme co-occurrence in a single post. These findings are consistent with recent studies suggesting that ice-flavored products are often promoted in various pod-style cartridge-based disposable (eg, Puff Bar) and refillable (eg, PUFF Krush, PHIX) e-cigarette products [11]. These disposable and refillable pod-style products are among the fastest-growing segments of the e-cigarette market [3,9,10]. To reduce the appeal of e-cigarette products among youth, the US FDA took regulatory action (February 2020) to remove prefilled cartridge-based e-cigarettes in flavors other than menthol and tobacco, and issued warning letters to 10 e-cigarette companies instructing them to stop selling flavored disposable e-cigarettes by July 2020 [7,8]. Although some companies (eg, JUUL) voluntarily removed all flavors except menthol and tobacco, other companies (including those producing JUUL-compatible pods) continued to manufacture them. Given that ice hybrid–flavored products may contain both cooling and fruity flavors, it is unclear how these flavors fit into current regulatory policies. While the US FDA is moving toward removing menthol from combustible tobacco products, it is critical that regulatory actions consider restricting any constituents that taste/smell like menthol and produce a similar cooling sensory experience (ie, ice flavors) [12,24].

Personal testimonials was another common theme in this study, while *positive sentiment* was the most common category among testimonials. These findings are consistent with previous studies showing that Twitter users are exposed to positive messages about vaping and tobacco [25,26]. Positive messages about tobacco on social media may be linked with tobacco product experimentation [21]. Given that the messages about the ice-flavored products on Twitter are positive and that user

XSL•FO
RenderX

conversations are not subject to federal policies, public health education campaigns informing the public on the harms of flavored e-cigarettes may be beneficial in countering these messages and reducing positive social norms about these products [21,26]. For instance, it could be possible to target individuals posting positive messages about ice-flavored e-cigarette products and deliver cessation messages directly to these individuals and their social networks. In addition, the US FDA regulations banning tobacco marketing on Twitter may help to limit the promotion of ice flavors to the public.

Limitations

This study was limited to the analysis of discussions on ice-flavored e-cigarette products and may not pertain to other e-cigarette flavors. In addition, relatively few (n=2001, 0.2%) tweets among all identified ice tweets were analyzed (n=976,347), which limits the generalizability of our results. Moreover, the word "cool" was one of the four terms used to identify ice-flavored e-cigarette posts. "Cool" is a colloquialism and is commonly used to express the enjoyment of someone or something. The inclusion of this search term resulted in a collection of many posts that were unrelated to e-cigarette flavors; nonetheless, we found this limitation unavoidable and worked diligently to analyze pertinent tweets. This study focused on the text of the Twitter posts but did not code website links or accompanying images. Previous studies demonstrated that there is value in examining both text and image. In other words, it is possible that additional themes would have emerged had we coded images. Findings may not generalize to other time periods or other social media platforms. Our findings may not extend to all Twitter users or to the population of the United States.

Conclusions

Our findings may offer valuable insights into the public's experience with and understanding of ice-flavored e-cigarette products. In this study, we found that Twitter discussions about ice-flavored e-cigarette products focused on marketing, personal testimonials, and ice-flavored disposable pod devices. Future studies should evaluate the relationship between exposure to personal testimonials of ice-flavored vaping products and curiosity, harm perceptions, and experimentation with these products among priority populations. Public health education campaigns informing the public on the harms of flavored e-cigarettes may be helpful in reducing positive social norms about ice-flavored products, while flavor regulatory actions may include ice flavors to prevent new products from circumventing current regulations. For instance, banning tobacco marketing posts on Twitter may reduce the exposure to ice-flavored e-cigarette product promotions. In addition, social media might be more influential than traditional marketing because participants can be actively engaged in the content they are viewing [21,22]. Close monitoring of ice flavor promotions on social media is needed to ensure that these promotions are not targeted toward minors and e-cigarette nonusers. Twitter has a policy prohibiting paid advertising of tobacco products to appear on its platform [27]. However, promotional posts related to ice-flavored products are still prevalent. Future studies should evaluate whether these promotions are sponsored by e-cigarette manufacturers or e-cigarette product distributors, or generated by general users.

Acknowledgments

The research reported in this publication was supported by a National Cancer Institute (NCI) and Food and Drug Administration (FDA) Center for Tobacco Products award (NCI/FDA grant U54CA180905, Mary Ann Pentz and Adam Leventhal, principal investigators).

The NCI or the FDA had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors took an active role in the manuscript. AG, J-PA, and JBU conceptualized and designed the study. AG, J-PA, JBU, and JV created the codebook for data analysis. AG and EG completed data coding. AG, J-PA, and JV acquired, analyzed, and interpreted the data. AG, EG, and JV drafted the initial manuscript. AG, JV, EG, JBU, MGK, and J-PA revised the manuscript for important intellectual content. JBU and J-PA obtained funding for this study. All authors approved the final manuscript submitted.

Conflicts of Interest

J-PA has received fees for consulting services in court cases pertaining to the content on social media platforms. He reports no other conflicts of interest. All other authors declare no competing interests.

Multimedia Appendix 1
List of keywords.

XSL•FO
RenderX

[DOCX File , 16 KB - jmir_v24i11e41785_app1.docx]

References

1. Fuentes AL, Crotty Alexander LE. Beware, vaping e-cigarettes around children is adversely impacting their lung health. Thorax 2022 Jul;77(7):638-639. [doi: 10.1136/thoraxjnl-2021-218168] [Medline: 35012999]
2. Johnston L, Miech R, O'Malley P, Bachman J, Schulenberg J, Patrick M. Monitoring the future national survey results on drug use, 1975-2020: overview, key findings on adolescent drug use. Institute for Social Research, University of Michigan. Ann Arbor; 2021. URL: https://files.eric.ed.gov/fulltext/ED611736.pdf [accessed 2021-11-19]
3. Galimov A, Leventhal A, Meza L, Unger JB, Huh J, Baezconde-Garbanati L, et al. Prevalence of disposable pod use and consumer preference for e-cigarette product characteristics among vape shop customers in Southern California: a cross-sectional study. BMJ Open 2021 Oct 05;11(10):e049604 [FREE Full text] [doi: 10.1136/bmjopen-2021-049604] [Medline: 34610937]
4. Rostron BL, Cheng Y, Gardner LD, Ambrose BK. Prevalence and reasons for use of flavored cigars and ENDS among US youth and adults: estimates from wave 4 of the PATH study, 2016-2017. Am J Health Behav 2020 Jan 01;44(1):76-81 [FREE Full text] [doi: 10.5993/AJHB.44.1.8] [Medline: 31783934]
5. Wang TW, Neff LJ, Park-Lee E, Ren C, Cullen KA, King BA. E-cigarette use among middle and high school students - United States, 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 18;69(37):1310-1312. [doi: 10.15585/mmwr.mm6937e1] [Medline: 32941408]
6. Zare S, Nemati M, Zheng Y. A systematic review of consumer preference for e-cigarette attributes: flavor, nicotine strength, and type. PLoS One 2018;13(3):e0194145 [FREE Full text] [doi: 10.1371/journal.pone.0194145] [Medline: 29543907]
7. FDA finalizes enforcement policy on unauthorized flavored cartridge-based e-cigarettes that appeal to children, including fruit and mint. US Food and Drug Administration. 2020. URL: https://www.fda.gov/news-events/press-announcements/fda-finalizes-enforcement-policy-unauthorized-flavored-cartridge-based-e-cigarettes-appeal-children [accessed 2022-07-03]
8. FDA notifies companies, including Puff Bar, to remove flavored disposable e-cigarettes and youth-appealing e-liquids from market for not having required authorization. US Food and Drug Administration. 2020. URL: https://tinyurl.com/mvw49jaf [accessed 2022-07-03]
9. Cwalina SN, McConnell R, Benowitz NL, Barrington-Trimis JL. Tobacco-free nicotine - new name, same scheme? N Engl J Med 2021 Dec 23;385(26):2406-2408 [FREE Full text] [doi: 10.1056/NEJMp2111159] [Medline: 34919358]
10. Leventhal A, Dai H, Barrington-Trimis J, Sussman S. 'Ice' flavoured e-cigarette use among young adults. Tob Control 2021 Jun 14:tobaccocontrol-2020-056416. [doi: 10.1136/tobaccocontrol-2020-056416] [Medline: 34127549]
11. Leventhal AM, Tackett AP, Whitted L, Jordt SE, Jabba SV. Ice flavours and non-menthol synthetic cooling agents in e-cigarette products: a review. Tob Control 2022 Apr 28:tobaccocontrol-2021-057073. [doi: 10.1136/tobaccocontrol-2021-057073] [Medline: 35483721]
12. Davis DR, Morean ME, Bold KW, Camenga D, Kong G, Jackson A, et al. Cooling e-cigarette flavors and the association with e-cigarette use among a sample of high school students. PLoS One 2021;16(9):e0256844 [FREE Full text] [doi: 10.1371/journal.pone.0256844] [Medline: 34469460]
13. Leventhal A, Cho J, Barrington-Trimis J, Pang R, Schiff S, Kirkpatrick M. Sensory attributes of e-cigarette flavours and nicotine as mediators of interproduct differences in appeal among young adults. Tob Control 2020 Nov;29(6):679-686 [FREE Full text] [doi: 10.1136/tobaccocontrol-2019-055172] [Medline: 31852818]
14. Kreslake JM, Wayne GF, Alpert HR, Koh HK, Connolly GN. Tobacco industry control of menthol in cigarettes and targeting of adolescents and young adults. Am J Public Health 2008 Sep;98(9):1685-1692. [doi: 10.2105/AJPH.2007.125542] [Medline: 18633084]
15. Krishnan-Sarin S, Green BG, Kong G, Cavallo DA, Jatlow P, Gueorguieva R, et al. Studying the interactive effects of menthol and nicotine among youth: an examination using e-cigarettes. Drug Alcohol Depend 2017 Nov 01;180:193-199 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.07.044] [Medline: 28915478]
16. Rosbrook K, Green BG. Sensory effects of menthol and nicotine in an e-cigarette. Nicotine Tob Res 2016 Jul;18(7):1588-1595 [FREE Full text] [doi: 10.1093/ntr/ntw019] [Medline: 26783293]
17. Barker JO, Vassey J, Chen-Sankey JC, Allem J, Cruz TB, Unger JB. Categorizing IQOS-related Twitter discussions. Int J Environ Res Public Health 2021 Apr 30;18(9):4836 [FREE Full text] [doi: 10.3390/ijerph18094836] [Medline: 33946546]
18. Allem J, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. Drug Alcohol Depend 2018 Sep 01;190:1-5 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.05.018] [Medline: 29958115]
19. Allem J, Dormanesh A, Majmundar A, Rivera V, Chu M, Unger JB, et al. Leading topics in Twitter discourse on JUUL and Puff Bar products: content analysis. J Med Internet Res 2021 Jul 19;23(7):e26510 [FREE Full text] [doi: 10.2196/26510] [Medline: 34279236]
20. Allem J, Escobedo P, Chu K, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. Tob Control 2017 Mar;26(2):226-229 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052757] [Medline: 26956467]

XSL•FO
RenderX

21. Unger JB, Urman R, Cruz TB, Majmundar A, Barrington-Trimis J, Pentz MA, et al. Talking about tobacco on Twitter is associated with tobacco product use. Prev Med 2018 Sep;114:54-56 [FREE Full text] [doi: 10.1016/j.ypmed.2018.06.006] [Medline: 29898418]
22. Soneji S, Yang J, Knutzen KE, Moran MB, Tan AS, Sargent J, et al. Online tobacco marketing and subsequent tobacco use. Pediatrics 2018 Feb;141(2):e20172927 [FREE Full text] [doi: 10.1542/peds.2017-2927] [Medline: 29295893]
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. FDA proposes rules prohibiting menthol cigarettes and flavored cigars to prevent youth initiation, significantly reduce tobacco-related disease and death. US Food and Drug Administration. 2022. URL: https://www.fda.gov/news-events/press-announcements/fda-proposes-rules-prohibiting-menthol-cigarettes-and-flavored-cigars-prevent-youth-initiation [accessed 2022-07-03]
25. 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]
26. Allem J, Ramanujam J, Lerman K, Chu K, Boley Cruz T, Unger JB. Identifying sentiment of hookah-related posts on Twitter. JMIR Public Health Surveill 2017 Oct 18;3(4):e74 [FREE Full text] [doi: 10.2196/publichealth.8133] [Medline: 29046267]
27. Tobacco and tobacco accessories. Twitter for Business. URL: https://business.twitter.com/en/help/ads-policies/ads-content-policies/tobacco-and-tobacco-accessories.html [accessed 2022-07-03]

Abbreviations

CBD: cannabidiol
FDA: Food and Drug Administration

Edited by A Mavragani; submitted 08.08.22; peer-reviewed by M Nali, K Minji, M Navarro; comments to author 27.09.22; revised version received 26.10.22; accepted 09.11.22; published 30.11.22.

Please cite as:
Galimov A, Vassey J, Galstyan E, Unger JB, Kirkpatrick MG, Allem JP
Ice Flavor–Related Discussions on Twitter: Content Analysis
J Med Internet Res 2022;24(11):e41785
URL: https://www.jmir.org/2022/11/e41785
doi:10.2196/41785
PMID:36449326

©Artur Galimov, Julia Vassey, Ellen Galstyan, Jennifer B Unger, Matthew G Kirkpatrick, Jon-Patrick Allem. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Lessons Learned From Transition of an In-Person to a Virtual Randomized Controlled Trial for Weight and Fitness Concerns in Active-Duty Service Members: Survey Study

Niloofar Afari[1,2,3], PhD; Natalie M Yarish[2], PhD; Jennalee S Wooldridge[1,2], PhD; Karla Materna[2], PhD; Jeffrey Hernandez[2], BS; Brian H Blanco[2], MSW; Angela L Camodeca[2], BS; Joshua J Peters[4], BS; Matthew S Herbert[1,2,3], PhD

[1]Department of Psychiatry, University of California, San Diego, La Jolla, CA, United States

[2]Veterans Affairs San Diego Healthcare System, San Diego, CA, United States

[3]Veterans Affairs Center of Excellence for Stress and Mental Health, San Diego, CA, United States

[4]Navy Medicine Readiness and Training Command, San Diego, CA, United States

Corresponding Author:
Niloofar Afari, PhD
Department of Psychiatry
University of California, San Diego
9500 Gilman Drive, 0737
La Jolla, CA, 92093
United States
Phone: 1 858 642 3387
Email: nafari@ucsd.edu

Abstract

Background: This paper describes and discusses the transition of and modifications to a weight management randomized controlled trial among active-duty military personnel from an in-person to a virtual format as a result of the COVID-19 pandemic. The original pragmatic cohort-randomized controlled trial was designed to compare the effectiveness of an 8-week group weight management program, ShipShape, to a version of ShipShape enhanced with acceptance and commitment therapy.

Objective: The objective of our study was to assess potential differences between in-person and virtual participation in participants' demographics, motivation, confidence, credibility, expectations, and satisfaction with the interventions; we also examined the pragmatics of the technology and participants' experiences in virtual-format intervention groups.

Methods: A total of 178 active-duty personnel who had failed or were at risk of failing their physical fitness assessment or were overweight or obese were enrolled in the study. In-person (n=149) and virtual (n=29) participants reported demographics, motivation, confidence, credibility, expectations, and satisfaction. Interventionists recorded attendance and participation in the group sessions. Independent-sample 2-tailed *t* tests and chi-square tests were used to compare the characteristics of the in-person and virtual participants. Pragmatics of the technology and participants' experiences in the virtual format were assessed through surveys and open-ended questions.

Results: Participants were 29.7 (SD 6.9) years old on average, 61.8% (110/178) female, and 59.6% (106/178) White and had an average BMI of 33.1 (SD 3.9) kg/m^2. Participants were highly motivated to participate and confident in their ability to complete a weight management program. A total of 82.6% (147/178) of all participants attended 5 of the 8 sessions, and participation was rated as "excellent" by interventionists in both formats. The interventions were found to be credible and to have adequate expectations for effectiveness and high satisfaction in both formats. There were no differences between in-person and virtual participants in any of these metrics, other than interventionist-rated participation, for which virtual participants had significantly higher ratings (P<.001). Technical satisfaction with the virtual sessions was rated as "good" to "very good," and participants were satisfied with the content of the virtual sessions. A word cloud of responses identified "mindfulness," "helpful," "different," "food," "binder," and "class" as concepts the virtual participants found most useful about the program.

Conclusions: Modifications made in response to the COVID-19 pandemic were successful, given the recruitment of active-duty personnel with similar demographic characteristics, attendance levels, and indicators of credibility, expectancy, and satisfaction

XSL•FO
RenderX

in the virtual format and the in-person format. This successful transition provides support for the use of virtual or digital weight management interventions to increase accessibility and reach among highly mobile active-duty personnel.

Trial Registration: ClinicalTrials.gov NCT03029507; https://clinicaltrials.gov/ct2/show/NCT03029507

(J Med Internet Res 2022;24(11):e37797) doi:10.2196/37797

KEYWORDS

virtualization; weight-loss intervention; lessons learned; military personnel; acceptance and commitment therapy

Introduction

To slow the spread of COVID-19, the World Health Organization declared a global pandemic and urged every country to take immediate and aggressive action on March 11, 2020 [1]. Stay-at-home orders were announced on March 19, 2020, that mandated all California residents to stay in their homes unless they needed to leave for essential work or activities. Like many other institutions, in-person research activities were halted, including our pragmatic randomized controlled trial (RCT) of a weight management program in active-duty military personnel that was funded by the National Institutes of Health (NIH) [2]. The RCT was designed to compare Navy's ShipShape (SS) weight management program to an acceptance and commitment therapy (ACT)-enhanced SS program. Over 8 months, our research team transitioned the protocols and materials for this RCT and the intervention groups from an in-person to a virtual format. This paper describes and discusses this transition and identifies lessons learned to make recommendations relevant to RCTs and interventions with active-duty personnel.

The US Navy implemented the evidence-based SS program to address weight management and fitness among active-duty personnel who had failed or were at risk of failing the Physical Fitness Assessment (PFA) or were overweight or obese. Given the growing research that supports incorporating ACT—a transdiagnostic cognitive behavioral therapy designed to increase psychological flexibility [3]—into weight loss interventions [4-6], we obtained NIH funding in 2016 to conduct such a study with active-duty personnel. The primary aim of this RCT was to compare the effectiveness of the standard SS program and an ACT plus SS program on weight loss as a primary outcome and to examine several secondary outcomes among active-duty personnel at the Naval Medical Center San Diego (NMCSD), a large military treatment facility in southern California. Due to the pragmatic nature of the RCT, which required conducting the study in a real-world setting, the RCT was implemented in the NMCSD Health and Wellness Department. Study and clinic staff worked together collaboratively to carry out the study procedures and deliver the SS and ACT plus SS group interventions.

The pandemic-related restrictions initiated in March 2020 led to the temporary suspension of NMCSD Health and Wellness Department in-person activities, including the study groups. Soon thereafter, the Navy's PFA requirements were suspended until at least June 2021, and NMCSD made the decision not to offer the SS program until then. The study team, in consultation with the NIH program officer and the study safety officer, determined that the study would continue through community recruitment and virtual study and intervention procedures. This decision was based on two rationales: (1) we anticipated that active-duty personnel, like many others, would continue to struggle with weight management through the pandemic, and (2) although there is growing evidence that technology-supported ACT can be effective in addressing health issues [7], including weight loss [8], we were not aware of any studies that had examined group-based virtual interventions for weight management among active-duty personnel. Further, examining the feasibility and successful implementation of a virtual study of active-duty personnel is particularly important given the limited number of RCTs of behavioral interventions with military personnel [9] and the well-known challenges of recruiting and retaining active-duty personnel for in-person studies [10].

This paper describes a series of modifications that were made to the study procedures and intervention protocols to transition the study from an in-person to a virtual format in the context of the COVID-19 pandemic. Additionally, we (1) assessed potential differences between in-person and virtual participants in demographics, motivation, confidence, credibility, expectations, and satisfaction with the interventions; and (2) examined the pragmatics of the technology and participants' experiences in virtual-format intervention groups.

Methods

Original Study Overview

Background

The design, aims, procedures, and interventions of the original pragmatic cohort RCT trial have been described in detail elsewhere [2]. In brief, eligible participants included active-duty personnel between the ages of 18 to 69 years who had failed or were at risk of failing the PFA or were overweight or obese, were free of physical limitations that would prevent participation in physical fitness activities, and were not pregnant or planning to become pregnant in the next 6 months. Before the pandemic, in-person recruitment occurred primarily through referrals from the Navy's Fitness Enhancement Program with the option to enroll in SS. Study staff met with interested participants after a Navy-facilitated orientation, briefly described the study, and asked them to complete a brief screening assessment. Eligible participants provided in-person consent after the orientation. If they consented, they were enrolled in the study and randomized to receive either an SS or ACT plus SS group intervention. A computer-generated randomization schedule was developed prior to the start of the data collection by the study statistician, and each cohort was randomly assigned to SS or ACT plus SS with equal probability.

XSL•FO
RenderX

Participants completed a battery of standardized, self-reported psychosocial measures and assessments of weight (as the primary outcome) and body fat percentage at week 1 (baseline), week 4 (midpoint), week 8 (posttreatment), a 3-month follow-up, and a 6-month follow-up. Baseline, midpoint, and posttreatment assessments were completed primarily in person, and measurements were taken by the SS coordinator or another member of the NMCSD Health and Wellness Department. The 3- and 6-month follow-up assessments were either scheduled to be in person or were completed remotely by mail or telephone, depending on the availability of the participant.

SS Intervention

The SS intervention was standard care at NMCSD and was based on guidelines from the Navy and Marine Corps Public Health Center [11]. Intervention topics included education about nutrition and physical activity; setting weight-loss goals; tracking eating habits; and initiating, monitoring, and maintaining physical activity levels. The group sessions were primarily led by the NMCSD SS coordinator (with 2 sessions led by an NMCSD social worker) and were held for 2 hours weekly over 8 weeks.

ACT Plus SS Intervention

The study team developed the ACT plus SS intervention by incorporating ACT principles and methods into the standard SS protocol. These concepts and strategies included (1) identifying thoughts, feelings, and bodily sensations surrounding weight loss efforts and overcoming limitations of previous efforts to control or eliminate negative thoughts or emotions, stress, or food cravings; (2) clarifying personal values that aligned with goals to improve health and quality of life; and (3) incorporating mindfulness exercises to increase present-moment awareness. Sessions were co-led by the NMCSD SS coordinator, who provided SS materials, and a study interventionist, who provided ACT content and experiential exercises. Sessions were held for 2 hours weekly over 8 weeks.

Study Transition to Virtual Format

Outline

Table 1 provides an outline of modifications made to recruitment, screening, consent, assessment, and intervention procedures.

Table 1. Transition from in-person to virtual recruitment, screening, consent, assessment, and intervention.

Stage	In-person format	Virtual format
Recruitment	Referrals primarily from the Navy's Fitness Enhancement Program. Recruits attended a ShipShape program orientation, and study staff explained the study.	The study was advertised through social media platforms, including Facebook, Instagram, and Twitter. Individuals in the San Diego community responded to the advertisements.
Screening	Study staff met with interested participants after a Navy-facilitated health and wellness orientation. Potential participants were screened immediately after orientation.	An online screening form asked potential participants questions about their active-duty status, overweight or obesity status, first name, phone number, and best time to call. Study staff conducted additional phone screening to determine the eligibility of interested participants, and then scheduled a virtual consent appointment.
Consent	Study staff reviewed the study procedures (including compensation, participation commitment, contacts, and study timelines) and provided the study consent form to interested and eligible participants in person. Interested and eligible participants provided in-person signed consent after the orientation.	Study staff reviewed the study procedures (including compensation, participation commitment, contacts, and study timelines) and provided the study consent form to interested and eligible participants during a virtual consent session. Interested and eligible participants provided virtual signed consent using approved telehealth technology. A study-materials box was mailed to participants ahead of the virtual orientation session, and all study procedures were reviewed in detail.
Assessment	Medical history and comprehensive demographic information were collected at the consent stage. The ShipShape coordinator or another Health and Wellness Department staff member weighed and measured the participants at the end of baseline, midpoint, and posttreatment sessions, as well as at the 3- and 6-month follow-up visits (if in person). Study staff administered assessment packets at the same time points.	Medical history and comprehensive demographic information were collected as part of the baseline packet. Participants were extensively trained on how to weigh and measure themselves and when to report that information. Study staff collected self-measured weight and other measurements through a Webex (Cisco Inc) private chat at relevant timepoints. Participants brought relevant assessment packets to the virtual intervention sessions, and the 3- and 6-month follow-up packets were completed by mail or telephone. Three brief questions to assess virtual audiovisual quality were asked through Webex polls at the end of each virtual intervention session. The qualitative and quantitative virtual format questionnaire assessed the participants' experiences with the virtual format.
Intervention	Participants received a binder with an intervention workbook during the first session. Intervention group sizes ranged from 6 to 13 participants. The ShipShape coordinator played recordings of motivational and informational content relevant to ShipShape. ShipShape materials and acceptance and commitment therapy activities were provided live and in person.	Participants received a binder through the mail prior to the start of the study. Intervention group sizes ranged from 4 to 9 participants. Recordings of the ShipShape coordinator and study interventionists were made for the ShipShape and acceptance and commitment therapy content, respectively, and were played during Webex sessions, based on the study arm. Study interventionists facilitated virtual sessions through a combination of live and recorded materials presented during Webex sessions.

XSL•FO
RenderX

Recruitment and Screening

Once in-person recruitment was no longer possible, recruitment for participants in the virtual format occurred through the community and by self-referral in response to advertisements within an 80.5-kilometer (50-mile) radius of San Diego. We contracted with BuildClinical LLC, a marketing company that specializes in advertisements for clinical studies through social media. An institutional review board-approved advertisement was posted on various social media platforms (eg, Facebook and Instagram). The advertisement provided brief information on the study and a link to ClinicalTrials.gov. A brief online screening form asked potential participants about their active-duty status, overweight or obesity status, first name, phone number, and the best time for them to receive a phone call. Study staff then contacted potential participants by phone to further determine eligibility.

Consent and Assessment

After screened participants were found to be eligible and interested, they scheduled virtual consent and study preparation sessions to review the study in detail, provide signed consent, and review the mailed study materials. Virtual consent and study preparation sessions were held either by phone or Webex (Cisco Inc), and signed consent was obtained using secure telehealth software approved by Veterans Affairs (VA).

The mailed study materials included a binder with comprehensive instructions on all study procedures, which were all completed remotely, and the relevant intervention binder, assessment packets, a digital scale, and other study materials (eg, a tape measure). Study staff provided extensive training to the participants on all at-home measurements, including how to weigh themselves and perform other measurements, prior to virtual intervention sessions 1, 4, and 8. The data were collected privately at the beginning of those sessions. Participants were also instructed to complete their relevant timepoint assessment packets at the beginning (baseline) or end (midpoint and posttreatment) of the intervention sessions. Study staff collected 3- and 6-month follow-up assessments through the mail or by telephone using a protocol that had been used with some of the prepandemic participants.

Intervention Transition to Virtual Format

Virtual Intervention Delivery Protocol

The NMCSD SS coordinator who had primarily led the SS sessions was reassigned to COVID-related duties. Thus, we had to make decisions about how to deliver the intervention materials virtually. We considered providing the materials asynchronously (without live discussion with a facilitator or group of participants) for participants to study at their own pace, allowing more flexibility for participants and a lower burden on facilitators. After discussions with stakeholders, we decided to remain as consistent as possible with the in-person format in terms of the facilitators. We chose to maintain a group format, which allows for the review of materials with a facilitator and opportunities for participants to share experiences and receive support from others. We opted to prepare video recordings of the SS coordinator presenting most of the SS content and a study interventionist presenting the ACT-related content. The virtual group sessions were then held live, with participants viewing the recordings interspersed with live discussion with study interventionists. To maintain live discussion of the SS material, one interventionist acted as the SS facilitator for the nonrecorded content in both the SS and ACT plus SS groups, while another interventionist facilitated the nonrecorded ACT content in the ACT plus SS group. All decisions were guided by the study's principal investigator.

Video recordings of the SS coordinator (for SS materials) and a study interventionist (for ACT materials) were made with Zoom (Zoom Inc). Recordings were then reviewed, edited, and organized by study staff. Study staff were also present at each session to manage the virtual technology and play the video recordings while the interventionist was focused on delivery of the intervention and attending to group processes. All recordings were reviewed for adherence, competency, content, and logistics by an ACT-trained clinical psychologist. As with the in-person sessions, 20% of the virtual sessions were selected at random and were attended by 2 members of the study team to assign adherence and competency ratings for both the SS and ACT plus SS live interventionists.

Session Structure

As with the in-person sessions, participants were required to attend sessions at designated times in the virtual format. Participants received instructions and a list of rules relevant to the virtual format (eg, logging in on time and keeping the camera on at all times). Given that video recordings are inherently less interactive than in-person interactions, the SS and ACT video recordings were followed by interventionist-led live content and discussions. To best support individual behavioral change, some content, such as monitoring hunger level and selecting a personal weekly action plan, was live. Interventionists asked questions about each participant and included interactive content approximately every 30 minutes to encourage engagement. Approximately the same amount of time was spent delivering recorded and live content and group discussions in both conditions.

Session Content

Overall, the content of the SS and ACT plus SS interventions remained the same with the transition to the virtual format. The transition from the in-person to the virtual format for the SS condition did not require many modifications, because there were fewer physically interactive activities embedded in the program. The SS coordinator led the in-person sessions with slide show presentations, relevant videos, and discussions. With the transition to the virtual format, these sessions were recorded, with the videos embedded in the recordings.

The transition from in-person to virtual delivery for the ACT plus SS condition required more modifications, because some of the interactive activities were not feasible in the virtual format. The objective was to make modifications to the ACT exercises and metaphors while still targeting the same ACT processes. For example, "tug of war" is an experiential activity that requires a physical prop and at least 2 individuals to demonstrate the ACT processes of acceptance, defusion, and committed action. The study team determined that it was not

XSL•FO
RenderX

possible to conduct "tug of war" virtually. This exercise was replaced with "unwanted party guest," an activity that includes a video and discussion and demonstrates the ACT processes in a similar way. Another example of an activity modification was "cravings and trigger foods." The study staff brought participant-identified trigger foods to the in-person sessions. For the virtual sessions, participants were instructed to provide their own trigger foods or an image of the food.

Measures

Demographics

Self-reported sociodemographic information was collected at either screening or baseline, including age (in years), sex (male or female), race (American Indian or Alaskan Native; Asian or Pacific Islander; Black or African American; or White, not of Hispanic origin), ethnicity (Hispanic or Latino), paygrade (enlisted or officer), living status (off base or on base), and relationship status (married, partnered, or in a significant relationship).

Anthropometrics

In-person measurements were collected by the NMCSD SS coordinator or another member of NMCSD. Weight and height were measured objectively using a stadiometer (Health O Meter 500KL; Pelstar, LLC). BMI was calculated as weight in kilograms divided by height in meters squared. Measurements were made of neck and waist circumference for males, and neck, waist, and hip circumference for females. Study staff calculated body fat percentage with a health and fitness Navy PFA app (Vandersoft, Navy PFA). Virtual participants weighed themselves using a study-provided calibrated digital scale (Body Smart Weight and BMI Digital Scale, Withings SA) and reported their weight and self-reported height to study staff, who calculated BMI. Self-measured neck and waist circumference for males and neck, waist, and hip circumference for females were used by study staff to calculate body fat percentage using the Navy PFA app.

Motivation and Confidence

Participants answered the following questions either at screening or baseline: "How motivated are you to participate in a structured weight management/fitness program?" and "How confident are you to complete a structured weight management/fitness program?" Responses to both questions ranged from not at all (0) to extremely (10). Total sum motivation and confidence scores were calculated.

Attendance and Participation

Attendance at each of the 8 group sessions was recorded. Participants who completed a minimum of 5 of 8 sessions were predefined as "completers" based on Health and Wellness Department practice. For those who were present at each session, interventionists scored their level of participation as poor (1), adequate (2), or excellent (3). An overall average participation score was calculated.

Credibility and Expectations

The 6-item Credibility and Expectations for Improvements (CEI) scale was administered to assess how logical the intervention seemed and how much the participants expected to benefit [12]. Individual items on this measure were transformed to standardized scores (*z* scores) and summed into 2 subscales: credibility and expectancy. Higher scores indicated greater credibility and expectancy.

Satisfaction

The 8-item Client Satisfaction Questionnaire (CSQ) was used to measure satisfaction with the interventions [13]. Responses were summed to calculate a total satisfaction score ranging from 0 to 32, with higher scores indicating greater satisfaction.

Pragmatics of Technology

To determine if the virtual sessions had adequate audio and video quality, virtual-format participants responded to 3 questions through Webex polls at the end of each virtual session. Questions were asked about audio quality, video quality, and overall satisfaction with the technical aspects of the session, with ratings ranging from very poor (1) to very good (5). An average score was calculated for each item across all participants and visits.

Evaluation of Virtual Format

Virtual participants completed an investigator-created virtual format questionnaire (VFQ) with both open-ended and multiple-choice items to examine their perceptions of virtual-format effectiveness after the completion of the intervention. Participants responded to the open-ended question "Overall, what did you find most useful in the course?" and a word cloud map was used to assess the most common words from participants' responses to this question (Multimedia Appendix 1). Five Likert-type items asked about satisfaction with the virtual format, with responses ranging from strongly dissatisfied (1) to strongly satisfied (5). An additional 4 items asked about the timing and frequency of sessions to assist with the design of future virtual interventions.

Statistical Analyses

Descriptive statistics were computed to characterize the samples. Data distributions were examined, and independent sample *t* tests (for continuous outcomes) and chi-square tests (for categorical outcomes) were used to compare the characteristics of in-person to virtual participants. Descriptive statistics were used to assess the participants' experience in the virtual format. A power analysis was not performed, as this was a secondary analysis of already collected data. All statistical tests were 2-tailed. The significance level was set at $P<.05$. Data analysis was performed in SPSS (version 28; IBM, Inc). Qualitative analysis of the open-ended VQF question was conducted using Atlas.ti Word Cloud (version 9, Atlas.ti Inc).

Ethics Approval

All methods and technology for the virtual format were approved by the Veterans Affairs San Diego Healthcare Systems institutional review board and research and development committee (H160150) and were Health Insurance Portability and Accountability Act–compliant. The original trial was registered on ClinicalTrials.gov (NCT03029507).

XSL•FO
RenderX

Results

Before the pandemic, a total of 248 potential participants were screened through the NMCSD, of whom 154 (62.1%) were enrolled and randomized to in-person groups. During the pandemic, there were 35 potential participants identified through BuildClinical, of whom 23 (65.7%) were enrolled and randomized to virtual groups. Five participants who had consented through the NMCSD immediately prior to the pandemic restrictions were also included in the virtual groups. The average group size was 10 (SD 2.4; range 6-13) for the in-person groups and 6 (SD 2.2; range 4-9) for the virtual groups.

Overall, there were 178 participants, who were on average 29.7 (SD 6.9) years old, 61.8% (110/178) female, 59.6% (106/178) White, and 28.7% (51/178) Hispanic or Latino. Most participants were enlisted (166/178, 93.3%), lived off base (133/178, 74.7%), and were married, partnered, or reported being in a significant relationship (126/178, 70.8%) (Table 2). The average weight was 94.7 kg (208.7 lbs) for the participants overall, 105.1 kg (231.8 lbs) for males, and 87.6 kg (193.1 lbs) for females; average BMI was 33.1 kg/m^2 (33.5 kg/m^2 for males and 32.9 kg/m^2 for females), which falls in the obese range. Average body fat percentage was 35.6% (27.2% for males and 40.9% for females), which also indicated obesity for both men and women. There were no significant differences between formats in participant characteristics.

On average, participants were highly motivated (mean score 8.1, SD 1.7) to participate and confident (mean score 7.6, SD 1.9) in their ability to complete a weight management program. Of the 149 participants in the in-person format, 123 (82.6%) completed at least 5 of the 8 sessions; 24 of 29 (82.8%) participants in the virtual format completed at least 5 of 8 sessions. Overall, participation was rated as “excellent” by interventionists in both formats, but participants in the virtual format had significantly higher ratings for participation (P<.001). The intervention was found to be credible and have adequate expectations for effectiveness across all participants, with no significant differences for credibility (P=.1) or expectancy (P=.07) across the 2 formats. The average intervention satisfaction score was high (mean score 29.3, SD 3.2), with no significant differences between formats (P=.82).

Across all virtual sessions, average ratings for audio quality (mean score 4.57, SD 0.74), visual quality (mean score 4.59, SD 0.76), and overall technical satisfaction (mean score 4.62, SD 0.71) were between “good” and “very good.” Sixteen of the 29 virtual participants completed the VFQ (Table 3). All respondents were “somewhat satisfied” or “strongly satisfied” with the video and live portions of the virtual classes. Most participants used the intervention binders “weekly” or “a few times a week” and reported being “always” or “mostly attentive” to the video recordings. Over half of the respondents reported using the skills learned in class “at least a few times a week” to “daily.” The word cloud generated from the responses of 16 participants to the open-ended question “Overall, what did you find most useful in the course?” identified 110 unique words. The most common 6 words reported were “mindfulness,” “helpful,” “different,” “food,” “binder,” and “class,” depicting the concepts participants found the most useful (Multimedia Appendix 1).

XSL•FO
RenderX

Table 2. Characteristics of in-person and virtual participants (N=178).

Characteristics	Total (N=178)	In-person format (N=149)	Virtual format (N=29)
Sociodemographics			
Age, mean (SD) years	29.7 (6.9)	28.7 (6.6)	34.4 (6.6)
Female, n (%)	110 (61.8)	94 (63.1)	16 (55.2)
Race, n (%)			
American Indian or Alaskan Native	8 (4.5)	5 (3.4)	3 (10.3)
Asian or Pacific Islander	15 (8.4)	14 (9.4)	1 (3.4)
Black or African American	44 (24.7)	39 (26.2)	5 (17.2)
White (not of Hispanic origin)	106 (59.6)	86 (57.7)	20 (69)
Hispanic/Latino ethnicity	51 (28.7)	40 (26.8)	11 (37.9)
Pay grade, n (%)			
Enlisted	166 (93.3)	140 (94)	26 (89.7)
Officer	12 (6.7)	9 (6)	3 (10.3)
Living status, n (%)			
Off base	133 (74.7)	108 (72.5)	25 (86.2)
On base	45 (25.3)	41 (27.5)	4 (13.8)
Married, partnered, or in a significant relationship, n (%)	126 (70.8)	106 (71.1)	20 (69)
Navy referral, n (%)	155 (87.1)	149 (100)	6 (20.7)
Health status[a], mean (SD)			
Weight	209.1 (37.0)	207.5 (36.4)	217.6 (39.8)
Body fat percentage	35.6 (8.9)	35.3 (8.6)	37.4 (10.1)
BMI	33.2 (3.9)	33.0 (4.0)	34.1 (3.9)
Study experience			
Motivation, mean (SD) score	8.1 (1.7)	8.2 (1.7)	7.8 (1.6)
Confidence, mean (SD) score	7.6 (1.9)	7.7 (1.9)	7.6 (1.7)
Attendance, n (%)	147 (82.6)	123 (82.6)	24 (82.8)
Participation, mean (SD) score	2.6 (0.5)	2.5 (0.5)	2.9 (0.4)[b]
Credibility and Expectations for Improvements Scale, mean (SD) score	–0.12 (4.1)	–0.13 (4.3)	–0.5 (3.5)
Credibility, mean (SD) score	–0.70 (1.9)	–0.13 (1.9)	0.20 (1.6)
Expectancy, mean (SD) score	–0.02 (2.7)	0.01 (2.8)	–0.20 (2.5)
Client Satisfaction Questionnaire, mean (SD) score	29.3 (3.2)	29.6 (3.2)	27.7 (3.0)

[a]N=148 for in-person group.

[b]$P<.001$.

XSL•FO
RenderX

Table 3. Responses to the virtual format questionnaire (N=16).

Questions (range of responses)	Responses				
	1, n (%)	2, n (%)	3, n (%)	4, n (%)	5, n (%)
How satisfied were you with the information provided in the video recordings? (1 to 5: “strongly dissatisfied” to “strongly satisfied”)	0 (0)	0 (0)	0 (0)	10 (63)	6 (38)
How satisfied were you with the live discussion portion of each class? (1 to 5: “strongly dissatisfied” to “strongly satisfied”)	0 (0)	0 (0)	0 (0)	8 (50)	8 (50)
How frequently did you use the binder (packet of written materials) outside of the classroom? (1 to 5: “never” to “daily”)	2 (13)	0 (0)	8 (50)	6 (38)	0 (0)
During the videos, how much of the time were you able to pay attention (not be distracted by family, environment, email, or other things)? (1 to 5: “rarely” to “always”)	0 (0)	1 (6)	0 (0)	11 (69)	4 (25)
If the video recordings were available to you outside of class, how likely would you have been able to watch the videos on your own before class? (1 to 5: “strongly unlikely” to “strongly likely”)	1 (6)	1 (6)	0 (0)	9 (56)	5 (31)
What would be the ideal total amount of time spent watching video recordings during each class? (1 to 5: “10-20 mins,” “20-30 mins,” “30-40 mins,” “40-50 mins,” and “50-60 mins,” respectively)	1 (6)	7 (44)	4 (25)	3 (19)	1 (6)
What would be the ideal class length? (1 to 5: “<30 mins,” “30-60 mins,” “60-90 mins,” “90-120 mins,” and left blank, respectively)	0 (0)	4 (25)	6 (38)	6 (38)	0 (0)
How many classes in a course would be ideal for you? (1 to 5: “1-2 classes,” “3-4 classes,” “4-6 classes,” “6-8 classes,” and “>8 classes,” respectively)	0 (0)	0 (0)	3 (19)	9 (56)	4 (25)
How often have you used the skills learned during this course? (1 to 5: “never” to “daily”)	0 (0)	2 (13)	4 (25)	7 (44)	3 (19)

Discussion

Principal Findings

This paper describes the transition of a weight management RCT among active-duty personnel from an in-person format to a virtual format following the onset of the COVID-19 pandemic. Our study adds to the few existing resources [14-17] on adapting recruitment procedures, protocols for study implementation, and lessons learned for transitioning in-person RCTs to virtual delivery. Major changes from the transition to a virtual format included recruitment, modifications to the study procedures, and intervention delivery. These modifications were deemed successful, as we were able to recruit active-duty personnel with similar demographic characteristics and attendance levels and similar indicators of credibility, expectancy, and satisfaction as the in-person format. Additionally, participants in the virtual format provided positive feedback regarding their experiences and study materials.

Lessons Learned for Recruitment and Logistics

Recruitment and retention of active-duty personnel in interventional or other longitudinal studies is especially challenging for a variety of reasons, such as the need to obtain command approval, restrictions on the use of incentives, transportation difficulties, frequent change in duty station, deployment and military exercises, and retirement or discharge from the service [10]. We found that community-focused advertisements on social media pages (through BuildClinical) were effective for recruitment, screening, and enrollment. Community-based recruitment broadened our efforts to include the region (as opposed to a specific military medical treatment facility) and was more flexible. This strategy allowed us to recruit quickly and at about the same or higher enrollment rate (23/35, 65.7%, vs 154/248, 62.1%) as the in-person recruitment strategy, which is consistent with previous research [18]. Additionally, study procedures were successfully completed remotely. Future research among active-duty personnel may benefit from using a virtual, community-based recruitment approach and remote study procedures to address the mobility and job demands of this population.

Lessons Learned About Adapting Intervention Protocols

Adapting the SS and ACT plus SS intervention protocols from in-person to virtual delivery took approximately 8 months. This development period ensured that we could integrate prerecorded content with live content while targeting key SS and ACT concepts. Some content was more easily translated to presentation via recordings (eg, mindfulness exercises, stress management, healthy eating behavior, and exercise information)

XSL•FO
RenderX

than other content (eg, experiential exercises that relied on live props). This was challenging, and it took some consideration to re-create these experiential experiences virtually while still targeting the key ACT concepts.

In the initial in-person session, clear behavior guidelines were reviewed and provided in writing, including rules such as "Attendance at all 8 sessions is expected." In the virtual format, guidelines regarding virtual group behavior were developed, provided to participants in their binders, and reviewed in the first session and as needed during the group sessions. For example, virtual guidelines included rules such as "Keep your camera on and have sufficient lighting so you can be seen" and "Unless you are speaking, please mute your audio." Setting behavioral guidelines early on and reminding participants of these guidelines were critical to the success of the virtual sessions.

Lessons Learned About Participants' Experiences in a Virtual Class

Other than significantly higher facilitator-rated participation for the virtual participants, participants in the virtual and in-person sessions were similar in demographic characteristics and study satisfaction levels, which is consistent with previous studies [19]. The higher levels of participation might have been due to the facilitators specifically requesting participation in the virtual format. It is also possible that the smaller group size (ranging from 4 to 9 people) of the virtual sessions allowed for more interaction [20]. Overall, the virtual participants reported good audio and visual quality and few technical difficulties, which is contrary to previous research with virtual interventions [21]. They also reported being satisfied with the video and live portions of each session, were attentive to the video recordings, and used the study and intervention materials (ie, the binder) frequently. We found that 100% of participants were "somewhat satisfied" to "strongly satisfied" with the discussion portions, reinforcing the usefulness of brief live content presentations and discussions. Results from the word cloud reflected critical aspects of SS and ACT, including an emphasis on mindfulness, food, and helpful strategies. This suggests that virtual delivery of the intervention materials was effective, and that they resonated with participants. Together, these findings indicate that the virtual transition was successful, and that SS and ACT plus SS can be delivered virtually for greater reach and accessibility. Future research can focus on a large-scale virtual RCT of ACT among active-duty personnel to determine its efficacy for weight management.

Limitations

Although the current study provides valuable insight into how to transition from an in-person RCT to a virtual format, there are a few limitations. First, before the pandemic, active-duty personnel were mandated by the Navy to attend SS, whereas after the pandemic, recruitment was from the community. Thus, it is possible that the virtual participants self-selected to participate in the virtual format instead of doing so based on a Navy mandate. Second, the samples of in-person and virtual participants were unequal, with a much smaller number of participants in the virtual groups; this may also have skewed the findings. Third, data collection strategies were heterogeneous and weight and height assessments were self-measured and self-reported in the virtual format, which may have increased bias [22]. Nevertheless, while we found no significant differences between the groups in demographic or anthropometric characteristics, suggesting that participants in the virtual groups were similar to participants in the in-person groups, our quantitative findings should be interpreted with caution. Additional research with larger samples is needed to determine optimal designs for virtual RCTs of active-duty personnel with overweight or obesity.

Conclusions

During COVID-19, our research team successfully transitioned a weight management RCT with active-duty personnel from an in-person to a virtual format, including aspects of study design such as recruitment, screening, assessments, procedures, and content delivery. The virtual format led to similar levels of satisfaction and attendance as the in-person format. The successful transition of this study to a virtual format provides support for the use of virtual interventions among active-duty personnel. Furthermore, virtual ACT-based weight management and fitness interventions may be a promising approach to increase the accessibility and reach of these interventions among highly mobile populations, such as active-duty personnel.

Acknowledgments

This study was supported by grant R01DK106415 from the National Institute of Diabetes and Digestive and Kidney Diseases. MSH is supported by the Veterans Affairs Rehabilitation Research and Development Service under Career Development Award 1IK2RX002807.

Disclaimer

The views expressed in this paper are solely those of the authors and do not reflect the official policy or position of the funding agency, Department of Veterans Affairs, US Army, US Navy, US Air Force, the Department of Defense, the United States government, or any institutions with which the authors are affiliated.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Word cloud map of the most common words from participants (N=16) in response to the open-ended question “Overall, what did you find most useful in the course?”.
[PNG File , 791 KB - jmir_v24i11e37797_app1.png]

Multimedia Appendix 2
CONSORT-eHEALTH checklist.
[PDF File (Adobe PDF File), 65 KB - jmir_v24i11e37797_app2.pdf]

References

1. Timeline: WHO's COVID-19 response. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline [accessed 2022-10-10]
2. Afari N, Cuneo JG, Herbert M, Miller I, Webb-Murphy J, Delaney E, et al. Design for a cohort-randomized trial of an acceptance and commitment therapy-enhanced weight management and fitness program for Navy personnel. Contemp Clin Trials Commun 2019 Sep;15:100408 [FREE Full text] [doi: 10.1016/j.conctc.2019.100408] [Medline: 31338482]
3. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 2006 Jan;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
4. Forman EM, Butryn ML, Manasse SM, Crosby RD, Goldstein SP, Wyckoff EP, et al. Acceptance-based versus standard behavioral treatment for obesity: Results from the mind your health randomized controlled trial. Obesity (Silver Spring) 2016 Oct;24(10):2050-2056 [FREE Full text] [doi: 10.1002/oby.21601] [Medline: 27670400]
5. Lawlor ER, Islam N, Bates S, Griffin SJ, Hill AJ, Hughes CA, et al. Third-wave cognitive behaviour therapies for weight management: A systematic review and network meta-analysis. Obes Rev 2020 Jul;21(7):e13013 [FREE Full text] [doi: 10.1111/obr.13013] [Medline: 32181957]
6. Manchón J, Quiles M, León E, López-Roig S. Acceptance and Commitment Therapy on physical activity: A systematic review. J Contextual Behav Sci 2020 Jul;17:135-143. [doi: 10.1016/j.jcbs.2020.07.008]
7. Herbert MS, Dochat C, Wooldridge JS, Materna K, Blanco BH, Tynan M, et al. Technology-supported Acceptance and Commitment Therapy for chronic health conditions: A systematic review and meta-analysis. Behav Res Ther 2022 Jan;148:103995. [doi: 10.1016/j.brat.2021.103995] [Medline: 34800873]
8. 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]
9. Cook WA, Doorenbos AZ, Bridges EJ. Trends in research with U.S. Military service member participants: a population-specific ClinicalTrials.gov review. Contemp Clin Trials Commun 2016 Aug 15;3:122-130 [FREE Full text] [doi: 10.1016/j.conctc.2016.04.006] [Medline: 27822569]
10. Cook WA, Doorenbos AZ. Indications of recruitment challenges in research with U.S. Military service members: a ClinicalTrials.gov review. Mil Med 2017 Mar;182(3):e1580-e1587 [FREE Full text] [doi: 10.7205/MILMED-D-16-00225] [Medline: 28290928]
11. ShipShape Participant Resources. Department of the Navy, Bureau of Medicine and Surgery. URL: https://www.med.navy.mil/Navy-Marine-Corps-Public-Health-Center/Population-Health/Health-Promotion-and-Wellness/ShipShape-Participant-Resources/ [accessed 2022-10-10]
12. Borkovec TD, Nau SD. Credibility of analogue therapy rationales. J Behav Ther Exp Psychiatry 1972 Dec;3(4):257-260. [doi: 10.1016/0005-7916(72)90045-6]
13. Larsen DL, Attkisson C, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: Development of a general scale. Eval Program Plann 1979 Jan;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6]
14. Lai B, Chiu C, Pounds E, Tracy T, Mehta T, Young H, et al. COVID-19 modifications for remote teleassessment and teletraining of a complementary alternative medicine intervention for people with multiple sclerosis: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Jul 03;9(7):e18415 [FREE Full text] [doi: 10.2196/18415] [Medline: 32540838]
15. Lalande K, Greenman PS, Bouchard K, Johnson SM, Tulloch H. The healing hearts together randomized controlled trial and the COVID-19 pandemic: a tutorial for transitioning from an in-person to a web-based intervention. J Med Internet Res 2021 Apr 06;23(4):e25502 [FREE Full text] [doi: 10.2196/25502] [Medline: 33729984]
16. Myers B, van der Westhuizen C, Pool M, Hornsby N, Sorsdahl KR. Responding to COVID-19 threats to trial conduct: lessons learned from a feasibility trial of a psychological intervention for South African adolescents. Trials 2021 Jul 09;22(1):440 [FREE Full text] [doi: 10.1186/s13063-021-05400-8] [Medline: 34243806]
17. Tang L, Broad J, Lewis R, Ma DW, Haines J, Guelph Family Health Study. Transitioning a home-based, motivational interviewing intervention among families to remote delivery during the COVID-19 pandemic: Key lessons learned. Patient Educ Couns 2021 Sep;104(9):2286-2291 [FREE Full text] [doi: 10.1016/j.pec.2021.02.043] [Medline: 33674132]

XSL•FO
RenderX

18. Moseson H, Kumar S, Juusola JL. Comparison of study samples recruited with virtual versus traditional recruitment methods. Contemp Clin Trials Commun 2020 Sep;19:100590 [FREE Full text] [doi: 10.1016/j.conctc.2020.100590] [Medline: 32637722]
19. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: a systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]
20. Biggs K, Hind D, Gossage-Worrall R, Sprange K, White D, Wright J, et al. Challenges in the design, planning and implementation of trials evaluating group interventions. Trials 2020 Jan 29;21(1):116 [FREE Full text] [doi: 10.1186/s13063-019-3807-4] [Medline: 31996259]
21. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]
22. Maukonen M, Männistö S, Tolonen H. A comparison of measured versus self-reported anthropometrics for assessing obesity in adults: a literature review. Scand J Public Health 2018 Jul 10;46(5):565-579. [doi: 10.1177/1403494818761971] [Medline: 29528773]

Abbreviations

ACT: acceptance and commitment therapy
CEI: Credibility and Expectations for Improvements
CSQ: Client Satisfaction Questionnaire
NIH: National Institutes of Health
NMCSD: Naval Medical Center San Diego
PFA: Physical Fitness Assessment
RCT: randomized controlled trial
SS: ShipShape
VA: Veterans Affairs
VFQ: virtual format questionnaire

Edited by R Kukafka; submitted 13.03.22; peer-reviewed by R Marshall, K Uludag; comments to author 08.04.22; revised version received 21.04.22; accepted 26.09.22; published 10.11.22.

Please cite as:
Afari N, Yarish NM, Wooldridge JS, Materna K, Hernandez J, Blanco BH, Camodeca AL, Peters JJ, Herbert MS
Lessons Learned From Transition of an In-Person to a Virtual Randomized Controlled Trial for Weight and Fitness Concerns in Active-Duty Service Members: Survey Study
J Med Internet Res 2022;24(11):e37797
URL: https://www.jmir.org/2022/11/e37797
doi:10.2196/37797
PMID:36201851

©Niloofar Afari, Natalie M Yarish, Jennalee S Wooldridge, Karla Materna, Jeffrey Hernandez, Brian H Blanco, Angela L Camodeca, Joshua J Peters, Matthew S Herbert. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Economic and Environmental Impact of Digital Health App Video Consultations in Follow-up Care for Patients in Orthopedic and Trauma Surgery in Germany: Randomized Controlled Trial

Jennifer Muschol[1], MSc; Martin Heinrich[2], Dr med; Christian Heiss[2], Prof Dr med Dr hc; Alher Mauricio Hernandez[3], Prof Dr; Gero Knapp[2], Dr med; Holger Repp[1], Dr med; Henning Schneider[4], Prof Dr med; Ulrich Thormann[2], PD Dr med; Johanna Uhlar[4]; Kai Unzeitig[2], Dr med; Christian Gissel[1], Prof Dr

[1]Department of Health Economics, Justus Liebig University Giessen, Giessen, Germany

[2]Department of Trauma, Hand and Reconstructive Surgery, University Hospital Giessen, Giessen, Germany

[3]Bioinstrumentation and Clinical Engineering Research Group, Bioengineering Department, Engineering Faculty, Universidad de Antioquia, Medellín, Colombia

[4]Institute of Medical Informatics, Justus Liebig University Giessen, Giessen, Germany

Corresponding Author:
Christian Gissel, Prof Dr
Department of Health Economics
Justus Liebig University Giessen
Klinikstrasse 29
Giessen, 35392
Germany
Phone: 49 641 9922070
Email: cg@phec.de

Abstract

Background: Following the Riyadh Declaration, digital health technologies were prioritized in many countries to address the challenges of the COVID-19 pandemic. Digital health apps for telemedicine and video consultations help reduce potential disease spread in routine health care, including follow-up care in orthopedic and trauma surgery. In addition to the satisfaction, efficiency, and safety of telemedicine, its economic and environmental effects are highly relevant to decision makers, particularly for the goal of reaching carbon neutrality of health care systems.

Objective: This study aims to provide the first comprehensive health economic and environmental analysis of video consultations in follow-up care after knee and shoulder interventions in an orthopedic and trauma surgery department of a German university hospital. The analysis is conducted from a societal perspective. We analyze both economic and environmental impacts of video consultations, taking into account the goal of carbon neutrality for the German health care system by 2030.

Methods: We conducted a prospective randomized controlled trial comparing follow-up care with digital health app video consultations (intervention group) to conventional face-to-face consultations in the clinic (control group). Economic impact included the analysis of travel and time costs and production losses. Examination of the environmental impact comprised the emissions of greenhouse gases, carbon monoxide, volatile hydrocarbons, nitrogen oxides, and particulates, and the calculation of environmental costs. Sensitivity analysis included calculations with a higher cost per ton of carbon dioxide equivalent, which gives equal weight to the welfare of present and future generations.

Results: Data from 52 patients indicated that, from the patients' point of view, telemedicine helped reduce travel costs, time costs, and production losses, resulting in mean cost savings of €76.52 per video consultation. In addition, emissions of 11.248 kg of greenhouse gases, 0.070 kg of carbon monoxide, 0.011 kg of volatile hydrocarbons, 0.028 kg of nitrogen oxides, and 0.0004 kg of particulates could be saved per patient through avoided travel. This resulted in savings of environmental costs between €3.73 and €9.53 per patient.

Conclusions: We presented the first comprehensive analysis of economic and environmental effects of telemedicine in the follow-up care of patients in orthopedic and trauma surgery in Germany. Video consultations were found to reduce the environmental footprint of follow-up care; saved travel costs, travel time, and time costs for patients; and helped to lower production losses. Our findings can support the decision-making on the use of digital health during and beyond the COVID-19 pandemic, providing

XSL•FO
RenderX

decision makers with data for both economic and environmental effects. Thanks to the pragmatic design of our study, our findings can be applied to a wide range of clinical contexts and potential digital health applications that substitute outpatient hospital visits with video consultations.

Trial Registration: German Clinical Trials Register DRKS00023445; https://tinyurl.com/4pcvhz4n

(J Med Internet Res 2022;24(11):e42839) doi:10.2196/42839

KEYWORDS

carbon neutrality; digital health; environmental impact; health economics; net-zero; orthopedic; sustainability; telemedicine; trauma surgery; video consultations

Introduction

Medical care does not always require patients' attendance in the hospital [1], mainly because digital health affords physicians and patients the opportunity to have synchronous video consultations online [2]. When used for outpatient follow-up care in orthopedic and trauma surgery, for example, video consultations can relieve patients of any restrictions on their mobility or of the need to travel long distances [3-5]. Patient satisfaction, physician satisfaction, and clinical outcomes often show comparable results between telemedicine and conventional face-to-face (F2F) examinations in the hospital, demonstrating that video consultations can be a safe and efficient alternative for patient care in orthopedic and trauma surgery [6-13].

After the outbreak of the COVID-19 pandemic, the role of digital health has been highlighted by the Riyadh Declaration [14]. Following the global pandemic response, there has been an increasing interest in telemedicine in clinical practice to reduce potential disease spread as well as in science, which is reflected in a growing number of literature reviews [2,15-20]. The number of clinical trials, however, remains limited. In particular, there are only a few health economic analyses of the use of telemedicine in orthopedic and trauma surgery follow-up care [15,21].

In addition to patient satisfaction and quality of care, the societal perspective needs to consider both economic and environmental effects in order to support stakeholders in deciding whether to implement telemedicine in orthopedic and trauma surgery. Following the United Nations Sustainable Development Goals, the 125th German Medical Assembly declared in 2021 that the German health care system should become carbon-neutral by 2030 [22]. One way of meeting this requirement might be the implementation of video consultations to supplement or substitute clinic consultations. Whether this is possible, however, must first be determined by investigations. A positive environmental impact of telemedicine has already been demonstrated in certain cases: for example, in the reduction of carbon monoxide, carbon dioxide, and nitric oxides [23-25]. However, analyses of the environmental impact of video consultations in the field of orthopedic and trauma surgery are limited, and no studies based on German data exist to date.

The aim of this study is to provide the first health economic analysis comparing telemedicine in the follow-up of patients in orthopedic and trauma surgery with knee and shoulder disorders with conventional F2F examinations in the clinic in Germany. The analysis focuses on the societal perspective, considering, on the one hand, the patients' point of view in terms of potential time and cost savings and, on the other hand, the environmental impact regarding potential savings of emissions and environmental costs.

Methods

Study Design

The data used for the health economic analysis were obtained by a prospective randomized controlled trial (RCT) conducted at a single German university hospital—University Hospital Giessen, Department of Trauma, Hand and Reconstructive Surgery, Level-1 trauma center—between September 2020 and April 2021. The RCT was reported according to the Consolidated Standards of Reporting Trials (CONSORT) [26]. Patients in orthopedic and trauma surgery were randomly assigned 1 to 1 to an intervention group or a control group for a single follow-up appointment. The intervention group did not attend a standard outpatient follow-up appointment in the clinic but had a real-time online video consultation with the treating physician instead. The control group, on the other hand, was treated conventionally and received a F2F examination in the clinic. In both the intervention group and the control group, the examinations were performed by the same physicians. The study population had already received conservative or surgical treatment for various knee and shoulder conditions in the clinic.

Ethical Considerations

Patients who were eligible for the study based on the inclusion and exclusion criteria in Textbox 1 were asked either at the clinic or by telephone if they wished to participate in the RCT. After a detailed verbal explanation of the study, including the conduct of a health economic analysis as part of the study, all study participants provided written informed consent. To protect the privacy of participating patients, pseudonymization of the study data took place. Study participants were not compensated for their participation. The local ethics committee of the University of Giessen approved the RCT (AZ 73/20), and the study was registered in the German Clinical Trials Register (DRKS00023445).

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria of the randomized controlled trial.

Inclusion criteria:

- 18 years or older
- Previous outpatient or inpatient stay at the clinic, with an operation or conservative therapy
- Need of a follow-up that does not require more than a visual examination
- Ownership of a computer, laptop, tablet, or smartphone with microphone and camera
- Stable internet connection
- Mental and physical ability to consent and to participate
- Sufficient knowledge of German in order to understand the declaration of consent
- Shoulder International Classification of Diseases, Tenth Revision (ICD-10) codes: M75.1, M75.6, M75.0, Z96.60, M75.4, M19.91, S43.1, S42.20, S42.00, M75.2, M75.3, and S43.0
- Knee ICD-10 codes: S83.53, S83.54, S83.2, S83.0, M22.0, M23.32, M23.35, M17.1, M17.5, M21.16, M21.06, S83.3, S83.44, S83.43, S82.18, S82.0, S72.3, S72.43, M25.56, M76.5, S83.6, S76.1, and S86.8

Exclusion criteria:

- Neurological diseases that preclude the use of digital devices
- Diagnosis of dementia, blindness, or deafness
- Need for presence in the clinic and on-site treatment and diagnostics (ie, imaging, laboratory, stitches, and drainage)
- Appointments where the patient has to be touched and moved by the treating physicians
- Lack of willingness to participate
- Failure to consent

Sample Size and Randomization

The sample size calculation of the underlying RCT was based on an a priori power analysis. As a conservative estimate, we used half of the effect size of 2.19 that was observed for the findings of patient satisfaction with telemedicine in a study by Sharareh and Schwarzkopf [8]. The effect size of 1.095 yielded 19 patients per study arm for a power of 90% in a 2-sided *t* test with a 5% significance level. To increase statistical power and to compensate for potential withdrawals and dropouts, missing responses, and a skewed distribution of results, the number of participants was expanded to 30 patients for each group. In total, 60 eligible patients were recruited for the study.

Using block randomization with randomly varying block sizes (ie, 4, 6, and 8), 30 patients were assigned to a follow-up with telemedicine (intervention group), and 30 patients were assigned to a conventional F2F follow-up in the clinic (control group). The parallel-design randomization and assignment process was performed independently of the treating physicians by study staff using sealed envelopes.

Course of the Study

The video consultations in the intervention group were browser based for physicians and multiplatform for patients, including a digital health app or browser-based software from a German telemedicine provider. The software complies with the legal requirements in Germany and is recognized by the National Association of Statutory Health Insurance Physicians. The university hospital paid a monthly fee for each physician to use the software. Video consultation procedures were deliberately kept as simple and as functional as possible to ensure that they would be viable in regular clinical practice: all video consultations were performed directly between the physicians in the clinic and the patients, regardless of their location. No other medical providers, such as local caregivers or others, were involved. Patients received written instructions on how to conduct the video consultation, and no additional clinical staff were required to assist the patients. This pragmatic study design appeared to be the most promising one for a health economic evaluation seeking to produce valid, generalizable results [27]. Patients in the intervention group did not have to bear any additional costs or out-of-pocket payments for using telemedicine, as the digital health app or browser-based software was free for them to use. They were only required to have a smartphone, tablet, laptop, or computer with a microphone and camera, and an adequate internet connection. The examination itself was paid for by their respective health insurance. Patients in the control group did not have to pay any additional costs either; their costs for an in-clinic follow-up appointment (eg, travel costs) were the same as those they would have paid outside of study participation.

After the follow-up appointments, patients in both the intervention and control groups completed questionnaires. These questionnaires included questions about the distance between the patients' homes and the clinic, the amount of time spent for the appointment (eg, travel and waiting time), and the potential need to be absent from work to attend the appointment. Further information on the study can be found in a previous publication by Muschol et al [13].

XSL•FO
RenderX

Statistical Analysis and Health Economic Evaluation

The RCT data are presented as mean and SD, median and IQR, or percentage. To compare the intervention and control groups, the Mann-Whitney *U* test was used for continuous variables and the Fisher exact test was used for categorical ones. Statistical significance was assumed at $P \leq .05$.

The health economic analysis was based on data collected from the questionnaires and other official, external data. The study design was guided by recommendations for health economic analyses in the context of eHealth interventions, and the study examined non–health care costs associated with the use of telemedicine from a societal perspective [27,28]. The analysis proceeded in two steps. In the first step, economic effects of the societal perspective were examined from the patients' point of view. This involved, firstly, calculating and comparing three types of non–health care costs associated with medical appointments:

1. Travel costs were calculated following recommendations for empirical standard costs for health economic evaluations in Germany [29].
2. Time costs were assessed by assigning monetary values to patients' travel time, waiting time, and total time spent on appointments based on Verbooy et al's [30] valuation approach to unpaid work and leisure time.
3. Production losses due to patients' absence from work while attending their appointments were computed using Germany's average gross hourly wage in 2021 and average working hours for all German full-time and part-time employees in 2019 [31,32].

When tallying total costs from a societal perspective, it was felt to be appropriate to differentiate between patients who were employed and patients who were not employed, given that production losses are only relevant for patients who are employed.

In the second step, the effects of the societal perspective were evaluated in the form of the environmental impact of telemedicine. The analysis of the environmental impact was conducted using data from the German Federal Environment Agency. It comprised three different aspects. First, the environmental impact in terms of greenhouse gases, carbon monoxide, volatile hydrocarbons, nitrogen oxides, and particulates was calculated by multiplying the average emissions per passenger-kilometer (pkm) by the kilometers patients traveled by car to and from the clinic. This calculation was based on an average car occupancy of 1.4 passengers, as the average emissions are specified by the Federal Environment Agency on the basis of this value [33]. A separate calculation of emissions from public transportation was not performed within the study because only 1 patient in the control group and 1 patient in the telemedicine group used or would have used public transportation. Second, the average environmental costs incurred per pkm by the patients' trips per car were calculated. For this purpose, the cost rate of the Federal Environment Agency of €195 per ton of carbon dioxide equivalent was applied (a currency exchange rate of €1=US $0.97 is applicable) [34,35]. This value is based on a higher weighting of the welfare of current versus future generations [35].

In a third step, the potential savings in emissions and environmental costs were estimated in a model calculation if 8 patients per week would conduct a video consultation instead of a clinic consultation, as was the case in our study [33-35].

Sensitivity Analysis

Finally, a sensitivity analysis was performed to evaluate the robustness of the findings. For the patients' point of view in the societal perspective, this analysis studied the effect of differentiating between full-time and part-time employment when calculating production losses [32]. For the environmental impact of the societal perspective, the sensitivity analysis considered the following:

1. A cost rate from the Federal Environment Agency for the calculation of the environmental costs of €680 per ton of carbon dioxide equivalent, which gives equal weight to the welfare of present and future generations [34,35].
2. A total of 16 patients with a video consultation per week for the analysis of potential savings in emissions and environmental costs [33-35].

For the calculation of the environmental costs, both €195 and €680 per ton of carbon dioxide equivalent were considered [34,35]. As the Federal Environment Agency reports both cost rates, the aim of the sensitivity analysis was to show how the equal weighting of the welfare of present and future generations (€680) compared to the higher weighting of the welfare of present versus future generations (€195) affects the environmental costs.

Results

General Findings

Of the 60 patients recruited—intervention group (n=30) and control group (n=30)—4 patients in each of the groups withdrew from the study. Thus, data from a total of 52 patients could be considered for the health economic evaluation, with several variables displaying a lower n value due to missing items on some patient questionnaires. The progress of the recruited patients through the trial is shown in a CONSORT flow diagram in Multimedia Appendix 1.

Demographic patient characteristics are shown in Table 1. No significant differences were observed between the telemedicine group and the control group.

Regarding the variables used for calculating costs, however, the differences between the groups were partially significant, as shown in Table 2. Treatment duration in the intervention group, at 8.23 minutes on average, was significantly shorter than that in the control group, at 10.92 minutes on average (P=.02). The average waiting time in the online waiting room for the telemedicine software was also significantly shorter than that experienced in the clinic (6.73 minutes vs 36.88 minutes, respectively; P<.001). The largest intergroup difference, however, was observed in total patient time spent per follow-up appointment. An appointment in the telemedicine group took an average of 21.92 minutes out of the patients' days, whereas an appointment in the control group required patients to spend 154.80 minutes on average (P<.001). There was no significant

XSL•FO
RenderX

difference between the potential travel distance and travel time the telemedicine group would have faced if required to travel to an in-clinic appointment and the actual travel distance and travel time faced by the control group. The groups also did not differ significantly in patients' absence from work due to their appointments. Nevertheless, of the employed patients, only 5% (1/20) were absent from work so they could attend the appointment in the telemedicine group, compared with 16% (3/19) in the control group, as shown with the Fisher exact test (P=.34). In the telemedicine group, 1 patient had to visit the clinic again for further treatment. As this would also have been required after an F2F consultation and, therefore, occurred independently of the video consultation, this additional visit was not included in the cost calculation.

Table 1. Demographic characteristics of patients.

Characteristics	Telemedicine group (n=26), n (%)	Control group (n=26), n (%)	*P* value[a]
Medical indication			.99
Knee	10 (38)	9 (35)	
Shoulder	16 (62)	17 (65)	
Age (years)			.36
18-40	7 (27)	5 (19)	
41-60	17 (65)	15 (58)	
>60	2 (8)	6 (23)	
Female	11 (42)	10 (38)	.99
Employed	20 (77)	19 (76)[b]	.99

[a]*P* values were based on the Fisher exact test.

[b]Percentage of n=25 due to missing item on questionnaire.

Table 2. Variables included for cost calculation.

Variables	Telemedicine group (n=26)			Control group (n=26)			*P* value[a]
	Participants, n (%)	Mean (SD)	Median (IQR)	Participants, n (%)	Mean (SD)	Median (IQR)	
Treatment duration (minutes)	26 (100)	8.23 (4.45)	6.00 (5-10)	25 (96)	10.92 (5.58)	10.00 (8-14.5)	.02
Travel distance (kilometers)	26 (100)	37.00 (32.06)	30.00 (10-46.25)	25 (96)	31.58 (22.62)	28.00 (15.5-45)	.65
Actual and potential travel time (minutes)	26 (100)	38.46 (21.72)	40.00 (18.75-46.25)	25 (96)	34.80 (20.89)	30.00 (20-40)	.42
Waiting time (minutes)	26 (100)	6.73 (6.84)	5.00 (1.75-10)	24 (92)	36.88 (27.54)	30.00 (15-48.75)	<.001
Total time spent on appointment (minutes)	26 (100)	21.92 (10.40)	22.50 (13.75-30)	25 (96)	154.80 (79.75)	150.00 (105-197.5)	<.001

[a]*P* values were based on the Mann-Whitney *U* test.

Patients' Perspectives

The cost calculation from the patients' point of view in the societal perspective showed that patients in the control group had to pay an average of €18.95 in travel costs, based on a cost of €0.30 for each kilometer travelled to and from the clinic, as shown in Table 3. There were no travel costs for patients in the telemedicine group because they did not have to attend the clinic. If they had had an in-clinic follow-up, however, their average travel costs would have been €22.20.

The time costs resulting from follow-up appointments in both groups were estimated at €16.00 per hour to account for both unpaid work time and leisure time that patients lost. The average cost of patients' travel time was €18.56 in the control group. Again, patients in the telemedicine group faced no travel time costs due to the trip they avoided. Yet, the potential cost of their travel time would have been €20.51. The increased waiting time in the clinic was reflected in time costs of €9.83 in the control group, compared with €1.79 in the intervention group.

The difference in time costs between the groups became even more pronounced when the total time patients spent on their follow-up appointments was valued. Whereas patients with a telemedical appointment had average total time costs of €5.85, those with an in-clinic appointment had total time costs of €41.28. In other words, a telemedical rather than an in-clinic follow-up appointment would have saved patients €35.43 in average time costs.

XSL•FO
RenderX

Finally, the production loss due to patients' absence from work while they were attending their appointments was calculated. This was based on an average hourly wage of €29.48 in Germany and an overall average of 6.96 working hours per day per full-time or part-time German employee. With 1 patient absent in the telemedicine group and 3 patients absent in the control group, total production losses were €205.18 and €615.54, respectively. With 20 employed patients in the telemedicine group and 19 employed patients in the control group, the costs due to lost production averaged €10.26 for a telemedical follow-up and €32.40 for an in-clinic one.

Taking employment status into account, the total costs of a follow-up appointment were €16.11 for an employed patient in the telemedicine group and €92.63 for an employed patient in the control group. For an unemployed patient, the total costs decreased to €5.85 in the telemedicine group and to €60.23 in the control group due to the irrelevant production loss. Multimedia Appendix 2 presents the cost calculations in detail.

Table 3. Cost calculation from the patients' perspective.

Costs	Telemedicine group	Control group	Difference
Travel costs (€[a]), mean (SD)	0 (0)	18.95 (13.57)	18.95
Travel time costs (€), mean (SD)	0 (0)	18.56 (11.14)	18.56
Waiting time costs (€), mean (SD)	1.79 (1.82)	9.83 (7.34)	8.04
Total time costs (€), mean (SD)	5.85 (2.77)	41.28 (21.27)	35.43
Production loss (€)	205.18	615.54	410.36

[a]A currency exchange rate of €1=US $0.97 is applicable.

Environmental Impact

To calculate the emissions saved in the telemedicine group due to the avoided trips to and from the clinic, 152 g/pkm for greenhouse gases, 0.94 g/pkm for carbon monoxide, 0.15 g/pkm for volatile hydrocarbons, 0.38 g/pkm for nitrogen oxides, and 0.006 g/pkm for particulates were applied based on an average car occupancy of 1.4 passengers. This led to the result that around 11.248 kg of greenhouse gases, 0.070 kg of carbon monoxide, 0.011 kg of volatile hydrocarbons, 0.028 kg of nitrogen oxides, and 0.0004 kg of particulates were saved per patient with the help of video consultations. Table 4 also shows the total emissions saved for the 26 patients in the telemedicine group. For example, as a result of the video consultations, emissions of 292.448 kg of greenhouse gases could be avoided in our study. The calculation of environmental costs saved in the telemedicine group is based on environmental costs of €0.05045 per pkm. This value represents the average environmental costs of gasoline and diesel powered cars. The use of telemedicine saved approximately €3.73 in environmental costs per patient, resulting in a total of €97.07 for all patients in our study. Finally, the potential savings can also be seen in the model calculation for 1 year if 8 patients per week had a video consultation instead of a clinic consultation, as was the case in our study. For this calculation, the average distance between the home of the patients in the telemedicine group and control group and the clinic was used. With a total of 384 patients who would not have to travel to the clinic each year due to video consultations, a total of 4009.88 kg of greenhouse gases, 24.80 kg of carbon monoxide, 3.96 kg of volatile hydrocarbons, 10.02 kg of nitrogen oxides, and 0.16 kg of particulates could be avoided. In addition, at €195 per ton of carbon dioxide equivalent, €1330.91 in environmental costs could be saved.

Table 4. Saved emissions and environmental costs in the telemedicine group.

Emissions and costs	Per patient	Total
Greenhouse gases (kg)	11.248	292.448
Carbon monoxide (kg)	0.070	1.809
Volatile hydrocarbons (kg)	0.011	0.289
Nitrogen oxides (kg)	0.028	0.731
Particulates (kg)	0.0004	0.012
Environmental costs (€[a])	3.73	97.07

[a]A currency exchange rate of €1=US $0.97 is applicable.

Sensitivity Analysis

In the subsequent sensitivity analysis, several adjustments were made. First, the cost calculation from the patients' point of view was modified to test the effect of alternative assumptions on the valuation of production losses. Assuming that all patients who were absent from work were employed full time (ie, 8.2 hours per day), the societal cost of lost production would have increased to €241.74 (mean €12.09, SD 54.05) in the telemedicine group and to €725.21 (mean €38.17, SD 90.56) in the control group. In contrast, assuming only part-time employment of 3.9 hours per day for all patients who were absent from work, the costs of lost production would have decreased to €114.97 (mean €5.75, SD 25.71) in the

XSL•FO
RenderX

telemedicine group and to €344.92 (mean €18.15, SD 43.07) in the control group. These assumptions would have changed the total costs for employed patients to €17.94 for full-time employees and €11.60 for part-time employees in the telemedicine group, as well as to €98.40 for full-time employees and €78.38 for part-time employees in the control group.

Second, the calculation of environmental costs was adjusted to the cost rate of €680 per ton of carbon dioxide equivalent, which increased the average environmental costs of gasoline and diesel cars to €0.12885 per pkm. Due to this adjustment, the environmental costs saved in the telemedicine group would have been €9.53 per patient and €247.91 in total.

In addition, if a total of 16 patients per week had a video consultation instead of a clinic consultation, approximate emissions of 8019.76 kg of greenhouse gases, 49.60 kg of carbon monoxide, 7.91 kg of volatile hydrocarbons, 20.05 kg of nitrogen oxides, and 0.32 kg of particulates could be saved. Environmental costs could furthermore be reduced by €2661.82, at €195 per ton of carbon dioxide equivalent, or by €6798.33, at €680 per ton of carbon dioxide equivalent.

Discussion

Principal Findings

This analysis of the economics of using telemedicine in follow-up care for patients in orthopedic and trauma surgery in a German university hospital showed that implementing video consultations enabled time and cost savings for patients, savings in environmental costs, and reductions in emissions.

Implications for Patients

Seen from the patients' point of view in the societal perspective of the health economic analysis, the use of telemedicine was not associated with additional costs (eg, out-of-pocket payments) for the patients in our study. On the contrary, compared with the control group, telemedical appointments resulted in cost savings due to the avoidance of travel and the reduction in time costs.

Previous economic evaluations by Buvik et al [36] and Ohinmaa et al [37] also showed that telemedicine saved travel time and travel distance—and, thus, travel costs—in sparsely populated Scandinavian countries even though patients had to travel to a local caregiver for their appointment [36,37]. Similarly, RCTs by Sathiyakumar et al [9] and Kane et al [12] found savings in travel distances and time spent as well, but these studies did not feature economic analyses [9,12]. Reducing travel burdens is an important societal benefit of telemedicine, as it can ensure better access to medical care. In particular, patients in rural regions and hospitals that seek to offer their medical services beyond their own region stand to benefit. At the same time, however, all patients must still be able to reach their local clinic when video consultations are not sufficient.

Since our trial ended in 2021, our analysis did not consider the energy pricing dynamics following the 2022 European energy crisis. Actual savings in travel costs could be far higher in future digital health deployments.

In addition, the results of the analysis showed that the average costs of lost production were lower for a video consultation compared to a clinical consultation, indicating that telemedicine may have a positive impact in this regard as well. The potential of telemedicine to reduce lost work time—and, thus, production losses—reported here is consistent with the findings of other RCTs [9,12,36,37].

From a societal point of view, the use of telemedicine saved average total costs for employed patients of €76.52 per follow-up appointment, ranging from €66.78 to €80.46 in the sensitivity analysis. Most likely, the real savings would be even higher, as patients often wish or require an accompanying person for a clinic consultation, and the cost and time savings of companions were not considered in the study. The finding that video consultations save overall costs compared with conventional F2F examinations in follow-up care is also confirmed by Buvik et al's [36] analysis. It should be noted, however, that in our calculation patient time lost due to a follow-up appointment was assigned a monetary value independently of any production losses, because including such time costs is strongly recommended in health economic methodology [28,30].

Implications for the Environment

In addition, from the environmental point of view in the societal perspective, our analysis showed that for each patient who received a video consultation instead of a clinic consultation, emissions of 11.248 kg of greenhouse gases, 0.070 kg of carbon monoxide, 0.011 kg of volatile hydrocarbons, 0.028 kg of nitrogen oxides, and 0.0004 kg of particulates could be saved due to avoiding traveling by car. International studies have also demonstrated the reduction of emissions through the use of telemedicine, although the level of individual emissions differs in the respective studies [38,39]. For example, in a study by Udayaraj et al [23], telemedicine led to a reduction of 3527 miles and saved 1035 kg of carbon dioxide for kidney transplant patients in the United Kingdom. A retrospective analysis of patients in vascular surgery in the United States by Paquette and Lin [24] found a reduction of 1632 kg of carbon dioxide; 42,867 g of carbon monoxide; and 3160 g of nitric oxides by performing a total of 146 telemedicine encounters. In addition, based on Spanish data, a study by Vidal-Alaball et al [25] showed an average reduction of 3248.3 g of carbon dioxide, 4.05 g of carbon monoxide, and 4.86 g of nitric oxides per patient in a telemedicine program that included different specialties.

In our study, up to 8 patients could be treated weekly via telemedicine, which can lead to an annual improvement in the environmental footprint for a single German university orthopedic and trauma surgery department alone. Although the performance of telemedicine is not suitable for all patients in orthopedic and trauma surgery, the reduction in emissions could be improved by increasing the number of patients treated by video consultations each week. If the number of patients were expanded to the 1903 hospitals in Germany and included specialties suitable for telemedicine, such as general and visceral surgeries or dermatology, the call of the 125th German Medical

XSL•FO
RenderX

Assembly in 2021 for a net-zero German health care system could be substantially supported [22].

In addition to the emission savings themselves, our study also showed that the introduction of telemedicine can also contribute to a reduction in environmental costs from the societal perspective.

Implications for Practice

This health economic analysis provides clinical evidence that can improve stakeholders' decision-making on implementing telemedicine both in and beyond the current COVID-19 pandemic. It was shown that the use of telemedicine in the follow-up care of orthopedic and trauma surgery benefits both patients and the environment from an economic perspective. Given the pragmatic design of this study, it can be expected that its main findings can be applied by decision makers in other clinical contexts as well.

When deciding whether to implement telemedicine, however, health care providers should consider other aspects besides the economic and environmental benefits. First, the quality of care provided by telemedicine must be ensured. Patient and physician satisfaction, efficiency, and the safety of the video consultations in terms of the same clinical outcomes achieved in F2F consultations play an important role. Various studies show that these goals can be achieved by introducing telemedicine in orthopedic and trauma surgery [6-12]. In addition, we have extensively analyzed patient and physician satisfaction, as well as quality of care for the study cohort in a previous publication [13]. Second, the costs of the technological infrastructure for telemedicine (eg, for electricity, internet connection, and hardware, such as computers and laptops with cameras and microphones) have to be considered. This infrastructure, however, is expected to be part of the standard equipment in most hospitals, as was the case in our study.

Limitations

This study also has some limitations that should be noted. First, although the results were primarily based on actual data collected in the course of an RCT, some assumptions had to be made to be able to calculate costs. Travel costs saved, for example, were calculated based on the assumption that patients have their video consultations at home. In fact, they could have them anywhere, meaning that patients' actual travel costs from that place to the hospital may well be higher or lower. The distance from home and the time spent on the appointments (eg, travel and waiting times) were furthermore queried via a questionnaire, and the actual distances and times could potentially differ slightly from the information provided by the patients. In addition, the original calculation of production loss lacked information on whether patients were employed full time or part time. For this reason, a sensitivity analysis sought to identify possible deviations and to evaluate the robustness of the findings.

Furthermore, given that data on time costs for German patients were missing in the literature, Verbooy et al's [30] valuation approach was used, which was based on Dutch data. However, assuming that the Dutch population is reasonably similar to the German one, this minor inconsistency appears unlikely to have distorted overall results.

Finally, one of the inclusion criteria of the study was patients' ownership of a technical device (smartphone, computer, etc) that allowed them to make video calls. This requirement could lead to socioeconomic inequalities being exacerbated, because only patients with adequate financial means might be able to benefit from cost savings due to telemedicine [40]. This inequity could not be avoided within the study, but it is an important issue with practical relevance and should be taken into account by policy makers.

Conclusions

The use of telemedicine was found to reduce the environmental footprint and to save travel costs, travel time, and time costs for patients, and it helped to lower production losses from a societal perspective compared to F2F consultations in Germany. Thus, telemedicine helps to reduce costs in multiple dimensions. These results were demonstrated in the first health economic analysis of the use of telemedicine in follow-up care for patients with knee and shoulder disorders in orthopedic and trauma surgery, based on data from Germany. Simultaneously, this study provided economic and environmental evidence supporting stakeholders, such as hospitals, patients, and policy makers, who may consider extending the use of telemedicine in and beyond the COVID-19 pandemic. In addition, these findings might be relevant beyond the medical specialty of orthopedic and trauma surgery; they could be applied to other clinical contexts and to a wide range of potential digital health applications that substitute outpatient hospital visits with video consultations.

Acknowledgments

The authors wish to thank Dr med Alexander Eicher for his support in organizing the study.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Consolidated Standards of Reporting Trials (CONSORT) flow diagram.
[DOCX File , 87 KB - jmir_v24i11e42839_app1.docx]

Multimedia Appendix 2
Detailed presentation of cost calculations.
[DOCX File , 30 KB - jmir_v24i11e42839_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 345 KB - jmir_v24i11e42839_app3.pdf]

References

1. Wood G, Naudie D, MacDonald S, McCalden R, Bourne R. An electronic clinic for arthroplasty follow-up: A pilot study. Can J Surg 2011 Dec;54(6):381-386 [FREE Full text] [doi: 10.1503/cjs.028510] [Medline: 21939609]
2. Lanham N, Bockelman K, McCriskin B. Telemedicine and orthopaedic surgery. JBJS Rev 2020 Jul;8(7):e20.00083 [FREE Full text] [doi: 10.2106/jbjs.rvw.20.00083]
3. Parisien RL, Shin M, Constant M, Saltzman BM, Li X, Levine WN, et al. Telehealth utilization in response to the novel coronavirus (COVID-19) pandemic in orthopaedic surgery. J Am Acad Orthop Surg 2020;28(11):e487-e492. [doi: 10.5435/jaaos-d-20-00339]
4. Jaenisch M, Kohlhof H, Touet A, Kehrer M, Cucchi D, Burger C, et al. Evaluation of the feasibility of a telemedical examination of the hip and pelvis - Early lessons from the COVID-19 pandemic. Z Orthop Unfall 2021 Feb;159(1):39-46 [FREE Full text] [doi: 10.1055/a-1289-0779] [Medline: 33327028]
5. Kolin DA, Carroll KM, Plancher K, Cushner F. Perspective of attending physicians on the use of telemedicine in an outpatient arthroplasty setting during the COVID-19 pandemic. HSS J 2021 Feb;17(1):31-35 [FREE Full text] [doi: 10.1177/1556331620979984] [Medline: 33967639]
6. Vuolio S, Winblad I, Ohinmaa A, Haukipuro K. Videoconferencing for orthopaedic outpatients: One-year follow-up. J Telemed Telecare 2003;9(1):8-11. [doi: 10.1258/135763303321159620] [Medline: 12641886]
7. Good DW, Lui DF, Leonard M, Morris S, McElwain JP. Skype: A tool for functional assessment in orthopaedic research. J Telemed Telecare 2012 Mar;18(2):94-98. [doi: 10.1258/jtt.2011.110814] [Medline: 22247538]
8. Sharareh B, Schwarzkopf R. Effectiveness of telemedical applications in postoperative follow-up after total joint arthroplasty. J Arthroplasty 2014 May;29(5):918-922.e1. [doi: 10.1016/j.arth.2013.09.019] [Medline: 24342278]
9. Sathiyakumar V, Apfeld J, Obremskey W, Thakore RV, Sethi MK. Prospective randomized controlled trial using telemedicine for follow-ups in an orthopedic trauma population. J Orthop Trauma 2015 Mar;29(3):e139-e145. [doi: 10.1097/bot.0000000000000189]
10. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Quality of care for remote orthopaedic consultations using telemedicine: A randomised controlled trial. BMC Health Serv Res 2016 Sep 08;16:483 [FREE Full text] [doi: 10.1186/s12913-016-1717-7] [Medline: 27608768]
11. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Patient reported outcomes with remote orthopaedic consultations by telemedicine: A randomised controlled trial. J Telemed Telecare 2018 Jul 04;25(8):451-459. [doi: 10.1177/1357633x18783921]
12. Kane LT, Thakar O, Jamgochian G, Lazarus MD, Abboud JA, Namdari S, et al. The role of telehealth as a platform for postoperative visits following rotator cuff repair: A prospective, randomized controlled trial. J Shoulder Elbow Surg 2020 Apr;29(4):775-783. [doi: 10.1016/j.jse.2019.12.004] [Medline: 32197766]
13. Muschol J, Heinrich M, Heiss C, Knapp G, Repp H, Schneider H, et al. Assessing telemedicine efficiency in follow-up care with video consultations for patients in orthopedic and trauma surgery in Germany: Randomized controlled trial. J Med Internet Res 2022 Jul 27;24(7):e36996 [FREE Full text] [doi: 10.2196/36996] [Medline: 35896015]
14. Al Knawy B, Adil M, Crooks G, Rhee K, Bates D, Jokhdar H, et al. The Riyadh Declaration: The role of digital health in fighting pandemics. Lancet 2020 Nov 14;396(10262):1537-1539 [FREE Full text] [doi: 10.1016/S0140-6736(20)31978-4] [Medline: 32976771]
15. Petersen W, Karpinski K, Backhaus L, Bierke S, Häner M. A systematic review about telemedicine in orthopedics. Arch Orthop Trauma Surg 2021 Oct;141(10):1731-1739 [FREE Full text] [doi: 10.1007/s00402-021-03788-1] [Medline: 33635399]
16. Foni NO, Costa LAV, Velloso LMR, Pedrotti CHS. Telemedicine: Is it a tool for orthopedics? Curr Rev Musculoskelet Med 2020 Dec;13(6):797-801 [FREE Full text] [doi: 10.1007/s12178-020-09680-6] [Medline: 33119841]
17. Moisan P, Barimani B, Antoniou J. Orthopedic surgery and telemedicine in times of COVID-19 and beyond: A review. Curr Rev Musculoskelet Med 2021 Apr;14(2):155-159 [FREE Full text] [doi: 10.1007/s12178-021-09693-9] [Medline: 33460020]

XSL•FO
RenderX

18. Haider Z, Aweid B, Subramanian P, Iranpour F. Telemedicine in orthopaedics during COVID-19 and beyond: A systematic review. J Telemed Telecare 2022 Jul;28(6):391-403 [FREE Full text] [doi: 10.1177/1357633X20938241] [Medline: 32762270]
19. McMaster T, Wright T, Mori K, Stelmach W, To H. Current and future use of telemedicine in surgical clinics during and beyond COVID-19: A narrative review. Ann Med Surg (Lond) 2021 Jun;66:102378 [FREE Full text] [doi: 10.1016/j.amsu.2021.102378] [Medline: 33996071]
20. Cole PA, Lezak BA, Schroder LK, Cole PA. Global orthopaedic trauma surgeons highlight telenomics during the COVID-19 era: A case for advancing telemedicine in orthopaedics. J Clin Orthop Trauma 2021 Jun;17:182-185 [FREE Full text] [doi: 10.1016/j.jcot.2021.03.015] [Medline: 33814860]
21. Oates EV, Lim GHC, Nevins EJ, Kanakala V. Are surgical patients satisfied with remote consultations? A comparison of remote versus conventional outpatient clinic follow-up for surgical patients: A systematic review and meta-analysis of randomized controlled trials. J Patient Exp 2021;8:1-7 [FREE Full text] [doi: 10.1177/23743735211035916] [Medline: 34377778]
22. Gießelmann K, Osterloh F. Deutscher Ärztetag. Klimaschutz im Gesundheitswesen: Klimaneutralität bis 2030 [Article in German]. aerzteblatt.de. 2021. URL: https://www.aerzteblatt.de/archiv/222006/Klimaschutz-im-Gesundheitswesen-Klimaneutralitaet-bis-2030 [accessed 2022-06-16]
23. Udayaraj UP, Watson O, Ben-Shlomo Y, Langdon M, Anderson K, Power A, et al. Establishing a tele-clinic service for kidney transplant recipients through a patient-codesigned quality improvement project. BMJ Open Qual 2019;8(2):e000427 [FREE Full text] [doi: 10.1136/bmjoq-2018-000427] [Medline: 31206050]
24. Paquette S, Lin JC. Outpatient telemedicine program in vascular surgery reduces patient travel time, cost, and environmental pollutant emissions. Ann Vasc Surg 2019 Aug;59:167-172. [doi: 10.1016/j.avsg.2019.01.021] [Medline: 31077768]
25. Vidal-Alaball J, Franch-Parella J, Lopez Seguí F, Garcia Cuyàs F, Mendioroz Peña J. Impact of a telemedicine program on the reduction in the emission of atmospheric pollutants and journeys by road. Int J Environ Res Public Health 2019 Nov 08;16(22):4366 [FREE Full text] [doi: 10.3390/ijerph16224366] [Medline: 31717386]
26. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
27. Bergmo TS. How to measure costs and benefits of eHealth interventions: An overview of methods and frameworks. J Med Internet Res 2015 Nov 09;17(11):e254 [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]
28. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. JAMA 2016 Sep 13;316(10):1093-1103. [doi: 10.1001/jama.2016.12195] [Medline: 27623463]
29. Krauth C, Hessel F, Hansmeier T, Wasem J, Seitz R, Schweikert B. [Empirical standard costs for health economic evaluation in Germany -- A proposal by the working group methods in health economic evaluation] [Article in German]. Gesundheitswesen 2005 Oct;67(10):736-746. [doi: 10.1055/s-2005-858698] [Medline: 16235143]
30. Verbooy K, Hoefman R, van Exel J, Brouwer W. Time is money: Investigating the value of leisure time and unpaid work. Value Health 2018 Dec;21(12):1428-1436 [FREE Full text] [doi: 10.1016/j.jval.2018.04.1828] [Medline: 30502787]
31. National accounts, domestic product. German Federal Statistical Office (Destatis). 2022. URL: https://www.destatis.de/EN/Themes/Economy/National-Accounts-Domestic-Product/Tables/domestic-product-important-economic-indicators.html [accessed 2022-09-16]
32. Quality of employment - Weekly working hours. German Federal Statistical Office (Destatis). 2021. URL: https://www.destatis.de/EN/Themes/Labour/Labour-Market/Quality-Employment/Dimension3/3_1_WeeklyHoursWorked.html [accessed 2022-09-16]
33. Emissionsdaten. Umweltbundesamt [Content in German]. 2022. URL: https://www.umweltbundesamt.de/themen/verkehr-laerm/emissionsdaten#hbefa [accessed 2022-09-16]
34. Gesellschaftliche Kosten von Umweltbelastungen. Umweltbundesamt [Content in German]. 2021. URL: https://tinyurl.com/pmzsa34p [accessed 2022-09-16]
35. Matthey A, Bünger B. Methodenkonvention 3.1 zur Ermittlung von Umweltkosten: Kostensätze. Umweltbundesamt [Content in German]. Dessau-Roßlau, Germany; 2020 Dec. URL: https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2020-12-21_methodenkonvention_3_1_kostensaetze.pdf [accessed 2022-09-16]
36. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of telemedicine in remote orthopedic consultations: Randomized controlled trial. J Med Internet Res 2019 Feb 19;21(2):e11330 [FREE Full text] [doi: 10.2196/11330] [Medline: 30777845]
37. Ohinmaa A, Vuolio S, Haukipuro K, Winblad I. A cost-minimization analysis of orthopaedic consultations using videoconferencing in comparison with conventional consulting. J Telemed Telecare 2002;8(5):283-289. [doi: 10.1177/1357633X0200800507] [Medline: 12396857]
38. Purohit A, Smith J, Hibble A. Does telemedicine reduce the carbon footprint of healthcare? A systematic review. Future Healthc J 2021 Mar;8(1):e85-e91 [FREE Full text] [doi: 10.7861/fhj.2020-0080] [Medline: 33791483]

XSL•FO
RenderX

39. Ravindrane R, Patel J. The environmental impacts of telemedicine in place of face-to-face patient care: A systematic review. Future Healthc J 2022 Mar 17;9(1):28-33 [FREE Full text] [doi: 10.7861/fhj.2021-0148] [Medline: 35372776]
40. OECD. Health at a Glance 2021: OECD Indicators. Paris, France: OECD Publishing; 2021. URL: https://read.oecd.org/10.1787/ae3016b9-en?format=pdf [accessed 2022-11-11]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
F2F: face-to-face
pkm: passenger-kilometer
RCT: randomized controlled trial

Edited by G Eysenbach, T Leung; submitted 20.09.22; peer-reviewed by S Dramburg, J Nielsen; comments to author 25.10.22; revised version received 01.11.22; accepted 04.11.22; published 24.11.22.

Please cite as:
Muschol J, Heinrich M, Heiss C, Hernandez AM, Knapp G, Repp H, Schneider H, Thormann U, Uhlar J, Unzeitig K, Gissel C Economic and Environmental Impact of Digital Health App Video Consultations in Follow-up Care for Patients in Orthopedic and Trauma Surgery in Germany: Randomized Controlled Trial
J Med Internet Res 2022;24(11):e42839
URL: https://www.jmir.org/2022/11/e42839
doi:10.2196/42839
PMID:36333935

©Jennifer Muschol, Martin Heinrich, Christian Heiss, Alher Mauricio Hernandez, Gero Knapp, Holger Repp, Henning Schneider, Ulrich Thormann, Johanna Uhlar, Kai Unzeitig, Christian Gissel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Associated With Telemedicine Use Among German General Practitioners and Rheumatologists: Secondary Analysis of Data From a Nationwide Survey

Felix Muehlensiepen[1,2*], MPH, Dr rer med; Pascal Petit[2*], PhD; Johannes Knitza[2,3], MD, MHBA; Martin Welcker[4], MD; Nicolas Vuillerme[2,5,6], PhD

[1]Center for Health Services Research, Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Rüdersdorf bei Berlin, Germany

[2]AGEIS, Université Grenoble Alpes, Grenoble, France

[3]Department of Internal Medicine 3, Friedrich-Alexander-University Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany

[4]Medizinisches Versorgungszentrum für Rheumatologie Dr M Welcker GmbH, Planegg, Germany

[5]Institut Universitaire de France, Paris, France

[6]LabCom Telecom4Health, Orange Labs & Université Grenoble Alpes, CNRS, Inria, Grenoble, France

[*]these authors contributed equally

Corresponding Author:
Felix Muehlensiepen, MPH, Dr rer med
Center for Health Services Research
Faculty of Health Sciences Brandenburg
Brandenburg Medical School Theodor Fontane
Seebad 82/83
Rüdersdorf bei Berlin, 15562
Germany
Phone: 49 15119126024
Email: Felix.Muehlensiepen@mhb-fontane.de

Abstract

Background: Previous studies have demonstrated telemedicine (TM) to be an effective tool to complement rheumatology care and address workforce shortage. With the outbreak of the SARS-CoV-2 pandemic, TM experienced a massive upswing. However, in rheumatology care, the use of TM stagnated again shortly thereafter. Consequently, the factors associated with physicians' willingness to use TM (TM willingness) and actual use of TM (TM use) need to be thoroughly investigated.

Objective: This study aimed to identify the factors that determine TM use and TM willingness among German general practitioners and rheumatologists.

Methods: We conducted a secondary analysis of data from a German nationwide cross-sectional survey with general practitioners and rheumatologists. Bayesian univariate and multivariate logistic regression analyses were applied to the data to determine which factors were associated with TM use and TM willingness. The predictor variables (covariates) that were studied individually included sociodemographic factors (eg, age and sex), work characteristics (eg, practice location and medical specialty), and self-assessed knowledge of TM. All the variables positively and negatively associated with TM use and TM willingness in the univariate analysis were then considered for Bayesian model averaging analysis after a selection based on the variance inflation factor (≤2.5). All analyses were stratified by sex.

Results: Univariate analysis revealed that out of 83 variables, 36 (43%) and 34 (41%) variables were positively or negatively associated (region of practical equivalence≤5%) with TM use and TM willingness, respectively. The Bayesian model averaging analysis allowed us to identify 13 and 17 factors of TM use and TM willingness, respectively. Among these factors, being female, having very poor knowledge of TM, treating <500 patients per quarter, and not being willing to use TM were negatively associated with TM use, whereas having good knowledge of TM and treating >1000 patients per quarter were positively associated with TM use. In addition, being aged 51 to 60 years, thinking that TM is not important for current and future work, and not currently using TM were negatively associated with TM willingness, whereas owning a smart device and working in an urban area were positively associated with TM willingness.

XSL•FO
RenderX

Conclusions: The results point to the close connection between health care professionals' knowledge of TM and actual TM use. These results lend support to the integration of digital competencies into medical education as well as hands-on training for health care professionals. Incentive programs for physicians aged >50 years and practicing in rural areas could further encourage TM willingness.

(J Med Internet Res 2022;24(11):e40304) doi:10.2196/40304

KEYWORDS

telemedicine; rheumatology; primary care; secondary analysis; health services research

Introduction

Telemedicine (TM) offers the opportunity to overcome spatial distances in health care delivery [1]. Thus, TM represents a promising way to support rheumatology care [2,3] in light of the rising worldwide burden of musculoskeletal diseases [4] and growing workforce shortage [5,6]. However, the effective implementation of TM in standard care is only possible if the end users are willing and able to use TM [7,8].

With the outbreak of the SARS-CoV-2 pandemic, physicians' face-to-face consultations declined considerably [9,10]. The possibility of contactless medical care is now more important. Advantageously, through TM, medical care could be provided, avoiding contacts and thus infections [11,12]. Hence, TM has received a tremendous upswing worldwide [13] and regionally [9,14]. Although the pandemic situation, involving social distancing and multiple lockdowns, provided an ideal environment for the implementation of TM, this momentum soon stagnated again [10,15]. Particularly in rheumatology, health care professionals' use and acceptance of TM fell short of expectations [10]. Apparently, other factors may play a role in the willingness to use TM (TM willingness) and actual use of TM (TM use) among general practitioners (GPs) and rheumatologists. Identifying these factors is a rather challenging task but could have implications for the development of TM strategies aiming to improve health outcomes and access to care and make health care delivery systems more efficient and cost-effective.

To gain a better understanding of these factors, we performed a secondary analysis using data from a nationwide cross-sectional survey conducted earlier in Germany [7]. Our objective was to identify the underlying factors associated with TM use and the TM willingness among German GPs and rheumatologists.

Methods

Overview

This work reports on findings from a secondary analysis of data collected as part of a cross-sectional, self-completed, and paper-based survey of German GPs and outpatient rheumatologists. The initial study was conducted from September to November 2018 and investigated the acceptance, opportunities, and obstacles to the implementation of TM. Of the 2395 questionnaires that were sent out, 497 (20.75%) were returned. Of the 497 responses, 12 (2.4%) were excluded from the data set because fewer than half of the questions were answered. The final response rates were 18.94% (437/2307) and 55% (48/88) for GPs and rheumatologists, respectively. The exact methodology applied for the nationwide survey has been described previously [7].

Regression Analysis

Both Bayesian univariate and multivariate logistic regression analyses were applied to the data to determine which factors were associated with TM use (question [Q]3) and TM willingness (Q4A), respectively. In total, 22 independent variables were considered for each univariate regression analysis (Multimedia Appendix 1). The individuals who missed providing information on age or gender or answers to Q3 (467/492, 5.1%) and Q4A (454/492, 7.7%) were excluded. Otherwise, missing values (no answer) were considered as a new category for the univariate regression analysis. For instance, Q28, "assigning physician or rheumatologist," previously had 2 categories and was revised to have 3 categories, "assigning physician," "rheumatologist," or "not answered". For statistical analysis, all the categorical variables having >2 modalities, for example, "yes," "no," or "do not know," were transformed into dummy or binary variables. For instance, Q21 was transformed into 3 dummy variables.

For each model, odds ratios (ORs) with 95% credible interval (CI) are presented. All the individual variables associated (positively or negatively) with TM use and TM willingness in the Bayesian univariate analysis were considered for analysis in the later Bayesian multivariate analysis (model selection) after variable selection. This variable selection was based on the region of practical equivalence (ROPE) percentage (ROPE%≤5) [16] and a subsequent selection based first on the variance inflation factor (VIF) [17]. Collinear covariates, with a VIF>2.5, were excluded in the multivariate models [18]. Finally, the determinants of TM use and TM willingness were identified through Bayesian model averaging (BMA) [19]. The "best" model (ie, model with the highest posterior probability) from BMA was detailed. All models were stratified by sex. In addition, determinant factors (question answers), defined as variables with a posterior probability of ≥10% with BMA, were identified and used to establish the profile of the individuals using or willing to use TM and the profile of the individuals not using or not willing to use TM using spider charts. For each determinant factor, the percentage of individuals who chose a specific answer was displayed on the spider chart. This percentage could range from 0 (the inner circular line, the closest to the center) if no individuals chose the specified answer for the considered question to 100 (the outer circular line, the farthest from the radar center) if all individuals answered the question with the specified answer. Green points and lines on the spider charts refer to the individuals who use or want to use

XSL•FO
RenderX

TM, whereas red points and lines correspond to the individuals not willing to use or not using TM. For each question, there were 3 possible situations. When the green and red points overlapped (were similar), it meant that there was no difference between the individuals whether they were using TM or not or willing or not to use TM, that is, the proportion of similar answers was high. When the green point was higher (higher percentage) than the red point, it indicated that the individuals using or willing to use TM chose the specified answer more often than those not willing to use or not using TM, which meant that this factor (question) had a positive impact on TM use or TM willingness. Finally, when the green point was lower (lower percentage) than the red point, it indicated that the individuals willing to use or using TM chose the specified answer less often than the individuals not willing to use or not using TM, which meant that this factor (question) had a negative impact on TM use or TM willingness.

All statistical analyses were performed using R software (version 4.1.2, R Foundation for Statistical Computing) for Windows 10. The *tidyverse* package (version 1.3.2) was used [20]. VIFs were calculated using the *car* package (version 3.1-0) [21]. Bayesian estimation was performed using the *rstanarm* package (version 2.21.1) [22,23]. Weakly informative priors (default priors in *rstarnarm*) were used. The default priors in *rstanarm* 2.21.1 are designed to be weakly informative. The Bayesian model adds priors (independent by default) to the coefficients of the generalized linear model. The Bayesian estimation was performed via the Markov chain Monte Carlo Bernoulli model, with 4 randomly initialized Markov chains, each for 2000 iterations (including a warm-up period of 1000 iterations that is discarded). Posterior distributions were described using the *bayestestR* package (version 0.12.1) [24]. The selection of the "best" model through BMA was undertaken using the *BMA* package (version 3.18.15) [25]. Regarding priors for BMA, we assumed that all candidate models were equally likely a priori (same prior weight). The spider charts were created using the *fmsb* package (version 0.7.3) [26].

Ethics Approval

Primary data collection was conducted in compliance with the current data protection regulations of the General Data Protection Regulation [27] and the Helsinki Declaration. All study participants were informed about the research project and provided written informed consent. Data were anonymized before analysis. The ethics committee of the Theodor Fontane Medical School in Brandenburg stated that no written consent was necessary owing to the noninterventional study design, which also applies to the secondary analysis.

Results

Population Characteristics

A total of 94.9% (467/492) and 92.3% (454/492) of individuals were selected for the analysis of TM use and TM willingness, respectively. Most participants (247/454, 54.4%) were female. Most individuals were GPs (408/454, 89.9%) and were aged between 51 and 60 years (215/454, 47.4%). Although most individuals were not using TM (344/454, 75.8%), two-thirds (282/454, 62.1%) were willing to use it in the future.

Bayesian Univariate Logistic Regression Analysis

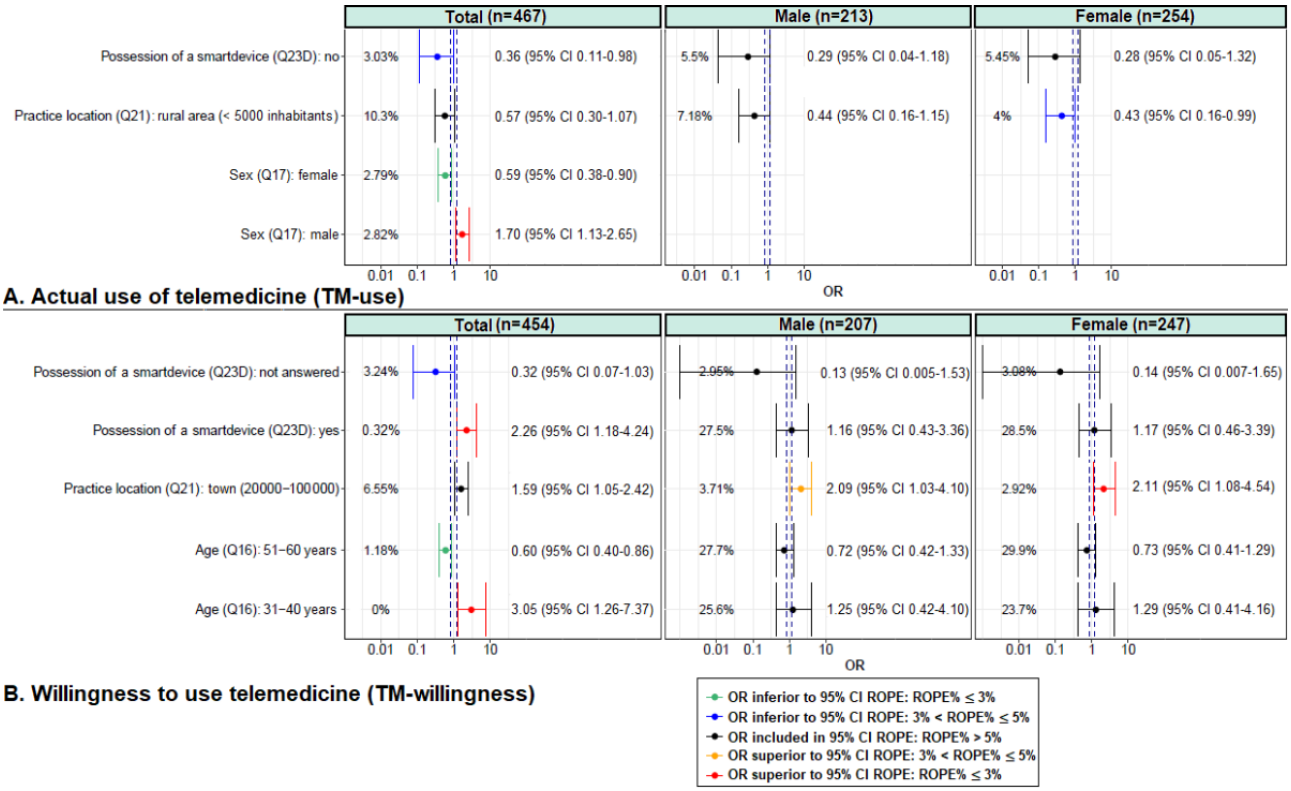
Only significant results are presented in the main text, but all the results can be found in the Multimedia Appendices 1-5 and Figures S1-S4 in Multimedia Appendix 6. A total of 26 questions were answered (83 answers) and analyzed using the univariate logistic regression analysis. Out of 83 variables, 36 (43%) and 34 (41%) variables were found to be positively or negatively associated (ROPE%≤5%) with TM use and TM willingness, respectively (Multimedia Appendix 2). Regarding sociodemographic factors (Figure 1), not owning a smart device (OR 0.36, 95% CI 0.11-0.99; ROPE%=3.0); being female (OR 0.59, 95% CI 0.38-0.90; ROPE%=2.8); and being female with a practice located in rural area (<5000 inhabitants; OR 0.43, 95% CI 0.16-0.99; ROPE%=4.0) were negatively associated with TM use, whereas being aged between 51 and 60 years (OR 0.60, 95% CI 0.40-0.86; ROPE%=1.2) was negatively associated with TM willingness. By contrast, being male (OR 1.70, 95% CI 1.13-2.65; ROPE%=2.8) was positively associated with TM use, whereas owning a smart device (OR 2.26, 95% CI 1.18-4.24; ROPE%=0.3); being aged 31 to 40 years (OR 3.05 95% CI 1.26-7.37; ROPE%=0); and having a practice located in town (20,000-100,000 inhabitants) were positively associated with TM willingness. For more details, please refer to Figures S1 and S2 in Multimedia Appendix 6.

Regarding work characteristics, being a rheumatologist, working in a medical care center, and treating >1000 patients per quarter were positively associated with TM use, whereas treating <500 patients per quarter and being an assigning physician were negatively associated (Multimedia Appendix 2 and Figures S3 and S4 in Multimedia Appendix 6).

Regarding the opinion and knowledge about TM, having at least good TM knowledge, thinking that TM is suitable for exchange in rheumatology, wanting to exchange information with specialists via TM, and thinking that TM is at least rather important for current and future work were positively associated with both TM use and TM willingness (Multimedia Appendix 2 and Figures S3-S6 in Multimedia Appendix 6). By contrast, having poor or very poor TM knowledge and thinking that TM is not important at all for current and future work were both negatively associated with both TM use and TM willingness. Individuals willing to use TM were strongly and positively associated with TM use.

XSL•FO
RenderX

Figure 1. Bayesian univariate logistic regression—Relationship between the actual use of telemedicine (TM use) or willingness to use telemedicine (TM willingness) and sociodemographic factors. The percentage indicates the region of practical equivalence (ROPE) percentage, that is, the probability that the considered credible factor values are not negligible. The dashed lines indicate the 95% credible interval (CI) of the ROPE. OR: odds ratio; Q: question.



BMA and Bayesian Multivariate Logistic Regression Analysis

A total of 6 BMA analyses were conducted, with 3 (both sexes, male, and female) for TM use and 3 for TM willingness. Figure 2 presents the determinants identified through BMA for the 6 analyses. Only variables with a posterior probability of ≥10% were considered determinant factors. A total of 16 answers were selected using BMA. Variables above the dashed horizontal line refer to factors positively associated with TM use or TM willingness (cells with color from light yellow to red). By contrast, variables under the dashed horizontal line refer to factors negatively associated with TM use or TM willingness (cells with colors from light green to dark blue). The value in each cell corresponds to the posterior probability that the considered variable is nonzero (in percentage). Darker the color, the higher the posterior probability percentage.

Regarding TM use, a total of 13 determinant factors (13 answers from 8 questions) were identified. Being female, having very poor knowledge of TM, treating <500 patients per quarter, thinking that TM is not important at all for current work, and not being willing to use TM were negatively associated with TM use. By contrast, having good or very good knowledge of TM, thinking that TM is important or very important for current work and at least rather not important for future work, treating >1000 patients per quarter, and thinking that TM is suitable for exchange in rheumatology were positively associated with TM use.

Regarding TM willingness, a total of 17 determinant factors (17 answers from 11 questions) were identified. Not wanting to exchange information with specialists using TM, thinking TM services have no place in the care process, being aged 51 to 60 years, thinking that TM is not important for current and future work, and not currently using TM were negatively associated with TM willingness. By contrast, owning a smart device, thinking that TM is at least rather not important for future work, thinking that TM is relevant in subareas in rheumatology, and thinking that there should be exchange with TM were positively associated with TM willingness.

For more details about the BMA analysis, please refer to Multimedia Appendix 4, which synthesizes BMA results for the top 5 models, as well as to Figures S7-S11 in Multimedia Appendix 6 for TM use and for TM willingness, which represent all the variables considered (in the y-axis) for the full list of models selected (in the x-axis). Blue color indicates variables negatively associated with TM use or TM willingness, whereas red color indicates variables that are positively associated.

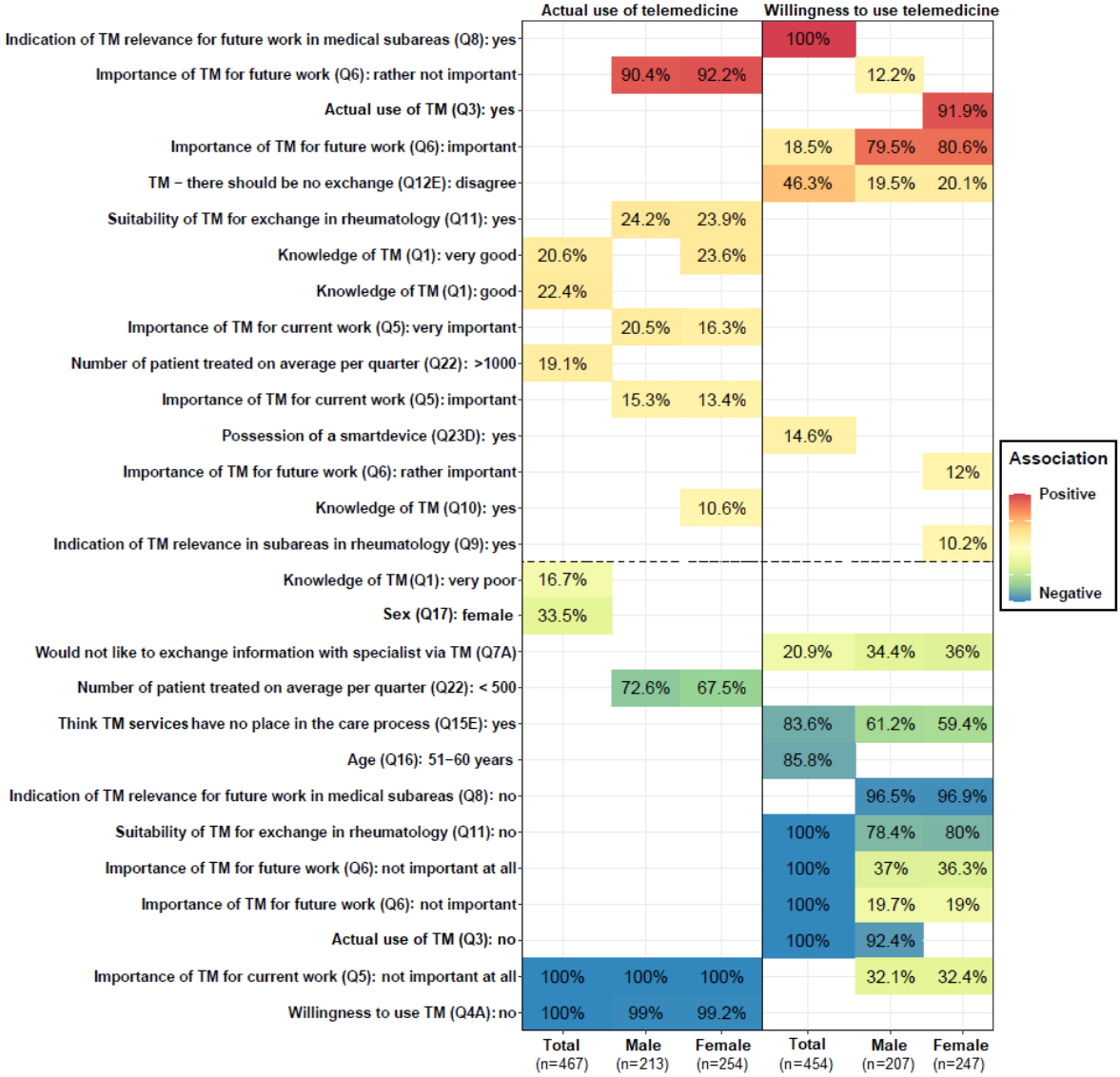
Results for the “best” model identified through BMA indicated that being female (OR 0.57, 95% CI 0.35-0.90; ROPE%=3.2); thinking that TM is not important at all for current work (OR 0.15, 95% CI 0.08-0.29; ROPE%=0); and not being willing to use TM (OR 0.22, 95% CI 0.10-0.38; ROPE%=0) were negatively associated with TM use for both sexes. When stratified by sex, it was found that treating <500 patients per quarter was negatively associated with TM use. Regarding TM willingness, being aged 51 to 60 years (OR 0.43, 95% CI 0.26-0.74; ROPE%=0); not using TM (OR 0.14, 95% CI

XSL•FO
RenderX

0.06-0.31; ROPE%=0); thinking that TM is not suitable for exchange in rheumatology (OR 0.13, 95% CI 0.05-0.35; ROPE%=0); and thinking that it is not important for future work (OR 0.13, 95% CI 0.05-0.35; ROPE%=0) were factors negatively associated with TM willingness for both sexes.

More details about the "best" models are available in Multimedia Appendix 5.

Figure 2. Determinants of the actual use of telemedicine (TM use) or willingness to use telemedicine (TM willingness) identified through the Bayesian model averaging analysis. A total of 28 answers from 16 questions were selected with Bayesian model averaging. The value in each cell corresponds to the posterior probability that the considered variable is nonzero (in percentage). Q: question.



Profile of TM Users or Individuals Willing to Use TM

Determinant factors, defined as variables with a posterior probability of ≥10% with BMA, were identified and used to establish the profile of individuals using or willing to use TM and the profile of individuals not using or not willing to use TM. Figure 3 presents the profiles identified based on gender.

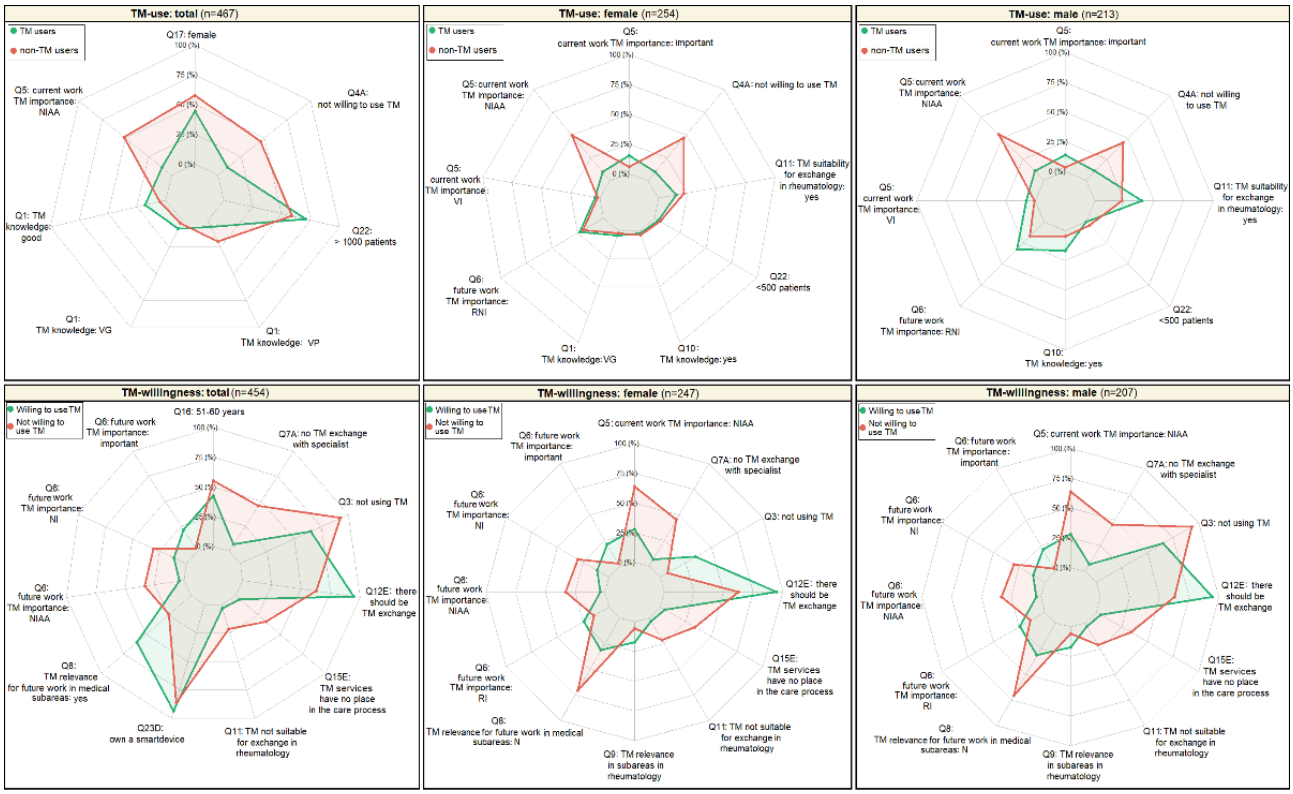
Regarding TM use, TM users more frequently had TM knowledge and treated, on average, more patients (>1000 patients per quarter) than non-TM users.

TM users were more often women, more often thought that TM is not important at all for current work, more frequently had very poor TM knowledge, and were less inclined to use TM compared with TM users.

Regarding TM willingness, the individuals who were willing to use TM owned a smart device and thought that there should be TM exchange more often than the individuals who were not willing to use TM. By contrast, the individuals not willing to use TM were more often aged 51 to 60 years and more frequently thought that TM is not suitable for exchange in rheumatology, is not important at all for current and future work, is not relevant for future work in medical subareas, and has no place in the care process. In addition, they used TM less often than the individuals who were willing to use TM.

XSL•FO
RenderX

Figure 3. Profile of telemedicine (TM) users versus nonusers and individuals willing to use TM versus those not willing to use TM using Bayesian model averaging (BMA). Variables displayed on the spider or radar chart correspond to factors selected with BMA that had a posterior probability of ≥10%. Percentages refer to the percentage of individuals with the answer specified for each question. NI: not important; NIAA: not important at all; RI: rather important; RNI: rather not important; VG: very good; VI: very important; VP: very poor; TM willingness: willingness to use telemedicine; Q: question.



Discussion

Overview

We performed a secondary analysis to identify factors associated with TM use and TM willingness on data collected as part of a cross-sectional, self-completed, and paper-based survey of German GPs and outpatient rheumatologists. The initial study [7] was conducted from September to November 2018, with the goal of exploring general acceptance, opportunities, and obstacles for the implementation of TM. The current secondary analysis was conducted to identify the most relevant factors affecting TM use and TM willingness to enable more effective TM strategies.

Principal Findings

Regarding the factors associated with TM use, our results revealed that having good or very good knowledge of TM and treating >1000 patients per quarter were positively associated with TM use. By contrast, being female, having very poor knowledge of TM, treating <500 patients per quarter, not owning a smart device, working in a rural area, thinking that TM is not important at all for current work, and not being willing to use TM were negatively associated with TM use.

Regarding the factors associated with TM willingness, owning a smart device, thinking that TM is relevant in subareas in rheumatology, working in urban areas, and thinking that there should be exchange with TM were positively associated with TM willingness. By contrast, not wanting to exchange information with specialists using TM, thinking that TM services have no place in the care process, being aged 51 to 60 years, thinking that TM is not important for current and future work, and not currently using TM were negatively associated with TM willingness.

Comparison With Prior Work

To the best of our knowledge, this is the first work analyzing specific factors influencing TM use and TM willingness among German GPs and rheumatologists. A major strength of this study lies in its ability to guide TM implementation strategies.

Our results underline the close connection between knowledge and technology use, as described by Paul Attewell [28]. According to his theory on technology diffusion and organizational learning, knowledge barriers—that is, the lack of knowledge about the technology and how this technology can be applied in an organizational setting—are in fact the reasons why technology diffusion remains low. Consistently, we found that having good or very good self-perceived knowledge of TM is positively associated with TM use, whereas having very poor knowledge is negatively associated with TM use. Similarly, a previous survey study identified the unawareness of suitable software solutions as the main factor that prevented rheumatologists from using electronic instead of paper-based questionnaires [29]. Concurringly, German rheumatologists were only aware of a fraction of the available rheumatology apps, limiting their use in clinical routine [30].

Tanriverdi and Iacono [31] extended Attewell's theory to a multidimensional concept including the economic,

XSL•FO
RenderX

organizational, and behavioral knowledge barriers that hamper the diffusion of TM. Our results support this multidimensionality. For instance, larger medical practices providing for more patients are more likely to use TM than smaller organizational units. Furthermore, in line with the results of Knörr et al [32], physicians in rural areas appear to use TM less frequently than physicians in urban areas, which seems counterintuitive and might also be due to the limited technical infrastructure in rural areas in Germany [8]. However, Vossen et al [33] reported a positive correlation between the traveling time to the treating rheumatologist and the willingness of German patients with rheumatoid arthritis to use video consultations.

In addition, the purchase of technology equipment, administration effort, and inadequate reimbursement (system) of TM services in Germany were identified as the main barriers to TM use in the primary analysis [7]. These barriers were later confirmed in a multiprofessional survey to impact TM use in other medical domains as well [34].

In line with the previous results reported by Alkureishi [35], our analysis results indicated a negative association between being female and TM use. We were surprised by this finding, as eHealth literacy was recently reported to be higher among women, both among health care professionals [36] and the overall German population [37]. Apparently, higher eHealth literacy does not translate directly into higher TM use. The reasons for the gender difference need to be specifically explored in further research, particularly as the proportion of women among physicians continues to increase in Germany [38]. Furthermore, the negative association between being aged 51 to 60 years and TM willingness is striking, as the average age of physicians in Germany is currently 54.2 years with an increasing trend [38]. This is linked to substantial concerns about increasing workforce shortage [5], which TM is actually intended to address [6,39]. However, a previous study on mobile health found no gender differences in patients with rheumatoid arthritis yet revealed a negative correlation with age [40]. Thus, the differences between the study findings may also be explained by specific TM approaches queried and terminology, which should be further researched.

Implications

Because TM use is closely intertwined with physicians' knowledge in this domain, we strongly support the integration of digital competencies into medical education and offering of dedicated training courses for physicians [41-43]. Continuous education in this area seems to be particularly important, as telemedical options continuously increase, including not only medical apps but also completely new procedures such as patient self-sampling. Health care professionals also seem concerned with an increasing workload due to increasing communication and transmitted information via TM [8]; education could help to implement the most successful TM strategies. As Tanriverdi and Iacono [31] discussed earlier, these training courses should also reflect on the multidimensionality of knowledge barriers by addressing the economic, organizational, and behavioral framework conditions of digital health implementation. Administrative, technical, and reimbursement requirements should be addressed first, as these have been reported as key barriers to the use of TM [7], just as they have recently been to the use of prescribed and regulated digital therapies in Germany [44].

Concomitantly, our data point to the importance of the organizational determinants of TM use. Although there are already numerous studies that point to the effectiveness of TM use [3], it remains unclear how TM needs to be integrated into organizational structures to ensure its effective and sustainable use in routine health care. Therefore, we recommend investigating the organizational and social factors of the implementation of TM and digital health in health care delivery.

Furthermore, our findings will inform private and public stakeholders on TM implementation. Public stakeholders, such as health policy makers, might use our findings to promote TM and upgrade infrastructure in rural areas. Specific target groups for incentive schemes could be female physicians aged 51 to 60 years in particular. Private stakeholders, such as TM companies or start-ups, might infer from our findings that health care professionals need low-threshold instructions on the use of their products. Finally, we recommend organizational and structural guidance, including setup, staff planning, billing of services, and administration, for the implementation of TM in routine health care delivery.

Limitations

The primary data on which this analysis was based were collected in 2018 before the SARS-CoV-2 outbreak. Owing to the need to reduce physical contact and thus minimize the risk of infection, TM use initially received a major uptake in global health care delivery [13]. Hence, more physicians and likely other subgroups will have tried TM by now [23], which has led to an increased use and awareness of TM in routine practice. Nevertheless, recent studies suggest that even after the SARS-CoV-2 outbreak, the same barriers continue to prevent widespread TM adoption [9,10,35,44,45]. However, a replication of the initial survey is essential to identify whether and how the identified factors have changed in the surveyed target group. Thus, the results from our study represent a baseline to future studies that would investigate the change in TM experience and perceptions due to the SARS-CoV-2 pandemic.

Apart from the aforementioned shortcomings, the limitations of the primary data still apply [7]. Only a relatively small proportion (44/454, 9.7%) of the survey sample are rheumatologists, which accounts for 7% of all of the rheumatologists in outpatient care in Germany [46]. Although the survey was directed at physicians from all over Germany, it was primarily physicians from Brandenburg who participated because of the recruitment strategy. We suspect a high potential of self-selection and nonresponse bias. Health care professionals in inpatient care as well as other professions involved in rheumatology care (eg, nurses) were not included in the survey. Furthermore, our results cover the perspectives of German physicians only. Their acceptance of TM might be strongly influenced by the specifics and policy drivers of the German health system. Previous studies reported [8,45] weak remuneration, high bureaucracy, and a lack of digital infrastructure to hamper TM use in Germany. Owing to these

XSL•FO
RenderX

influences, the transferability of our results to other countries and health care systems may be limited. Finally, physician engagement is an important factor in the adoption of telehealth into routine care delivery, but it represents only one side of the coin. The patient perspective and TM willingness represent the other side that needs to be investigated as a priority.

Regarding the statistical analysis, we used a Bayesian approach to conduct the secondary analysis of the aforementioned survey. A practical limitation of the Bayesian approach is that it requires the specification of prior distributions both on the parameters of each model and on the distribution of the models themselves. Because we had no a priori assumption, we used weakly informative priors. Choosing another prior distribution may have had substantial influence on the outcome [47,48]. Regarding the variable and model selections, a 3-step approach was used. First, all the individual variables associated (positively or negatively) with the use of or TM willingness in the Bayesian univariate analysis were selected based on the ROPE percentage (ROPE%≤5). Choosing a different ROPE percentage threshold may have yielded different results. Then, we performed a conservative selection based on the VIF (VIF≤2.5) to deal with potential variable multicollinearity. Finally, we used the remaining variables with BMA for model selection and identification of determinants. BMA was chosen in particular because it reduces overconfidence and is relatively robust against model misspecification [47,49-51]. Markov chain Monte Carlo was used to deal with the intractable computational challenge of BMA that comes from the candidate model enumeration [52].

Conclusions

TM use is intertwined with health care professionals' knowledge of TM. Limited knowledge restricts the implementation of TM in rheumatology care. Dedicated education courses could provide the necessary knowledge and improve TM uptake. These courses need to reflect on the multidimensionality of knowledge barriers by addressing the economic, organizational, and behavioral framework conditions of TM implementation.

TM willingness is associated with age and practice location, and incentive programs for advanced physicians practicing in rural areas have the potential to increase the implementation of TM in standard care.

Acknowledgments

The authors would like to thank the participants in the survey, their teams, and all other supporters of TeleRheumaBB. They also owe special gratitude to KV Consult- und Managementgesellschaft mbH, which initiated the study in the first place. This work is a part of the PhD thesis of FM (AGEIS, Université Grenoble Alpes, Grenoble, France). This work was supported by the French National Research Agency (France) under the framework of the Investissements d'avenir program (ANR-10-AIRT-05 and ANR-15-IDEX-02). This publication was funded by the Brandenburg Medical School Open Access Publication Fund supported by the German Research Association. The sponsors had no involvement in the review and approval of the manuscript for publication. This work forms a part of a broader transnational and interdisciplinary collaboration between Université Grenoble Alpes (France), Universitätsklinikum Erlangen-Nürnberg (Germany), and Brandenburg Medical School (Germany).

Authors' Contributions

All the authors were involved in drafting the article and critically revising it for important intellectual content, and all the authors approved the final version to be submitted for publication. FM had full access to all the data in the study and took responsibility for the integrity of the data and accuracy of the data analysis. FM, JK, NV, and PP conceptualized and designed the study. FM, MW, and NV acquired data. FM, JK, NV, and PP analyzed and interpreted the data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Regression analysis—variables.
[PDF File (Adobe PDF File), 291 KB - jmir_v24i11e40304_app1.pdf]

Multimedia Appendix 2
List of all the variables positively and negatively associated (region of practical equivalence≤5%) with the actual use of telemedicine use and willingness to use telemedicine in the Bayesian univariate logistic regression analysis.
[PDF File (Adobe PDF File), 308 KB - jmir_v24i11e40304_app2.pdf]

Multimedia Appendix 3
Bayesian univariate logistic regression analysis results.
[XLSX File (Microsoft Excel File), 49 KB - jmir_v24i11e40304_app3.xlsx]

Multimedia Appendix 4

XSL•FO
RenderX

Bayesian model averaging results.
[XLSX File (Microsoft Excel File), 28 KB - jmir_v24i11e40304_app4.xlsx]

Multimedia Appendix 5
Bayesian multivariate logistic regression analysis results for the best model.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v24i11e40304_app5.xlsx]

Multimedia Appendix 6
Bayesian univariate logistic regression figures.
[DOCX File , 2414 KB - jmir_v24i11e40304_app6.docx]

References

1. WMA Statement on the Ethics of Telemedicine. World Medical Association. 2018. URL: https://www.wma.net/policies-post/wma-statement-on-the-ethics-of-telemedicine/ [accessed 2022-06-10]
2. Piga M, Cangemi I, Mathieu A, Cauli A. Telemedicine for patients with rheumatic diseases: systematic review and proposal for research agenda. Semin Arthritis Rheum 2017 Aug;47(1):121-128. [doi: 10.1016/j.semarthrit.2017.03.014] [Medline: 28420491]
3. McDougall JA, Ferucci ED, Glover J, Fraenkel L. Telerheumatology: a systematic review. Arthritis Care Res (Hoboken) 2017 Oct;69(10):1546-1557 [FREE Full text] [doi: 10.1002/acr.23153] [Medline: 27863164]
4. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: a systematic analysis of the World Health Organization Burden of Diseases Database. Ann Rheum Dis 2019 Jun;78(6):844-848. [doi: 10.1136/annrheumdis-2019-215142] [Medline: 30987966]
5. Al Maini M, Adelowo F, Al Saleh J, Al Weshahi Y, Burmester G, Cutolo M, et al. The global challenges and opportunities in the practice of rheumatology: white paper by the World Forum on Rheumatic and Musculoskeletal Diseases. Clin Rheumatol 2015 May;34(5):819-829 [FREE Full text] [doi: 10.1007/s10067-014-2841-6] [Medline: 25501633]
6. Ward IM, Schmidt TW, Lappan C, Battafarano DF. How critical is tele-medicine to the rheumatology workforce? Arthritis Care Res (Hoboken) 2016 Oct;68(10):1387-1389 [FREE Full text] [doi: 10.1002/acr.22853] [Medline: 26866514]
7. Muehlensiepen F, Knitza J, Marquardt W, Engler J, Hueber A, Welcker M. Acceptance of telerheumatology by rheumatologists and general practitioners in Germany: nationwide cross-sectional survey study. J Med Internet Res 2021 Mar 29;23(3):e23742 [FREE Full text] [doi: 10.2196/23742] [Medline: 33690147]
8. Muehlensiepen F, Knitza J, Marquardt W, May S, Krusche M, Hueber A, et al. Opportunities and barriers of telemedicine in rheumatology: a participatory, mixed-methods study. Int J Environ Res Public Health 2021 Dec 13;18(24):13127 [FREE Full text] [doi: 10.3390/ijerph182413127] [Medline: 34948737]
9. Bruch D, Muehlensiepen F, Alexandrov A, Konstantinova Y, Voß K, Ronckers C, et al. The impact of the COVID-19 pandemic on professional practice and patient volume in medical practices: a survey among German physicians and psychotherapists. Z Evid Fortbild Qual Gesundhwes 2021 Nov;166:27-35 [FREE Full text] [doi: 10.1016/j.zefq.2021.08.001] [Medline: 34474990]
10. Richter JG, Chehab G, Reiter J, Aries P, Muehlensiepen F, Welcker M, et al. Evaluation of the use of video consultation in German rheumatology care before and during the COVID-19 pandemic. Front Med (forthcoming) 2022 Nov 07 [FREE Full text] [doi: 10.3389/fmed.2022.1052055]
11. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
12. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Jun;26(5):309-313 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]
13. Omboni S, Padwal RS, Alessa T, Benczúr B, Green BB, Hubbard I, et al. The worldwide impact of telemedicine during COVID-19: current evidence and recommendations for the future. Connect Health 2022 Jan 04;1:7-35 [FREE Full text] [doi: 10.20517/ch.2021.03] [Medline: 35233563]
14. El Aoufy K, Melis MR, Bellando Randone S, Blagojevic J, Bartoli F, Fiori G, et al. The positive side of the coin: Sars-Cov-2 pandemic has taught us how much Telemedicine is useful as standard of care procedure in real life. Clin Rheumatol 2022 Feb;41(2):573-579 [FREE Full text] [doi: 10.1007/s10067-021-05975-2] [Medline: 34739619]
15. Veränderung der vertragsärztlichen Leistungsinanspruchnahme während der COVID-Krise. Zentralinstitut für die kassenärztliche Versorgung in Deutschland. 2021 Feb 16. URL: https://www.zi.de/fileadmin/images/content/Publikationen/Trendreport_3_Leistungsinanspruchnahme_COVID_final.pdf [accessed 2022-06-10]
16. Kruschke JK. Rejecting or accepting parameter values in Bayesian estimation. Adv Methods Pract Psychol Sci 2018 May 08;1(2):270-280. [doi: 10.1177/2515245918771304]
17. Midi H, Sarkar SK, Rana S. Collinearity diagnostics of binary logistic regression model. J Interdisciplinary Math 2010 Jun;13(3):253-267. [doi: 10.1080/09720502.2010.10700699]

XSL•FO
RenderX

18. Zuur AF, Iono EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 2010 Mar;1(1):3-14. [doi: 10.1111/j.2041-210x.2009.00001.x]
19. Depaoli S, Lai K, Yang Y. Bayesian model averaging as an alternative to model selection for multilevel models. Multivariate Behav Res 2021 Jul 03;56(6):920-940. [doi: 10.1080/00273171.2020.1778439] [Medline: 32619364]
20. Brilleman S, Crowther M, Moreno-Betancur M, Buros Novik J, Wolfe R. Joint longitudinal and time-to-event models via Stan. GitHub. 2018. URL: https://github.com/stan-dev/stancon_talks/ [accessed 2022-06-10]
21. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov 21;4(43):1686. [doi: 10.21105/joss.01686]
22. Fox J, Weisberg S. An R companion to applied regression. 3rd edition. Sage Publications. 2019. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/ [accessed 2022-06-10]
23. Goodrich B, Gabry J, Ali I, Brilleman S. rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.1. 2020. URL: https://mc-stan.org/rstanarm [accessed 2022-06-10]
24. 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. [doi: 10.21105/joss.01541]
25. Raftery A, Hoeting J, Volinsky C, Painter I, Yeung KY. BMA: Bayesian Model Averaging. Package for Bayesian model averaging and variable selection for linear models, generalized linear models and survival models (cox regression). Version 3.18.15. GitHub. 2021. URL: https://github.com/hanase/BMA [accessed 2022-06-10]
26. Nakazawa M. fmsb: Functions for Medical Statistics Book with some Demographic Data. The Comprehensive R Archive Network. 2022 Mar 1. URL: https://CRAN.R-project.org/package=fmsb [accessed 2022-06-10]
27. The European Parliament and the Council of the European Union. Consolidated text: 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) (Text with EEA relevance). EUR-Lex. URL: https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04 [accessed 2022-11-24]
28. Attewell P. Technology diffusion and organizational learning: the case of business computing. Organ Sci 1992 Feb;3(1):1-19. [doi: 10.1287/orsc.3.1.1]
29. Krusche M, Klemm P, Grahammer M, Mucke J, Vossen D, Kleyer A, et al. Acceptance, usage, and barriers of electronic patient-reported outcomes among German rheumatologists: survey study. JMIR Mhealth Uhealth 2020 Jul 20;8(7):e18117 [FREE Full text] [doi: 10.2196/18117] [Medline: 32390592]
30. Knitza J, Vossen D, Geffken I, Krusche M, Meyer M, Sewerin P, Arbeitskreis Junge Rheumatologen. Use of medical apps and online platforms among German rheumatologists : results of the 2016 and 2018 DGRh conference surveys and research conducted by rheumadocs. Z Rheumatol 2019 Nov;78(9):839-846. [doi: 10.1007/s00393-018-0578-3] [Medline: 30542914]
31. Tanriverdi H, Iacono CS. Diffusion of telemedicine: a knowledge barrier perspective. Telemed J 1999;5(3):223-244. [doi: 10.1089/107830299311989] [Medline: 10908437]
32. Knörr V, Dini L, Gunkel S, Hoffmann J, Mause L, Ohnhäuser T, et al. Use of telemedicine in the outpatient sector during the COVID-19 pandemic: a cross-sectional survey of German physicians. BMC Prim Care 2022 Apr 23;23(1):92 [FREE Full text] [doi: 10.1186/s12875-022-01699-7] [Medline: 35461212]
33. Vossen D, Knitza J, Klemm P, Haase I, Mucke J, Kernder A, et al. Acceptance of video consultation among patients with inflammatory rheumatic diseases depends on gender and location-results of an online survey among patients and physicians. Z Rheumatol (forthcoming) 2021 Aug 27 [FREE Full text] [doi: 10.1007/s00393-021-01052-w] [Medline: 34448915]
34. Peine A, Paffenholz P, Martin L, Dohmen S, Marx G, Loosen SH. Telemedicine in Germany during the COVID-19 pandemic: multi-professional national survey. J Med Internet Res 2020 Aug 05;22(8):e19745 [FREE Full text] [doi: 10.2196/19745] [Medline: 32568724]
35. Alkureishi MA, Choo ZY, Lenti G, Castaneda J, Zhu M, Nunes K, et al. Clinician perspectives on telemedicine: observational cross-sectional study. JMIR Hum Factors 2021 Jul 09;8(3):e29690 [FREE Full text] [doi: 10.2196/29690] [Medline: 34184994]
36. Shiferaw KB, Mehari EA. Internet use and eHealth literacy among health-care professionals in a resource limited setting: a cross-sectional survey. Adv Med Educ Pract 2019 Jul 25;10:563-570 [FREE Full text] [doi: 10.2147/AMEP.S205414] [Medline: 31440113]
37. TechnikRadar 2022. Koerber Stiftung. 2022. URL: https://koerber-stiftung.de/projekte/technikradar/technikradar-2022/ [accessed 2022-06-10]
38. Health Data. Kassenärztliche Bundesvereinigung. URL: https://gesundheitsdaten.kbv.de/cms/html/16396.php [accessed 2022-06-10]
39. Ravindran V, Kataria S. Digital health in rheumatology. Ann Rheum Dis 2019 Oct;78(10):e103. [doi: 10.1136/annrheumdis-2018-214146] [Medline: 30049829]
40. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]

XSL•FO
RenderX

41. Khurana MP, Raaschou-Pedersen DE, Kurtzhals J, Bardram JE, Ostrowski SR, Bundgaard JS. Digital health competencies in medical school education: a scoping review and Delphi method study. BMC Med Educ 2022 Feb 26;22(1):129 [FREE Full text] [doi: 10.1186/s12909-022-03163-7] [Medline: 35216611]
42. Yeoh SA, Young K, Putman M, Sattui S, Conway R, Graef E, Global Rheumatology Alliance. Rapid adoption of telemedicine in rheumatology care during the COVID-19 pandemic highlights training and supervision concerns among rheumatology trainees. ACR Open Rheumatol 2022 Feb;4(2):128-133 [FREE Full text] [doi: 10.1002/acr2.11355] [Medline: 34791821]
43. Almufleh A, Lee C, Tsang MY, Gin K, Tsang TS, Nair P. The need for telemedicine integration into adult cardiology training curricula in Canada. Can J Cardiol 2021 Jun;37(6):929-932. [doi: 10.1016/j.cjca.2021.03.001] [Medline: 33992489]
44. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. There's an app for that, but nobody's using it: insights on improving patient access and adherence to digital therapeutics in Germany. Digit Health 2022 Jul 3;8:20552076221104672 [FREE Full text] [doi: 10.1177/20552076221104672] [Medline: 35811758]
45. Mühlensiepen F, Kurkowski S, Krusche M, Mucke J, Prill R, Heinze M, et al. Digital health transition in rheumatology: a qualitative study. Int J Environ Res Public Health 2021 Mar 05;18(5):2636 [FREE Full text] [doi: 10.3390/ijerph18052636] [Medline: 33807952]
46. Froschauer S, Muth T, Bredow L, Feist E, Heinemann-Dammann SP, Zinke S, et al. Versorgungsatlas rheumatologie : ansätze und konzepte zur verbesserung der versorgung in der ambulanten rheumatologie. Z Rheumatol 2021 Nov;80(9):819-826. [doi: 10.1007/s00393-021-01072-6] [Medline: 34535817]
47. Hinne M, Gronau QF, van den Bergh D, Wagenmakers EJ. A conceptual introduction to Bayesian model averaging. Adv Methods Pract Psychol Sci 2020 Jun 2;3(2):200-215. [doi: 10.1177/2515245919898]
48. 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 Feb;25(1):35-57 [FREE Full text] [doi: 10.3758/s13423-017-1343-3] [Medline: 28779455]
49. Genell A, Nemes S, Steineck G, Dickman PW. Model selection in medical research: a simulation study comparing Bayesian model averaging and stepwise regression. BMC Med Res Methodol 2010 Dec 06;10:108 [FREE Full text] [doi: 10.1186/1471-2288-10-108] [Medline: 21134252]
50. Mu Y, See I, Edwards JR. Bayesian model averaging: improved variable selection for matched case-control studies. Epidemiol Biostat Public Health 2019;16(2):e13048 [FREE Full text] [doi: 10.2427/13048] [Medline: 31772926]
51. Wang D, Zhang W, Bakhai A. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. Stat Med 2004 Nov 30;23(22):3451-3467. [doi: 10.1002/sim.1930] [Medline: 15505893]
52. Lu Z, Lou W. Bayesian approaches to variable selection: a comparative study from practical perspectives. Int J Biostat 2021 Mar 24;18(1):83-108. [doi: 10.1515/ijb-2020-0130] [Medline: 33761580]

Abbreviations

BMA: Bayesian model averaging
CI: credible interval
GP: general practitioner
OR: odds ratio
ROPE: region of practical equivalence
TM use: actual use of telemedicine
TM willingness: willingness to use telemedicine
TM: telemedicine
VIF: variance inflation factor
Q: question

Edited by R Kukafka; submitted 14.06.22; peer-reviewed by S Stones, D Ankersen; comments to author 11.07.22; revised version received 20.07.22; accepted 28.09.22; published 30.11.22.

Please cite as:
Muehlensiepen F, Petit P, Knitza J, Welcker M, Vuillerme N
Factors Associated With Telemedicine Use Among German General Practitioners and Rheumatologists: Secondary Analysis of Data From a Nationwide Survey
J Med Internet Res 2022;24(11):e40304
URL: https://www.jmir.org/2022/11/e40304
doi:10.2196/40304
PMID:36449333

XSL•FO
RenderX

©Felix Muehlensiepen, Pascal Petit, Johannes Knitza, Martin Welcker, Nicolas Vuillerme. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Care and Electronic Patient Communication During COVID-19: Cross-sectional Study of Inequities Across a Canadian Tertiary Cancer Center

Amir H Safavi[1,2], MSc, MD; Mike Lovas[3,4], MASc; Zhihui Amy Liu[5,6], PhD; Sheena Melwani[3], MHI; Tran Truong[7,8], MHSc; Shayla Devonish[3], MHI; Nazek Abdelmutti[5], MSc; Ambreen Sayani[9], MD, PhD; Danielle Rodin[1,2], MPH, MD; Alejandro Berlin[1,2], MSc, MD

[1]Radiation Medicine Program, Princess Margaret Cancer Centre, Toronto, ON, Canada
[2]Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada
[3]Smart Cancer Care Program, Princess Margaret Cancer Centre, Toronto, ON, Canada
[4]Healthcare Human Factors, University Health Network, Toronto, ON, Canada
[5]Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
[6]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[7]Data Science Program, Princess Margaret Cancer Centre, Toronto, ON, Canada
[8]Techna Institute, University Health Network, Toronto, ON, Canada
[9]Women's College Research Institute, Women's College Hospital, Toronto, ON, Canada

Corresponding Author:
Alejandro Berlin, MSc, MD
Radiation Medicine Program
Princess Margaret Cancer Centre
700 University Ave
7th Floor - RMP
Toronto, ON, M5G 2M9
Canada
Phone: 1 4169462983
Email: alejandro.berlin@rmp.uhn.ca

Abstract

Background: Virtual care (VC) visits (telephone or video) and email-based patient communication have been rapidly adopted to facilitate cancer care during the COVID-19 pandemic. Inequities in access and patient experience may arise as these digital health tools become prevalent.

Objective: We aimed to characterize inequities in access and patient-reported experience following adoption of digital health tools at a tertiary cancer center during the COVID-19 pandemic.

Methods: We designed a cross-sectional study of outpatients with visits from September to December 2020. Patient characteristics and responses to an email-based patient-experience survey were collated. Inequities in access were assessed across three pairs of comparison groups: (1) patients with VC and in-person visits, (2) patients with and without documented email addresses, and (3) responders and nonresponders to the survey. Inequities in patient-reported experience were assessed among survey responders. Demographics were mapped to area-level averages from national census data. Socioeconomic status was mapped to area-level dimensions of the Canadian Index of Multiple Deprivation. Covariate balance between comparison groups was assessed using standardized mean differences (SMDs), with SMD≥0.2 indicating differences between groups. Associations between patient experience satisfaction scores and covariates were assessed using multivariable analyses, with $P<.05$ indicating statistical significance.

Results: Among the 42,194 patients who had outpatient visits, 62.65% (n=26,435) had at least one VC visit and 31.15% (n=13,144) were emailable. Access to VC and email was similar across demographic and socioeconomic indices (SMD<0.2). Among emailable patients, 21.84% (2870/13,144) responded to the survey. Survey responsiveness was similar across indices, aside from a small difference by age (SMD=0.24). Among responders, 24.4% received VC and were similar to in-person responders across indices (SMD<0.2). VC and in-person responders had similar satisfaction levels with all care domains surveyed (all $P>.05$).

XSL•FO
RenderX

Regardless of visit type, patients had variable satisfaction with care domains across demographic and socioeconomic indices. Patients with higher ethnocultural composition scores were less satisfied with the cultural appropriateness of their care (odds ratio [OR] 0.70, 95% CI 0.57-0.86). Patients with higher situational vulnerability scores were less satisfied with discussion of physical symptoms (OR 0.67, 95% CI 0.48-0.93). Patients with higher residential instability scores were less satisfied with discussion of both physical (OR 0.81, 95% CI 0.68-0.97) and emotional (OR 0.86, 95% CI 0.77-0.96) symptoms, and also with the duration of their visit (OR 0.85, 95% CI 0.74-0.98; *P*=.02). Male patients were more satisfied with how their health care provider had listened to them (OR 1.64, 95% CI 1.11-2.44; *P*=.01).

Conclusions: Adoption of VC and email can equitably maintain access and patient-reported experience in cancer care across demographics and socioeconomic indices. Existing health inequities among structurally marginalized patients must continue to be addressed to improve their care experience.

(J Med Internet Res 2022;24(11):e39728) doi:10.2196/39728

KEYWORDS

digital health; telehealth; telemedicine; eHealth; oncology; cancer care; virtual care; health inequities; health inequality; digital divide; COVID-19; electronic mail; cross sectional; engagement; satisfaction; patient reported; experience

Introduction

Virtual care (VC), referring to the delivery of care using information and telecommunications technologies [1], has been adopted at cancer centers during the COVID-19 pandemic to promote adherence to physical distancing and other public health measures [2-5]. At the Princess Margaret Cancer Centre (Toronto, Ontario), VC visits were rapidly implemented 12 days after declaration of the pandemic to reduce in-person visits by 50% while maintaining continuity of care [2]. The hospital-wide VC platform developed in-house was also leveraged to email patient-reported experience surveys to all patients after VC or in-person visits, allowing quality improvement data collection to continue during the pandemic [2].

As digital health tools such as VC and email-based patient communications are increasingly adopted, it is possible they may mitigate or exacerbate existing health inequities [6]. Differential access to and benefit from digital resources is termed the digital divide [7-9]. Digital divides are modulated by the social determinants of health, including age, gender, income, housing, rurality, race, and language. Some have noted the potential benefits of VC and email for enhancing equitable care [10-12], including increased health care utilization among racialized minorities and those with travel restrictions. Others have expressed concerns about digital divides in access to VC due to structural marginalization [13-17], particularly among marginalized populations known to be more prone to adverse oncologic outcomes [18,19]. These concerns underscore the importance of investigating inequities following the adoption of digital health tools and addressing their impact on oncologic care [20,21]. Herein, we aimed to characterize inequities in access and patient-reported experience following adoption of digital health tools at a Canadian tertiary cancer center.

Methods

Study Design

A cross-sectional study was designed. Inequities in access were assessed across three pairs of comparison groups: (1) patients with one or more VC visits and those with only in-person visits; (2) emailable and nonemailable patients (effectively, patients with and without documented email addresses); and (3) responders and nonresponders to an email-based, patient-reported experience survey. Inequities in patient-reported experience were assessed among survey responders. We characterized digital divides in access to and use of VC and email using group comparisons by demographics and clinic type. We characterized inequities in patient experience among all patients by identifying associations between satisfaction with care and demographics or visit type. This study was conducted in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines.

Setting and Participants

All patients with outpatient VC or in-person clinic visits at the Princess Margaret Cancer Centre, University Health Network (Toronto, Ontario, Canada) from September 1 through December 31, 2020, were included in this study. This center conducts approximately 2000 outpatient visits daily, including close to 1000 ambulatory clinic encounters and 1000 ambulatory procedures and treatments [2]. We extracted study data, including age, gender, postal codes, clinic type, and visit type (in-person vs VC), and completed patient-reported experience surveys from the electronic medical record system and the hospital-wide VC platform.

Characterization of Demographics

The following demographics were selected for use in this study: age, gender, income, low-income status, area type, and the four dimensions of the Canadian Index of Multiple Deprivation (CIMD), namely residential instability (RI), economic dependency (ED), ethnocultural composition (EC), and situational vulnerability (SV) (Multimedia Appendix 1). Age and gender data were extracted from the electronic medical record, while the other demographics were derived from dissemination area (DA)-level data reported in the Statistics Canada 2016 census. Each individual's DA was captured by linking their postal code to Statistics Canada 2016 census data using Postal Code Conversion File (PCCF+) version 7B.

Individual patient income was estimated using the neighborhood income per single-person equivalent, a household size–adjusted measure of household income (before tax), based on 2016 Census Summary profile data at the DA level. Low-income

XSL•FO
RenderX

status was determined by comparing neighborhood income per single-person equivalent to Statistics Canada low-income cutoffs in 2016, based on area type (rural vs small vs medium vs large population center) and family size of one person. Per Statistics Canada, rural was defined as a population less than 1000, small population was defined as 1-29,999, medium population was defined as 30-99,999, and large population was defined as 100,000 or more. Socioeconomic status was mapped to the four DA-level dimensions of the CIMD (RI, ED, EC, and SV). For each dimension, we provided the CIMD in two forms: factor scores (higher scores correspond to more marginalized areas) and quintiles (a value of 1 corresponds to the least deprived area and 5 corresponds to the most deprived area). The constituent elements of each CIMD dimension are significantly correlated with only one dimension and are described in Multimedia Appendix 1.

Patient-Reported Experience Survey

We emailed an adapted version of Your Voice Matters [22], a validated patient-reported experience survey provincially mandated by Ontario Health, to all patients with documented email addresses at the Princess Margaret Cancer Centre after each outpatient VC or in-person clinic visit starting in September 2020. The survey was available in eight languages: English, French, Simplified Chinese, Traditional Chinese, Spanish, Portuguese, Italian, and Vietnamese. The adapted survey included new questions regarding the utilization of VC [23] (Multimedia Appendix 1, section 1.2; Q3-5), as well as existing questions regarding satisfaction with various care domains such as discussion of physical and emotional symptoms and cultural appropriateness of care. The survey is found in Multimedia Appendix 1 (section 1.2).

Statistical Methods

Summary statistics were calculated to describe demographics and clinic types of the full cohort and comparison groups. Categorical variables are summarized as numbers (percentages) and continuous variables are summarized as means (SD), medians (IQR), deciles (for income), and quintiles (for RI, ED, EC, and SV). The first completed survey from each respondent was utilized for this analysis. A sensitivity analysis was performed to assess the influence of intrapatient correlation among all survey responses (if there were multiple responses from the same respondent) on the overall results. Within survey responders, a large proportion reported their satisfaction as 4 or 5 out of 5, and therefore responses were dichotomized for each question as “satisfied” (reported 4 or 5) and “not satisfied” (reported 1, 2, or 3). Standardized mean differences (SMDs) were calculated for group comparisons by demographic variable and clinic types; SMDs of 0.2, 0.5, and 0.8 were considered small, medium, and large differences, respectively. Multivariable logistic regression was used to assess associations between satisfaction scores and demographic variables, as well as visit type, with income on a log scale. Statistical significance was judged at $P<.05$. Complete case analyses were performed to address missing survey responses. To correct for bias due to nonresponse, multivariable models for outcome variables were fitted with potential predictors of nonresponse as covariates [24]. Analyses were performed using R version 4.0.2 (R Foundation for Statistical Computing).

Ethics Statement

This study was reviewed by the institutional research ethics board; ethics approval was obtained, along with a waiver for written informed consent (University Health Network Quality Improvement Review Committee #21-0148).

Results

Participant Characteristics

From September 2020 through December 2020, 42,194 patients had outpatient clinic visits (Table 1). The median age of the full cohort was 64 (IQR 52-73) years and 51.7% of the patients self-identified as male. The majority of patients lived in a large urban population center. The mean income of the full cohort was CAD $62,400 (SD CAD $27,700), with CAD $1=~US $0.75. Approximately 7% of patients were classified in a low-income category. For the full cohort, the most common clinical specialties visited were genitourinary, head and neck, and gastrointestinal oncology. Full cohort characteristics are listed in Table S1 of Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Patient demographic characteristics of the full cohort (N=42,194).

Characteristics	Value
Age (years)	
Mean (SD)	61.6 (15.5)
Median (IQR)	64 (52 to 73)
Gender, n (%)	
Female	20,388 (48.32)
Male	21,806 (51.68)
Area[a], n (%)	
Rural area	2830 (6.71)
Small population center	2127 (5.04)
Medium population center	1494 (3.54)
Large urban population center	35,743 (84.71)
BTIPPE[b] (CAD $[c])×1000	
Mean (SD)	62.4 (27.7)
Median (IQR)	58.4 (46.2 to 72.6)
1st decile, n (%)	3594 (8.52)
2nd decile, n (%)	3668 (8.69)
3rd decile, n (%)	3888 (9.21)
4th decile, n (%)	3925 (9.30)
5th decile, n (%)	3795 (8.99)
6th decile, n (%)	3899 (9.24)
7th decile, n (%)	3750 (8.89)
8th decile, n (%)	4103 (9.72)
9th decile, n (%)	4822 (11.43)
10th decile, n (%)	6750 (16.00)
Low income[d], n (%)	
No	39,173 (92.84)
Yes	3021 (7.16)
Residential instability[e]	
Score, mean (SD)	0.3 (1.2)
Score, median (IQR)	–0.1 (–0.6 to 1.2)
1st quintile, n (%)	6747 (15.99)
2nd quintile, n (%)	7025 (16.65)
3rd quintile, n (%)	7100 (16.83)
4th quintile, n (%)	7500 (17.78)
5th quintile, n (%)	13,822 (32.76)
Economic dependency[e]	
Score, mean (SD)	–0.1 (1.1)
Score, median (IQR)	–0.2 (–0.8 to 0.4)
1st quintile, n (%)	10,764 (25.51)
2nd quintile, n (%)	8685 (20.58)
3rd quintile, n (%)	7983 (18.92)

XSL•FO
RenderX

Characteristics	Value
4th quintile, n (%)	7398 (17.53)
5th quintile, n (%)	7364 (17.45)
Ethnocultural composition[e]	
Score, mean (SD)	0.5 (1.1)
Score, median (IQR)	0.3 (–0.4 to 1.2)
1st quintile, n (%)	2663 (6.31)
2nd quintile, n (%)	4616 (10.94)
3rd quintile, n (%)	7606 (18.03)
4th quintile, n (%)	12,195 (28.90)
5th quintile, n (%)	15,114 (35.82)
Situational vulnerability[e]	
Score, mean (SD)	–0.3 (0.8)
Score, median (IQR)	–0.5 (–0.8 to 0.1)
1st quintile, n (%)	14,348 (34.00)
2nd quintile, n (%)	9219 (21.85)
3rd quintile, n (%)	7277 (17.25)
4th quintile, n (%)	6126 (14.52)
5th quintile, n (%)	5224 (12.38)
Clinic type, n (%)	
Genitourinary	8885 (21.06)
Head and neck	4594 (10.89)
Gastrointestinal	4220 (10.00)
Breast	3478 (8.24)
Gynecologic	3316 (7.86)
Other	17,701 (41.95)

[a]Rural area was defined as a population of less than 1000, small population was defined as 1-29,999 people, medium population was defined as 30-99,999 people, and large population was defined as 100,000 or more people.

[b]BTIPPE: before-tax neighborhood income per single-person equivalent.

[c]CAD $1=~US $0.75.

[d]Low-income status refers to neighborhood income per single-person equivalent below the Statistics Canada low-income cutoffs in 2016, based on area type (rural vs small vs medium vs large population center) and family size of one person.

[e]Higher factor scores and quintiles correspond to more marginalized areas.

Among the full cohort (N=42,194), 26,435 patients (62.65%) had at least one VC visit and 13,144 patients (31.15%) were emailable (Table 2). Among the emailable patients, 2870 (21.84%) responded to the survey. The majority of patients (97%) completed the survey once; the first or only completed surveys were included in the subsequent analysis. Sensitivity analysis accounting for multiple completed surveys by the minority of patients (3%) did not identify undue influence from intrapatient correlation. The percentage of missing responses was low (<5%) for most survey questions (Tables S6-S10 and S14 of Multimedia Appendix 1). Almost all responders (97.9%) completed the survey in English (Table S5 of Multimedia Appendix 1). Among responders, 36.1% reported that they were explicitly provided the option to have an in-person or VC visit (Table S9 of Multimedia Appendix 1).

In-person and VC visits were reported by 73.2% and 24.4% of respondents, respectively (Table 3). A visit type was not reported by 2.4% of responders. Among VC respondents, 84.6% reported having a phone visit, while 15.4% reported having a video visit (Table S9 of Multimedia Appendix 1). Additional group characteristics are found in Table 2, Table 3, and Multimedia Appendix 1 (section 1.3).

XSL•FO
RenderX

Table 2. Effect size measurements of differences by visit type, access to email, and survey responsiveness.

Characteristics	Visit type			Access to email			Survey responsiveness		
	In-person only (n=15,759)	≥1 virtual visit (n=26,435)	SMD[a]	Nonemailable (n=29,050)	Emailable (n=13,144)	SMD	Nonresponder (n=10,274)	Responder (n=2870)	SMD
Age									
Mean (SD)	60.9 (15.4)	62 (15.5)	0.072	61.9 (15.6)	60.9 (15.4)	0.063	60.1 (15.8)	63.7 (13.4)	0.241
Median (IQR)	63.0 (52.0-72.0)	64.0 (53.0-73.0)	—[b]	64 (53-73)	63 (52-72)	—	62 (51-71)	65 (56-73)	—
Gender, n (%)									
Female	8353 (53)	12,035 (45)	0.15	13,798 (47)	6590 (50)	0.053	5077 (49)	1513 (53)	0.066
Male	7406 (47)	14,400 (55)	—	15,252 (53)	6554 (50)	—	5197 (51)	1357 (47)	—
Area type	—	—	0.097	—	—	0.047	—	—	0.008
BTIPPE[c] (CAD $[d])×1000									
Mean (SD)	60.9 (26.1)	63.3 (28.6)	0.09	62 (27.7)	63.3 (27.8)	0.045	62.5 (27.4)	66.3 (29)	0.137
Median (IQR)	57.6 (45.4-70.8)	58.8 (46.5-73.4)	—	58.1 (45.8-72.1)	59 (46.8-73.4)	—	58.4 (46.3-72.6)	61.4 (49-76.5)	—
Decile distribution	—	—	0.088	—	—	0.068	—	—	0.182
Low income[e], n (%)									
No	14,816 (94)	24,357 (92)	0.074	26,993 (93)	12,180 (93)	0.01	9552 (93)	2628 (92)	0.053
Yes	943 (6)	2078 (8)	—	2057 (7)	964 (7)	—	722 (7)	242 (8)	—
Residential instability[f]									
Score, mean (SD)	0.3 (1.2)	0.3 (1.2)	0.055	0.3 (1.2)	0.3 (1.2)	0.021	0.3 (1.2)	0.4 (1.2)	0.023
Quintile distribution	—	—	0.071	—	—	0.022	—	—	0.046
Economic dependency[f]									
Score, mean (SD)	–0.1 (1.1)	–0.1 (1.1)	0.009	–0.1 (1.1)	–0.1 (1)	0.046	–0.1 (1)	–0.1 (1.1)	0.045
Quintile distribution	—	—	0.015	—	—	0.049	—	—	0.05
Ethnocultural composition[f]									
Score, mean (SD)	0.5 (1.1)	0.5 (1)	0.012	0.5 (1.1)	0.5 (1)	0.006	0.5 (1.1)	0.3 (1)	0.186
Quintile distribution	—	—	0.081	—	—	0.057	—	—	0.177
Situational vulnerability[f]									
Score, mean (SD)	–0.3 (0.8)	–0.3 (0.8)	0.069	–0.3 (0.8)	–0.4 (0.7)	0.078	–0.3 (0.8)	–0.4 (0.7)	0.156
Quintile distribution	—	—	0.07	—	—	0.082	—	—	0.15
Clinic type	—	—	1.112	—	—	0.612	—	—	0.247

[a]SMD: standardized mean difference.

[b]Not applicable.

[c]BTIPPE: before-tax neighborhood income per single-person equivalent.

[d]CAD $1=~US $0.75.

[e]Low-income status refers to neighborhood income per single-person equivalent below the Statistics Canada low-income cutoffs in 2016, based on area type (rural vs small vs medium vs large population center) and family size of one person.

[f]Higher factor scores and quintiles for correspond to more marginalized areas.

XSL•FO
RenderX

Table 3. Effect size of differences by visit type among survey responders.

Characteristics	In-person visit (n=2101)	Virtual visit (n=700)	SMD[a]
Age			
Mean (SD)	63.3 (13.7)	64.8 (13)	0.109
Median (IQR)	65 (56-73)	67 (58-74)	—[b]
Gender, n (%)			
Female	1163 (55.4)	319 (45.6)	0.197
Male	938 (44.6)	381 (54.4)	—
Area type	—	—	0.093
BTIPPE[c] (CAD $[d])×1000			
Mean (SD)	65.7 (28.3)	69.1 (30.9)	0.116
Median (IQR)	61 (48.4-76.5)	64 (51.5-77.2)	—
Decile distribution	—	—	0.176
Low income[e], n (%)			
No	1931 (91.9)	633 (90.4)	0.052
Yes	170 (8.1)	67 (9.6)	
Residential instability[f]			
Score, mean (SD)	0.4 (1.2)	0.3 (1.2)	0.06
Quintile distribution	—	—	0.097
Economic dependency[f]			
Score, mean (SD)	–0.1 (1.1)	–0.1 (1.1)	0.034
Quintile distribution	—	—	0.1
Ethnocultural composition[f]			
Score, mean (SD)	0.4 (1)	0.2 (0.9)	0.183
Quintile distribution	—	—	0.18
Situational vulnerability			
Score, mean (SD)	–0.4 (0.7)	–0.5 (0.7)	0.093
Quintile distribution	—	—	0.109
Clinic type	—	—	0.646

[a]SMD: standardized mean difference.

[b]Not applicable.

[c]BTIPPE: before-tax neighborhood income per single-person equivalent.

[d]CAD $1=~US $0.75.

[e]Low-income status refers to neighborhood income per single-person equivalent below the Statistics Canada low-income cutoffs in 2016, based on area type (rural vs small vs medium vs large population center) and family size of one person.

[f]Higher factor scores and quintiles for correspond to more marginalized areas.

Characterization of Digital Divides in Access to and Use of VC and Email

VC and in-person patients were similar across demographics, with negligible differences in age, gender, area, and income, as well as deprivation indices, including RI, ED, EC, and SV (Table 2). Clinic types differed between VC and in-person patients (SMD=1.112); the most common clinic types among VC patients were genitourinary (29.5%), gastrointestinal (11.0%), and breast (10.3%) oncology (Multimedia Appendix 1, Table S11). Emailable and nonemailable patients were similar across demographics (SMD<0.2). Clinic types differed between emailable and nonemailable patients (SMD=0.612); the most common clinic types among emailable patients were genitourinary (16.6%), gastrointestinal (11.0%), and breast (10.9%) oncology (Multimedia Appendix 1, Table S12). Survey responders and nonresponders were similar across demographics, aside from a small difference in age (SMD=0.241). Clinic types differed between responders and nonresponders (SMD=0.247); the most common clinic types

XSL•FO
RenderX

among responders were genitourinary (15.1%), gynecologic (12.4%), and breast (12.2%) oncology (Multimedia Appendix 1, Table S13). VC and in-person respondents were also similar across demographics (SMD<0.2) (Table 3). Clinic types differed between VC and in-person respondents (SMD=0.646); the most common clinic types among VC respondents were genitourinary (29.1%), breast (11.9%), and lymphoma (9.4%) oncology (Multimedia Appendix 1, Table S14).

Inequities in Patient Experiences Across Visit Types

VC and in-person respondents had similar satisfaction levels with all care domains surveyed (all *P*>.05; Table 4), although VC respondents were less satisfied with their experience overall compared to in-person respondents (Table 4). Regardless of visit type, structurally marginalized patients were less satisfied with their care (Figure 1; Multimedia Appendix 1 Tables S15-S24). Patients with higher EC scores were less likely to characterize their care as culturally appropriate (odds ratio [OR] 0.7, 95% CI 0.57-0.89; *P*<.001). Patients with higher SV scores were less satisfied with discussion of physical symptoms (OR 0.67, 95% CI 0.48-0.93; *P*=.02). Patients with higher RI scores were less satisfied with discussion of physical (OR 0.81, 95% CI 0.68-0.97; *P*=.02) and emotional (OR 0.86, 95% CI 0.77-0.96; *P*=.009) symptoms. Patients with higher RI scores were also less satisfied that enough time had been spent with them during their visit (OR 0.85, 95% CI 0.74-0.98; *P*=.02). Male patients were more satisfied with how their health care provider had listened to them (OR 1.64, 95% CI 1.11-2.44; *P*=.01). Older patients were more satisfied with six of nine care domains surveyed.

Table 4. Associations between visit type (virtual vs in-person) and satisfaction with care quality domains.

Domain	OR[a] (95% CI)[b]	*P* value
HCP[c] listened to what you had to say?	0.83 (0.55-1.27)	.40
HCP discussed any of your physical symptoms?	0.93 (0.62-1.39)	.72
HCP discussed any of your emotional worries or concerns?	1.12 (0.87-1.45)	.37
HCP spent enough time with you?	1 (0.73-1.36)	.98
HCP let you ask questions?	0.88 (0.6-1.29)	.51
HCP explained things in a way you could easily understand?	1.07 (0.72-1.6)	.74
HCP involved you in decisions (choices) about your care in the way that you wanted?	1.09 (0.76-1.54)	.65
HCP provided care that you felt was appropriate given your ethnic/cultural background?	0.97 (0.62-1.51)	.88
HCP treated you with respect?	1.23 (0.69-2.2)	.48
Overall experience at your last visit?	0.68 (0.49-0.94)	.02

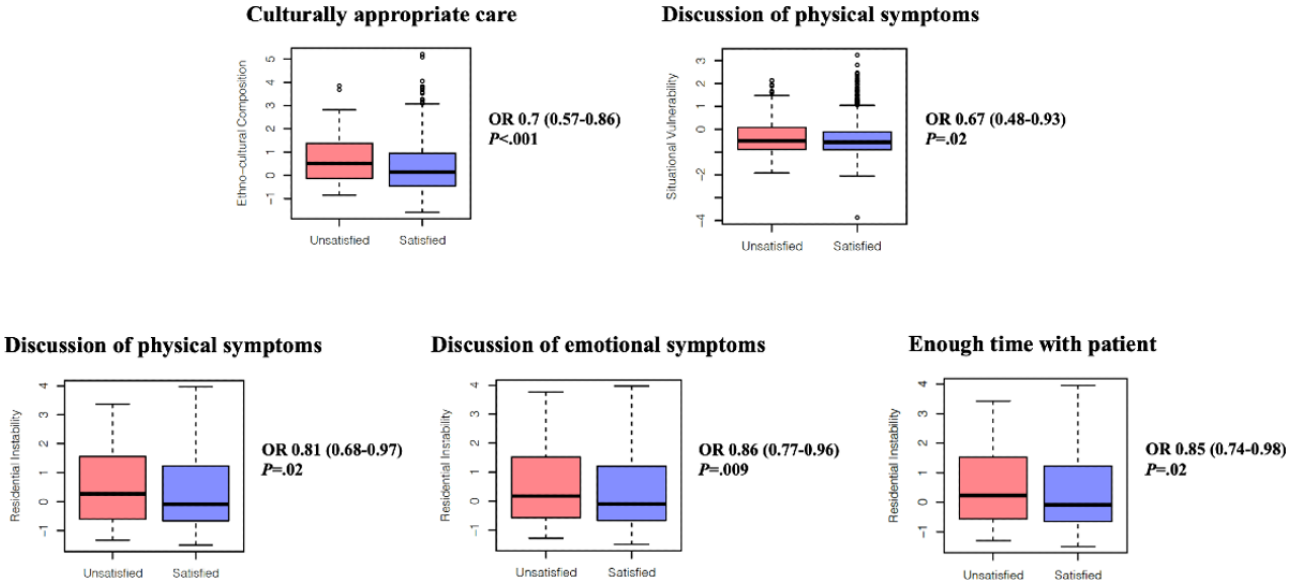
[a]OR: odds ratio.

[b]Values obtained from multivariable analyses of associations between demographics/visit type and satisfaction with surveyed care domains (see Multimedia Appendix 1 page 17-20); an odds ratio greater than 1 corresponds with greater satisfaction among patients with virtual care visits than those with in-person visits.

[c]HCP: health care provider.

XSL•FO
RenderX

Figure 1. Associations between demographics and satisfaction with care quality domains. Values obtained from multivariable analyses of associations between demographics/visit type and satisfaction with surveyed care domains (see Multimedia Appendix 1 page 17-20). An odds ratio (OR)>1 for residential instability, ethnocultural composition, and situational vulnerability corresponds to increasing satisfaction with increasing marginalization. Box plots display the median value (central bar) and the IQR (lower and upper hinges denote the 75th and 25th percentiles, respectively). Whiskers indicate the minimum and maximum values within a distance of 1.5 times the IQR from the lower and upper hinges, respectively. Outlier data beyond the boundary of the whiskers are depicted by circles.



Discussion

Principal Results

Our study quantitatively shows that VC and email are digital health tools that can equitably maintain patient access and experience at a publicly funded tertiary cancer center. Among the 42,194 patients analyzed, VC and email accessibility and survey responsiveness were similar across demographics and socioeconomic indices. These findings are encouraging, as continuing adoption of digital health tools in cancer care may improve clinical trial enrollment [20,25], patient-reported outcome collection [20,26], and timely participation in expert supportive care among underserved populations. Regardless of visit type, patients structurally marginalized by ethnocultural, situational, and residential status, as well as gender, were less satisfied with their care. These results reinforce the reality that social determinants of health have tangible impacts on the patient experience, and necessitate further characterization using targeted questionnaires and focus groups of patients, community members, and advocates, and engagement with these stakeholders in designing solutions for mitigating these inequities. These outcomes may also suggest that for the majority of patients, pre-existing social inequities, rather than the adoption of VC, may contribute to worse patient experiences. Thus, our work highlights the imperative for proactive and continuous quantification of social determinants of health to improve equitability of the patient experience.

Comparison With Prior Work

Our work presents findings that may appear incongruent with other contemporary published studies. A recent analogous study based on administrative claims data from the United States found differential use of telemedicine by socioeconomic status among 16,006 newly diagnosed cancer patients during the COVID-19 pandemic [27]. Compared to our findings, the different conclusions in this study may be attributable to the differences between publicly funded and private health care ecosystems, regionality (large Canadian urban setting vs United States–wide), as well as differences in the demographic and oncologic characteristics, which together determine the systemic inequities faced by the patient populations in these studies. These factors highlight the challenges of generalizing findings between health care settings. Nonetheless, we believe that the granularity of demographic characterization provided in our study may promote transferability of findings to other tertiary cancer centers located in large, diverse urban centers.

Future Directions

Some findings of our study may warrant further investigation. First, the majority (83.7%) of VC was provided by phone instead of by video. Here, we did not include a comparison between patients with phone and video visits, as a robust assessment of factors influencing the allocation of VC modalities would also require characterization of provider and disease characteristics. This future analysis is required, as previous work has suggested that marginalized patients participate in video visits less often than nonmarginalized peers due to limitations of technological access and literacy [15]. Second, we found differential access to and use of email and VC by clinic type, echoing findings of other studies [16]. While differential VC use by clinic type may be attributable to clinical reasoning regarding the added benefit of in-person oncologic assessment to VC-amenable biochemical and/or radiographic surveillance in certain clinical contexts, differential access to email and survey responsiveness by clinic type are agnostic to provider preferences and may be indicative of unaccounted intersectional digital divides among subpopulations with lower socioeconomic status [28]. Thus, additional work is required to characterize differences in uptake of email and VC across disease sites, and identify inequitable factors serving as barriers to access and use among specific

XSL•FO
RenderX

patient populations. Third, we found that increasing age was associated with greater survey responsiveness and increased satisfaction with several care domains surveyed. This result challenges conventional assumptions about the engagement and satisfaction of older patients with digital health tools. Additional work is required to characterize these age-related differences and tailor custom interventions to improve the patient experience across age groups.

Limitations

Our study design is dependent on patients having documented postal codes, as well as Statistics Canada's definitions of deprivation indices. As such, these results, while representative at scale, likely do not reflect the experience of populations who are living on societal margins due to precarious housing. Study of these experiences will require targeted engagement with patients and advocates to gather qualitative and quantitative data about their experiences with email and VC use in the health care setting. Our study has additional limitations to consider. First, patients were not stratified by access to and literacy with technologies required for VC and email, such as personal computers, phones, and high-speed home internet; as a result, our study does not account for the impact of these factors on utilization of and satisfaction with digital health tools in cancer care. Second, although the proportion of individuals without knowledge of English is incorporated in the EC index of the CIMD and language of survey completion was collected, our study's methodology precluded explicit characterization of the linguistic literacy of nonemailable patients and nonresponders; as such, this may be an unaccounted driver of digital divides. Third, our analysis is not intersectional. Individuals occupying intersecting social identities may have different experiences than members of each individual demographic group they may belong to [29], and thus they may be subject to unique digital divides not captured in our study. Fourth, the adapted patient-reported experience survey used in this study has not been validated. The original survey is a validated instrument [22], and questions added to the survey regarding VC are unlikely to impact its validity [23] (Multimedia Appendix 1; section 1.2, Q3-5); nonetheless, this potential limitation can be noted. Although more work is needed to identify the full scope of digital divides, our study provides encouraging evidence that the rapid systemic adoption of digital health tools during the COVID-19 pandemic equitably maintained access to, use of, and satisfaction with health care participation among numerous demographic indices.

Conclusions

Our cross-sectional study showed that VC and email are digital health tools that can maintain patient access and experience across patient demographics, which are similar regardless of emailability, digital survey responsiveness, and visit type. Although satisfaction is similar among VC and in-person patients, patients structurally marginalized by ethnocultural, situational, and residential status remain less satisfied with their care. To increase equitable participation in cancer care, digital health tools should be carefully deployed in concert with targeted interventions designed to further characterize the experiences of structurally marginalized patients, proactively identify at-risk patients, and implement practical solutions.

Acknowledgments

SM has received funding from the Princess Margaret Cancer Foundation (University Health Network).

Authors' Contributions

AS contributed to the conception and design of the work, acquisition and analysis of the data, and drafting of the manuscript. ML contributed to the conception and design of the work and the acquisition of the data. ZAL contributed to the acquisition and analysis of the data. SM and AB contributed to the conception and design of the work. All of the authors interpreted the data, revised the manuscript critically for important intellectual content, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Four dimensions of Canadian Index of Multiple Deprivation, Your Voice Matters survey, and comprehensive study data (Tables S1-S24).
[DOCX File , 109 KB - jmir_v24i11e39728_app1.docx]

References

1. Pan American Health Organization. Framework for the implementation of a telemedicine service. Washington, DC: Pan American Health Organization; 2016.
2. Berlin A, Lovas M, Truong T, Melwani S, Liu J, Liu ZA, et al. Implementation and outcomes of virtual care across a tertiary cancer center during COVID-19. JAMA Oncol 2021 Apr 01;7(4):597-602 [FREE Full text] [doi: 10.1001/jamaoncol.2020.6982] [Medline: 33410867]

XSL•FO
RenderX

3. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly T, et al. Trends in the use of telehealth during the emergence of the COVID-19 pandemic - United States, January-March 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 30;69(43):1595-1599. [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]
4. Liu R, Sundaresan T, Reed ME, Trosman JR, Weldon CB, Kolevska T. Telehealth in oncology during the COVID-19 outbreak: bringing the house call back virtually. JCO Oncol Pract 2020 Jun;16(6):289-293. [doi: 10.1200/OP.20.00199] [Medline: 32364826]
5. Schrag D, Hershman DL, Basch E. Oncology practice during the COVID-19 pandemic. JAMA 2020 May 26;323(20):2005-2006. [doi: 10.1001/jama.2020.6236] [Medline: 32282023]
6. Eyrich NW, Andino JJ, Fessell DP. Bridging the digital divide to avoid leaving the most vulnerable behind. JAMA Surg 2021 Aug 01;156(8):703-704. [doi: 10.1001/jamasurg.2021.1143] [Medline: 33769441]
7. van Dijk J. The digital divide. Cambridge, UK: Polity; 2020.
8. Hargittai E. Second-level digital divide: differences in people's online skills. First Monday 2002 Apr;7(4):- [FREE Full text] [doi: 10.5210/fm.v7i4.942]
9. van Deursen AJ, van Dijk JA. The digital divide shifts to differences in usage. New Media Soc 2013 Jun 07;16(3):507-526. [doi: 10.1177/1461444813487959]
10. Qian L, Sy LS, Hong V, Glenn SC, Ryan DS, Morrissette K, et al. Disparities in outpatient and telehealth visits during the COVID-19 pandemic in a large integrated health care organization: retrospective cohort study. J Med Internet Res 2021 Sep 01;23(9):e29959 [FREE Full text] [doi: 10.2196/29959] [Medline: 34351865]
11. Hoehn RS, Zureikat AH. Cancer disparities in the COVID-19 era. J Surg Oncol 2020 Sep;122(3):371-372 [FREE Full text] [doi: 10.1002/jso.26043] [Medline: 32452031]
12. Shaverdian N, Gillespie EF, Cha E, Kim SY, Benvengo S, Chino F, et al. Impact of telemedicine on patient satisfaction and perceptions of care quality in radiation oncology. J Natl Compr Canc Netw 2021 Jan 04;19(10):1174-1180 [FREE Full text] [doi: 10.6004/jnccn.2020.7687] [Medline: 33395627]
13. Perumalswami CR, Chen E, Martin C, Goold SD, De Vries R, Griggs JJ, et al. "I'm being forced to make decisions I have never had to make before": Oncologists' experiences of caring for seriously ill persons with poor prognoses and the dilemmas created by COVID-19. JCO Oncol Pract 2022 Jan;18(1):e89-e97. [doi: 10.1200/OP.21.00119] [Medline: 34324387]
14. Manz C, Baxter NN, duPont NC, Markham MJ, Drumheller C, Durakovic L, et al. Patterns of telehealth utilization during the COVID-19 pandemic and preferences for post-pandemic telehealth use: A national survey of oncology clinicians. J Clin Oncol 2021 May 20;39(15_suppl):1580-1580. [doi: 10.1200/jco.2021.39.15_suppl.1580]
15. Jewett P, Vogel RI, Ghebre RG, Rao A, Hui JYC, Parsons H, et al. Telehealth: reducing or increasing cancer care disparities? J Clin Oncol 2021 May 20;39(15_suppl):1582. [doi: 10.1200/jco.2021.39.15_suppl.1582]
16. Marks VA, Hsiang WR, Nie J, Umer W, Haleem A, Galal B, et al. Accessibility of telehealth services for cancer care at cancer hospital in the United States. J Clin Oncol 2021 May 20;39(15_suppl):6535. [doi: 10.1200/JCO.2021.39.15_suppl.6535]
17. Greenberg-Worisek A, Ferede L, Balls-Berry J, Marigi I, Valentin Mendez E, Bajwa N, et al. Differences in electronic personal health information tool use between rural and urban cancer patients in the United States: secondary data analysis. JMIR Cancer 2020 Aug 10;6(2):e17352 [FREE Full text] [doi: 10.2196/17352] [Medline: 32773369]
18. Singh GK, Jemal A. Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950-2014: over six decades of changing patterns and widening inequalities. J Environ Public Health 2017;2017:2819372. [doi: 10.1155/2017/2819372] [Medline: 28408935]
19. Ward E, Jemal A, Cokkinides V, Singh GK, Cardinez C, Ghafoor A, et al. Cancer disparities by race/ethnicity and socioeconomic status. CA Cancer J Clin 2004 Mar 01;54(2):78-93 [FREE Full text] [doi: 10.3322/canjclin.54.2.78] [Medline: 15061598]
20. Elston Lafata J, Smith AB, Wood WA, Fitzpatrick B, Royce TJ. Virtual visits in oncology: enhancing care quality while designing for equity. JCO Oncol Pract 2021 May;17(5):220-223. [doi: 10.1200/OP.20.00645] [Medline: 33539180]
21. Zon RT, Kennedy EB, Adelson K, Blau S, Dickson N, Gill D, et al. Telehealth in oncology: ASCO standards and practice recommendations. JCO Oncol Pract 2021 Sep;17(9):546-564. [doi: 10.1200/OP.21.00438] [Medline: 34319760]
22. Your Voice Matters: we are listening; Frequently asked questions – a resource for clinicians and cancer centre staff. Ontario Health. URL: https://www.cancercareontario.ca/sites/ccocancercare/files/assets/YVM%20Clinician%20FAQ.pdf [accessed 2022-09-17]
23. Frequently asked questions about the "Your Voice Matters" survey – information for clinicians and staff. Ontario Health. URL: https://www.cancercareontario.ca/sites/ccocancercare/files/assets/YVM_FAQ_Staff.docx [accessed 2022-09-18]
24. Gelman A. Struggles with survey weighting and regression modeling. Statist Sci 2007 May 1;22(2):153-164. [doi: 10.1214/088342306000000691]
25. Waterhouse DM, Harvey RD, Hurley P, Levit LA, Kim ES, Klepin HD, et al. Early impact of COVID-19 on the conduct of oncology clinical trials and long-term opportunities for transformation: findings From an American Society of Clinical Oncology Survey. JCO Oncol Pract 2020 Jul 12;16(7):417-421. [doi: 10.1200/op.20.00275]

XSL•FO
RenderX

26. Patt D, Wilfong L, Hudson KE, Patel A, Books H, Pearson B, et al. Implementation of electronic patient-reported outcomes for symptom monitoring in a large multisite community oncology practice: dancing the Texas two-step through a pandemic. JCO Clin Cancer Inform 2021 May;5:615-621 [FREE Full text] [doi: 10.1200/CCI.21.00063] [Medline: 34085537]
27. Katz AJ, Haynes K, Du S, Barron J, Kubik R, Chen RC. Evaluation of telemedicine use among US patients with newly diagnosed cancer by socioeconomic status. JAMA Oncol 2022 Jan 01;8(1):161-163. [doi: 10.1001/jamaoncol.2021.5784] [Medline: 34792526]
28. Conway DI, McMahon AD, Smith K, Black R, Robertson G, Devine J, et al. Components of socioeconomic risk associated with head and neck cancer: a population-based case-control study in Scotland. Br J Oral Maxillofac Surg 2010 Jan;48(1):11-17. [doi: 10.1016/j.bjoms.2009.03.020] [Medline: 19481316]
29. Crenshaw K. Demarginalizing the intersection of race and sex: a Black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics [1989]. In: Bartlett K, Kennedy R, editors. Feminist legal theory. New York: Routledge; 1991:24.

Abbreviations

CIMD: Canadian Index of Multiple Deprivation
DA: dissemination area
EC: ethnocultural composition
ED: economic dependency
OR: odds ratio
PCCF+: Postal Code Conversion File
RI: residential instability
SMD: standardized mean difference
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
SV: situational vulnerability
VC: virtual care

Edited by G Eysenbach; submitted 20.05.22; peer-reviewed by S Xu, P Li, B Wolfe-Piggott; comments to author 30.08.22; revised version received 18.09.22; accepted 20.10.22; published 04.11.22.

Please cite as:
Safavi AH, Lovas M, Liu ZA, Melwani S, Truong T, Devonish S, Abdelmutti N, Sayani A, Rodin D, Berlin A
Virtual Care and Electronic Patient Communication During COVID-19: Cross-sectional Study of Inequities Across a Canadian Tertiary Cancer Center
J Med Internet Res 2022;24(11):e39728
URL: https://www.jmir.org/2022/11/e39728
doi:10.2196/39728
PMID:36331536

©Amir H Safavi, Mike Lovas, Zhihui Amy Liu, Sheena Melwani, Tran Truong, Shayla Devonish, Nazek Abdelmutti, Ambreen Sayani, Danielle Rodin, Alejandro Berlin. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Socioeconomic Indicators Linked to Parent Health-Related Technology Use: Cross-sectional Survey

Madison P McCall[1], BS; Megan T Hineline[2], BA; Margaret T Anton[3], PhD; April Highlander[1], MA; Deborah J Jones[1], PhD

[1]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Department of Psychology, Appalachian State University, Boone, NC, United States

[3]AbleTo, Inc, New York City, NY, United States

Corresponding Author:
Madison P McCall, BS
Department of Psychology and Neuroscience
University of North Carolina at Chapel Hill
235 E Cameron Avenue
Chapel Hill, NC, 27599
United States
Phone: 1 919 843 2351
Email: mmccall@unc.edu

Abstract

Background: Despite the prevalence of parent health information seeking on the internet and its impact on parenting behavior, there is a paucity of research on parents of young children (ages 3 to 8 years). Given the importance of this developmental period, exploring how family socioeconomic indicators linked to the digital divide and health inequities affect parent proxy- and self-seeking is critical to further understanding variability in health information seeking and associated outcomes.

Objective: This study aimed to explore parental health-related technology use (HTU), the process by which parents engage in support, advice, and information-seeking behavior related to their (self-seeking) and their children's (proxy seeking) health across a range of hardware devices (eg, tablet, wearable, smartphone, laptop, and desktop computer) and sources (eg, search engines, mobile applications, social media, and other digital media).

Methods: A cross-sectional study including 313 parents and guardians of children ages 3 to 8 years recruited through Amazon Mechanical Turk (MTurk) was conducted. Parents were asked to complete a self-administered questionnaire on a broad range of parenting and parent-related constructs, including sociodemographic information, technology device ownership, and engagement in and use, features, and perceptions of HTU. Descriptive and bivariate analyses (chi-square tests) were performed to identify patterns and investigate associations between family socioeconomic indicators and parent HTU.

Results: The overwhelming majority (301/313, 96%) of parents of young children reported engaging in HTU, of which 99% (300/301) reported using search engines (eg, Google), followed by social media (62%, 188/301), other forms of digital media (eg, podcasts; 145/301, 48%), and mobile applications (114/301, 38%). Parents who engaged in HTU reported seeking information about their child's behavior and discipline practices (260/313, 83%), mental or physical health (181/313, 58%), and academic performance (142/313, 45%). Additionally, nearly half (134/313, 43%) of parents reported searching for advice on managing their stress. Among parents who reported using each source, an overwhelming majority (280/300, 93%) indicated that search engines were a helpful online source for proxy- and self-seeking, followed by social media (89%, 167/188), other digital media (120/145, 83%), and mobile apps (87/114, 76%). Among parents who reported using any technology source, approximately one-fifth reported that technology sources were most comfortable (61/311, 20%), most understanding (69/311, 22%), and most influential toward behavior change (73/312, 23%) compared to traditional sources of health information–seeking, including mental health professionals, other health care professionals, school professionals, community leaders, friends, and family members. Indicators of family socioeconomic status were differentially associated with frequency and perceptions of and search content associated with parent HTU across technology sources.

Conclusions: The findings of this study underscore critical considerations in the design and dissemination of digital resources, programs, and interventions targeting parent and child health, especially for families in traditionally underserved communities.

(J Med Internet Res 2022;24(11):e37455) doi:10.2196/37455

XSL•FO
RenderX

KEYWORDS

parenting; child; health behavior; information seeking; health-related technology use; health information; digital health; mobile health; socioeconomic status; accessibility

Introduction

In the past decade, researchers have shown an increased interest in parental online health information seeking (OHIS), the process by which parents search for health information using the internet, including search engines, forums, and social networking [1,2]. OHIS has been linked to various aspects of individual and family functioning, including parenting behavior, perceived social support, and health status [3-6]. While parents search for information related to their own health (ie, self-seeking), they are even more likely to use the internet for health information related to their children (ie, proxy seeking). Indeed, data from the past several years revealed that 75% to 90% of parents have searched for health information related to their child [1].

Despite the widespread prevalence of parent health information–seeking on the internet, there is a paucity of research among parents of young children ages 3 to 8 years [1]. Research indicates that up to one-third (15% to 30%) of young children experience social, emotional, and behavioral problems [6-9]. Further, difficulties during this critical developmental period can persist into adolescence and adulthood, increasing the risk for long-term academic, occupational, and physical and mental health difficulties [10,11], especially for children in traditionally underserved communities with less access to quality care [12]. Given the importance of early development in child and family health, exploring how sociodemographic characteristics linked to the digital divide and health inequities affect parent proxy- and self-seeking is critical to further understanding variability in health information–seeking behaviors in the community [13,14].

Accordingly, this study addresses 2 underdeveloped research areas with parents of young children. First, the bulk of work has focused on clinical or treatment-seeking samples of parents with specific presenting issues (eg, attention deficit hyperactivity disorder, hearing loss) or circumstances (eg, after childbirth, during a visit to a pediatric outpatient clinic). However, parents' recognition of health-related concerns outside of the traditional health care system may depend on the extent to which they perceive a mismatch between their child's functioning and the socially and culturally relevant contexts (eg, school, home) in which they engage in daily life. Further, such perceptions may prompt parents to search for health-related content within broader domains (eg, child academic performance, parental discipline) of child and family functioning. Considering the information-seeking behaviors of parents of young children experiencing chronic illnesses or acute health problems may not generalize to other parents, studies with non–treatment-seeking samples are critical to understanding health information needs, seeking behaviors, and outcomes across diverse families.

Second, prior studies investigating parent OHIS have been limited to internet use, defined broadly and inconsistently across studies [1,15]. Considering the increasing adoption and use of other consumer technologies (eg, mobile apps and wearables) for health-related reasons and long-standing disparities in broadband access and connectivity, there is a need to extend current work to account for parent use of a variety of information and communications technologies [16-19]. Accordingly, we refer to parental health-related technology use (HTU) as the process by which parents engage in support, advice, and information-seeking behavior related to their (self-seeking) and their children's (proxy seeking) health across a broader range of devices (ie, tablets, wearables, smartphones, laptops, and desktop computers) and sources (ie, search engines, mobile applications, social media, and other digital media).

Building upon these gaps in the literature, this study aims to describe HTU among parents of young children, including the frequency and perceived usefulness of and search content associated with parent HTU in a non–treatment-seeking sample. In addition, resources (eg, parent access to technology devices) and perceptions (eg, comfortability) that may influence parent engagement in HTU are examined. Finally, whether patterns vary by parent, child, and household-level sociodemographic characteristics is explored.

Methods

Participant Recruitment

Parents and guardians of children ages 3 to 8 years old were recruited through Amazon Mechanical Turk (MTurk) to complete a survey on a broad range of parenting and parent-related constructs. Parents consented online before completing study measures in compliance with university-approved institutional review board (IRB) procedures. Upon confirming eligibility criteria, respondents were asked to select their youngest child in the specified age range to be referred to as the target child throughout the survey. All demographic variables and questionnaires were completed regarding the selected target child.

Additional measures were included to increase confidence in a participant pool that provides responses comparable to traditional samples (eg, [20-22]). To ensure attention to survey responses, 4 attention check questions were included throughout the survey (eg, "For data quality purposes, please select Sometimes") and were assessed as part of the inclusion criteria. Additionally, respondents with duplicate IP addresses, geolocations, and MTurk IDs were excluded from analyses in accordance with recommendations for studies using MTurk samples. As with other crowdsourcing platforms, MTurk duplicates typically reflect multiple entries from the same individual or household or, most prominently, "bot" (ie, computer programs that can automatically complete surveys) or "farmer" respondents (ie, individuals using server farms or commercial data centers to evade MTurk's screening procedures). Furthermore, these respondents are linked to lower-quality data [20]. Finally, a random numerical code was

XSL•FO
RenderX

provided to eligible participants (ie, parents of children ages 3 to 8 years old living in the United States) upon completion of the study to facilitate participant payment of US $2.

Ethics Approval

This study (17-0722) was approved by the institutional review board of the University of North Carolina at Chapel Hill.

Measures

Sociodemographic Characteristics

Parents reported sociodemographic information for their family, including the age, race (eg, White, African American/Black, Asian or Pacific Islander, American Indian/Alaska Native, or multiracial), and ethnicity (eg, Hispanic/Latino) of both the respondent (ie, parent or caregiver) and target child. Multiple indicators of family socioeconomic status were also collected, including annual household income, parent employment status (eg, full-time employment, part-time employment, unemployed but looking for work, nonworking, and retired), parent educational attainment (eg, less than high school or General Education Diploma [GED], high school graduate or GED, some college, associate's degree, bachelor's degree, master's degree, and doctorate), and perceived financial difficulty. Finally, parents also reported their household composition, marital status, relationship to the target child, and the target child's health status (ie, prior diagnosis of or treatment for developmental delays).

Technology Device Ownership, Access, and Use

Parents reported their access to and frequency of using common technology devices (ie, desktop computer, laptop computer, smartphone, tablet, wearable) measured on a 6-point Likert scale ranging from 0 (never) to 5 (more than once daily).

Search Content

Parents reported the content of their search for health-related information, advice, or support, focusing on 3 broad domains of proxy seeking (child academics, behavior, and mental and physical health) and 1 domain of self-seeking (parent stress and stress management).

Frequency and Usefulness of Parent Health-Related Technology Use

Parents indicated their use and perceptions of particular technology-enabled sources (eg, search engines, mobile apps, social media, and other forms of digital media) to search for parenting advice and health-related information for their children. Parents reported the frequency of using each source (ie, "When you are looking for parenting advice, information, and/or support, how often do you turn to each of the following potential sources?") using a 4-point Likert scale ranging from 0 (never) to 3 (frequently). Although the usefulness of particular sources has been evaluated inconsistently in the literature on parent HTU (eg, [23,24]), researchers often use a single-item measure to capture the construct (eg, "How useful do you feel the internet is in helping you make decisions about your health?") [25]. Similarly, parents reported the usefulness of a source (ie, "How helpful or useful did you find the parenting advice, information, and/or support you received from these sources?") using a 4-point Likert scale ranging from 0 (not at all helpful) to 3 (very helpful).

Statistical Analysis

Descriptive statistics were used to summarize family characteristics, parent ownership of or access to consumer technology devices, and parent engagement in and perceptions of HTU. Chi-square tests were conducted to compare proportions of and determine associations between device ownership and characteristics of HTU (eg, search content, frequency of use, and usefulness of technology sources) across groups defined by parent educational attainment (<bachelor's degree vs ≥bachelor's degree), perceived financial difficulty (none to mild vs moderate to severe), and low-income status, as determined by the federal poverty level (FPL), which accounts for annual household income and the number of people in the household (<200% FPL vs ≥200% FPL). Importantly, while "low-income" has been defined inconsistently in the literature, the FPL is typically used to determine eligibility for services, including those related to child health and development (eg, Head Start, Children's Health Insurance Program) [26]. While income eligibility varies by state and service, 200% FPL has been mandated as an upper limit for participation in several government services (eg, Children's Health Insurance Program, Subsidized Child Care Assistance Program), and incomes below 200% FPL account for a significant proportion of families in the United States who experience increased financial burden and economic insecurity [27]. Indeed, nearly 17% of children in the United States live in poverty, with approximately 7% (New Hampshire) to 56% (Puerto Rico) living in households below 200% FPL across the United States [28,29]. Of note, sociodemographic characteristics were included in analyses based on their theoretical relevance, as indicated in the previous research [13,18]. Missing values were excluded from analyses. Statistical analyses were conducted using SPSS version 26 software.

Results

Participants

Of the 657 respondents who completed the survey, 344 were removed from analyses for screening ineligibility (eg, families without a child in the specified age range or living outside of the United States, n=116), missed attention check questions (n=86), or duplicate IP addresses, geolocations, or MTurk IDs (n=142), yielding a total of 313 for analyses. Parents ranged in age from 19 to 57 years with a mean parental age of 34.19 (SD 7.11) years. Three-fifths (186/313, 59.4%) of parents self-identified as female. Slightly more than half (176/313, 56.2%) of parents obtained a bachelor's degree or higher (ie, master's or doctorate), and most were employed full- or part-time (280/313, 89.5%). Most parents were also married (243/313, 77.6%) and the biological parent of the target child (281/313, 89.8%). According to the parent report, approximately half (153/312, 49%) of the target children were female, and their mean age was 4.67 (SD 1.37) years. The racial and ethnic identity of most parents was White and non-Hispanic/Latino (230/312, 73.7%), followed by 8.7% (27/312) African American or Black, 4.5% (14/312) Asian American, 0.6% (2/312)

XSL•FO
RenderX

American Indian/Alaska Native, and 3.2% (10/312) multiracial. For 10.5% (33/313) of the children, the parent's self-reported race or ethnicity differed from that of the child. Nevertheless, the majority of children identified as White and non-Hispanic/Latino (207/312, 66.3%), followed by 7.4% (23/312) African American, 3.8% (12/312) Asian American, 0.6% (2/312) American Indian/Alaska Native, and 13.1% (41/312) multiracial. Additionally, 11.8% (37/313) of parents and 15.3% (48/313) of children identified as Hispanic or Latino. The annual combined household income ranged from US $6000 to $380,000 with a median of $60,000 (SD $41,180). Finally, 56.2% (176/313) of families reported living in suburban areas, followed by 25.2% (79/313) in urban areas and 18.5% (58/313) in rural areas. Compared to the general population of parents in the United States, the recruited sample included slightly more college-educated and lower-income participants and a comparable percentage of women and married parents [30,31]. Additionally, parents were less racially and ethnically diverse than the general population of parents in the United States but slightly more so than has been reported in previous studies with parents using MTurk samples.

Device Ownership and Use

Parents reported owning a variety of technology devices, including a smartphone (276/313, 88.2%), laptop (276/313, 88.2%), tablet (243/313, 77.6%), desktop computer (193/313, 61.6%), and wearable device (100/313, 31.9%; Table 1). All (100%, 313/313) parents reported owning or having access to at least 1 technology device at home, and the majority (283/313, 90.4%) of parents reported access to multiple devices. Only 2 (0.64%) parents indicated not having access to a computer (desktop or laptop) at home, and both reported having access at work, school, or another setting (eg, library). Of the 37 (11.8%) participants that reported not having access to a smartphone, 12 (32.4%) reported having access to a smartphone at work, school, or another setting. Chi-square analyses revealed no statistically significant associations between parent educational attainment or perceived financial difficulty and access to technology devices. Families in low-income households were significantly less likely to own or have access to a wearable (χ^2_1=4.7, *P*=.03), but not any other technology device.

Table 1. Technology device ownership and access.

Demographics	Overall (N=313), n (%)	Desktop (n=193)		Laptop (n=276)		Smartphone (n=276)		Tablet (n=243)		Wearable (n=100)	
		Value, n (%)	χ^2(*df*), *P* value	Value, n (%)	χ^2(*df*), *P* value	Value, n (%)	χ^2(*df*), *P* value	Value, n (%)	χ^2(*df*), *P* value	Value, n (%)	χ^2(*df*), *P* value
Parent educational attainment			0.120 (*1*), .729		0.005 (*1*), .945		0.178 (*1*), .673		0.010 (*1*), .921		0.849 (*1*), .357
<Bachelor's degree	137 (43.8)	83 (60.6)		121 (88.3)		122 (89.1)		106 (77.4)		40 (29.2)	
≥Bachelor's degree	176 (56.2)	110 (62.5)		155 (88.1)		154 (87.5)		137 (77.8)		60 (34.1)	
Household income			0.246 (*1*), .620		2.676 (*1*), .102		0.147 (*1*), .701		0.075 (*1*), .785		4.744 (*1*), .034[a]
<200% FPL[b]	95 (30.4)	57 (60.0)		88 (92.6)		85 (89.5)		73 (76.8)		22 (23.2)	
≥200% FPL	216 (69.0)	136 (63.0)		186 (86.1)		190 (88.0)		169 (78.2)		77 (35.7)	
Perceived financial difficulty			0.028 (*1*), .868		0 (*1*), .998		0.591 (*1*), .442		0.127 (*1*), .721		1.563 (*1*), .211
None to mild	220 (70.3)	135 (61.4)		194 (88.2)		196 (89.1)		172 (78.2)		75 (34.1)	
Moderate to severe	93 (29.7)	58 (62.4)		82 (88.2)		80 (86.0)		71 (76.3)		25 (26.9)	

[a]*P*<.05.

[b]FPL: federal poverty level.

Among parents who reported access to a computer at home or another setting, 100% reported using their laptop or desktop device at least once monthly, with most reporting using their desktop computer (104/155, 67.1%) or laptop (117/177, 66.1%) more than once daily. Over half (31/55, 56.4%) of the parents reported using a wearable device multiple times during the day, and only 7% (4/55) reported using their wearable less than every 3 days. Approximately 31.1% (28/90) of parents who reported access to a tablet at home or another setting reported using their device more than once daily, and nearly a quarter (22/90, 24.4%) reported using their tablet once weekly or less. Most (177/196, 90.3%) parents reported using their smartphone multiple times per day, and no parents reported using their smartphone less than every 3 days.

XSL•FO
RenderX

The frequency of smartphone use was significantly lower for families in low-income households (χ^2_2=9.8, *P*=.007) and with parents reporting moderate to severe financial difficulty (χ^2_2=7.8, *P*=.021). Additionally, parents experiencing moderate to severe financial difficulty used their desktop computer (X^2_5=11.5, *P*=.042), laptop (X^2_5=12.4, *P*=.015), and tablet (X^2_5=23.9, *P*<.001) less frequently than their peers. The frequency of using any technology device did not vary significantly by parent educational attainment. Notably, parent age was not significantly correlated with the frequency of using any technology device.

Parent HTU

Most parents (301/313, 96.2%) reported using technology sources to search for parenting advice and health-related information for their children. Parents who engaged in health-related technology use reported using search engines (eg, Google; 300/301, 99.7%), social media (188/301, 62.5%), other forms of digital media (eg, podcasts; 145/301, 48.2%), and mobile applications (114/301, 37.9%). Approximately one-third (91/301, 30.2%) of parents reported using all 4 sources for proxy- and self-seeking.

There were no significant differences between parents who did and did not report engaging in HTU via mobile apps, social media, and other digital media across parent educational attainment (Table 2). Parents in low-income households were significantly more likely to report using mobile apps (χ^2_1=4.7, *P*=.030) and social media (X^2_1=4.9, *P*=.026) for health-related reasons, but not other forms of digital media. Parents reporting moderate to severe financial difficulty were significantly more likely to report using mobile apps (X^2_1=5.5, *P*=.019), social media (X^2_1=4.2, *P*=.040), and other digital media (X^2_1=7.3, *P*=.007) in comparison to their peers. Given that most participants reported using search engines, the technology source was not included in the chi-square analyses.

While the frequency of engagement in HTU varied across sources, parents reported more frequent use of search engines on average, followed by social media, mobile apps, and other digital media (Figure 1). The frequency of parent use was not significantly associated with self-reported parent educational attainment. The mean frequency of social media use (χ^2_3=16.4, *P*<.001) was significantly greater for parents in low-income households, and the use of social media (χ^2_3=11.9, *P*=.008) and other digital media (χ^2_3=10.4, *P*=.016) was also increased for parents reporting moderate to severe financial difficulty.

Among parents who reported using each source, an overwhelming majority (280/300, 93.3%) indicated that search engines were a useful online source for proxy- and self-seeking, followed by social media (167/188, 88.8%), other digital media (120/145, 82.8%), and mobile apps (87/114, 76.3%). Parents in low-income households also rated other digital media as more useful than their peers (χ^2_3=9.19, *P*=.027). Perceived financial difficulty and parent educational attainment were not significantly associated with the perceived usefulness of any technology source.

Table 2. Parent engagement in health-related technology use (HTU).

Demographics	Overall (N=313), n (%)	Social media (n=188)		Other media (n=145)		Mobile apps (n=114)	
		Value, n (%)	χ^2(*df*), *P* value	Value, n (%)	χ^2(*df*), *P* value	Value, n (%)	χ^2(*df*), *P* value
Parent educational attainment			0.027 (*1*), .868		1.560 (*1*), .212		0.001 (*1*), .981
<Bachelor's degree	137 (43.8)	83 (60.6)		58 (42.3)		50 (36.5)	
≥Bachelor's degree	176 (56.2)	105 (59.7)		87 (49.4)		64 (36.4)	
Household income			4.983(*1*), .026[a]		0.982 (*1*), .322		4.714 (*1*), .030[a]
<200% FPL[b]	95 (30.4)	66 (69.5)		48 (50.5)		43 (45.3)	
≥200% FPL	216 (69.0)	121 (56.0)		96 (44.4)		70 (32.4)	
Perceived financial difficulty			4.226 (*1*), .040[a]		7.332 (*1*), .007[c]		5.504 (*1*), 0.019[a]
None to mild	220 (70.3)	124 (56.4)		91 (41.4)		71 (32.3)	
Moderate to severe	93 (29.7)	64 (68.8)		54 (58.1)		43 (46.2)	

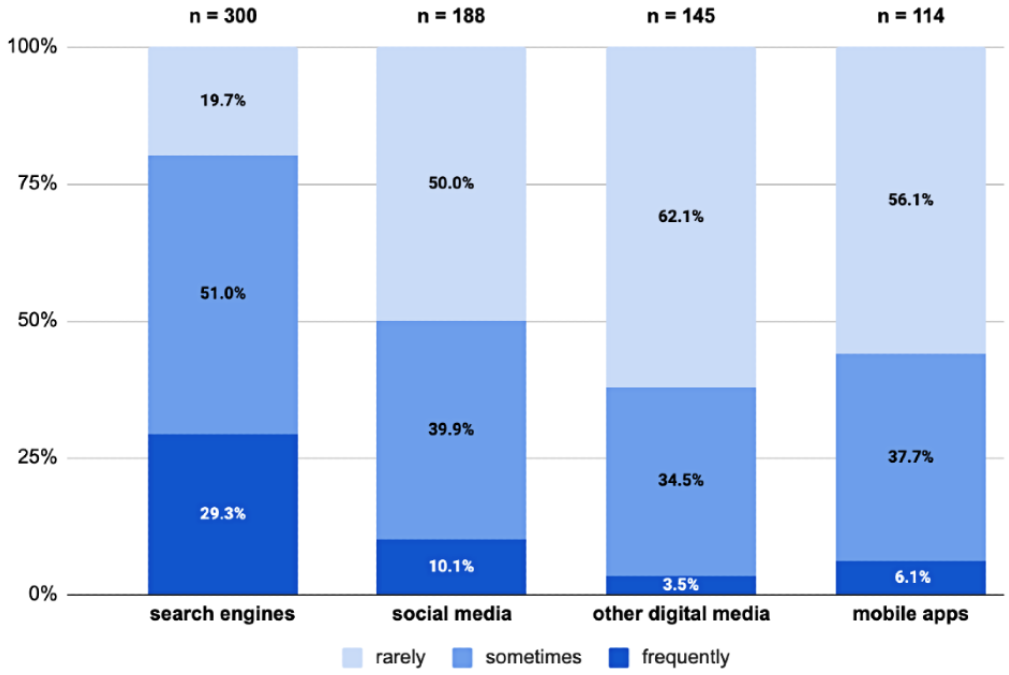
[a]*P*<.05.

[b]FPL: federal poverty level.

[c]*P*<.01.

XSL•FO
RenderX

Figure 1. Frequency of parent health-related technology use (HTU) across technology sources.



Search Content

Parents who engaged in HTU reported seeking information about their child's behavior and discipline practices (260/313, 83.1%), mental and physical health (181/313, 57.8%), and academic performance (142/313, 45.4%). Additionally, 42.8% (134/313) of parents reported searching for advice on managing their stress. Parents in low-income households were significantly less likely to search for health-related information or advice about their child's physical and mental health (χ^2_1=5.0, P=.025) and more likely to search for content about parent stress and stress management (χ^2_1=12.2, P<.001; Table 3). Parents reporting moderate to severe financial difficulty were also more likely to search for the latter (χ^2_1=4.2, P=.041). Parent educational attainment was not significantly associated with any search content.

Table 3. Parent proxy and self-seeking content areas.

Demographics	Overall (N=313), n (%)	Child behavior/discipline (n=260)		Child academic performance (n=142)		Child physical/mental health (n=181)		Parent stress/stress management (n=134)	
		Value, n (%)	$\chi^2(df)$, P value	Value, n (%)	$\chi^2(df)$, P value	Value, n (%)	$\chi^2(df)$, P value	Value, n (%)	$\chi^2(df)$, P value
Parent educational attainment			0.059 (*1*), .808		0.424 (*1*), .515		1.452(*1*), .228		3.695 (*1*), .055
<Bachelor's degree	137 (43.77)	113 (82.48)		65 (47.45)		74 (54.01)		67 (48.91)	
≥Bachelor's degree	176 (56.23)	147 (83.52)		77 (43.75)		107 (60.80)		67 (38.07)	
Household income			0.085 (*1*), .770		1.013 (*1*), .314		5.018 (*1*), .025[a]		12.231 (*1*), <.001[b]
<200% FPL[c]	95 (30.35)	80 (84.21)		39 (41.05)		46 (48.42)		55 (57.89)	
≥200% FPL	216 (69.01)	179 (82.87)		102 (47.22)		134 (62.04)		79 (36.57)	
Perceived financial difficulty			0.061 (*1*), .805		0.629 (*1*), .428		0.309 (*1*), .578		4.186 (*1*), .041[a]
None to mild	220 (70.29)	182 (82.73)		103 (46.82)		125 (56.82)		86 (39.09)	
Moderate to severe	93 (29.71)	78 (83.87)		39 (41.94)		56 (60.22)		48 (51.61)	

[a]P<.05

[b]P<.01

[c]FPL: federal poverty level.

XSL•FO
RenderX

Perceptions

Among parents who reported using any technology source, approximately one-fifth reported that technology sources were the most comfortable (61/311, 19.6%), most understanding (69/311, 22.2%), and most influential toward behavior change (73/312, 23.4%) compared to traditional sources, including mental health professionals, other health care professionals, school professionals, community leaders, friends, and family members. For perceived understanding, the majority of parents (48/69, 69.6%) referenced search engines, followed by social media (19/69, 27.5%) and other digital media and mobile apps (both less than 1/69, 2%). Similarly, for perceived comfortability, most parents listed search engines (42/61, 68.9%) and social media (18/61, 29.5%), and fewer mentioned mobile apps (1/61, 1.6%) and other digital media (0/61, 0%). Finally, in terms of parenting behavior change, search engines accounted for 73.97% (54/73), followed by social media (16/73, 21.91%), other digital media (2/73, 2.74%), and mobile apps (1/73, 1.36%). Perceived financial difficulty, but not any other socioeconomic status (SES) indicator, was significantly associated with perceptions of technology sources for health information seeking, such that parents experiencing moderate to severe difficulty were more likely to perceive engagement in HTU as the most understanding (χ^2_1=14.2, *P*<.001), most comfortable (χ^2_1=7.9, *P*=.005), and most likely to lead to behavior change (χ^2_1=7.3, *P*=.007) compared to traditional sources (Table 4).

Table 4. Parent perceptions of health-related technology use (HTU).

Demographics	Overall (N=313), n (%)	Most understanding (n=69)		Most comfortable (n=61)		Most parenting behavior change (n=73)	
		Value, n (%)	χ^2(*df*), *P* value	Value, n (%)	χ^2(*df*), *P* value	Value, n (%)	χ^2(*df*), *P* value
Parent educational attainment			2.773 (*1*), .096		0.528 (*1*), .468		0.735 (*1*), .391
<Bachelor's degree	137 (43.8)	36 (26.7)		29 (21.5)		35 (25.7)	
≥Bachelor's degree	176 (56.2)	33 (18.8)		32 (18.2)		38 (21.6)	
Household income			0.978(*1*), .323		1.144 (*1*), .285		0.074 (*1*), .785
<200% FPL[a]	95 (30.4)	24 (25.5)		22 (23.4)		23 (24.2)	
≥200% FPL	216 (69.0)	44 (20.5)		39 (18.1)		49 (22.8)	
Perceived financial difficulty			14.169 (*1*), <.001[b]		7.851 (*1*), .005[b]		7.298 (*1*), .007[b]
None to mild	220 (70.3)	36 (16.4)		34 (15.5)		42 (19.2)	
Moderate to severe	93 (29.7)	33 (35.9)		27 (29.3)		31 (33.3)	

[a]FPL

[b]*P*<.01

Discussion

Principal Findings

Given the increased prevalence of parent health-related technology use in recent years, this study aimed to explore family socioeconomic factors associated with this parenting behavior in a diverse sample of parents of young children. Considering that several developmental, socioemotional, and behavioral problems emerge in early childhood, understanding parent HTU use during this period has numerous clinical and public health implications. Indeed, children from lower SES households are more likely to experience reduced health quality and are less likely to have access to traditional health care services than children from higher SES households, and the relationship between these disparities and long-standing structural barriers is well established [32]. Further, research suggests similar barriers persist in access to technology devices and broadband, which may also challenge recent efforts to leverage technology to address health disparities [17,33,34]. Thus, understanding patterns and perceptions of parental HTU is critical for efforts to democratize digital health for parents of young children.

In the past decade, there has been a significant increase in technology device ownership in the United States, most substantially among smartphones and tablets [19]. Recruited families displayed a slightly higher percentage of smartphone, computer, tablet, and wearable device ownership and access in comparison to recent surveys of US adults [16,19], which may be reflective of our focus on parents (rather than adults in general), recruitment methods (eg, telephone interviews vs Amazon Mechanical Turk), or the inclusion of families in analyses with access to devices in other settings (eg, work, school, or library). Over three-fifths of parents endorsed ownership or access to a smartphone, tablet, and desktop or laptop computers, which did not vary across educational attainment, perceived financial difficulty, or household income. However, fewer than a third of parents reported access to a wearable device, and families with a lower income were significantly less likely to own a wearable (23% vs 36%).

Importantly, the majority of parents of young children reported using their laptop (150/177, 85%) or desktop computer (130/155, 84%) and wearable devices (45/55, 82%) daily, and the overwhelming majority of parents reported using their smartphone more than once per day. In contrast, only half of

XSL•FO
RenderX

the parents reported using their tablet daily. Some, but not all, indicators of SES were significantly associated with how often parents used their smartphone (income and perceived financial difficulty), tablet (perceived financial difficulty), and desktop computer (parent educational attainment), with parents without a bachelor's degree, those experiencing moderate to severe financial difficulty, and those in lower-income households using their technology devices less frequently than their peers.

Regarding engagement in HTU among parents of young children, our findings were congruent with the high rates observed in previous studies of parent health information seeking via the internet [1]. However, these results extended the existing research by examining differential engagement across technology sources (eg, search engines, mobile apps, social media, and other digital media) in general and across sociodemographic groups. Consistent with previous research, nearly all parents in our study endorsed the use of search engines. In general, fewer parents reported using social media for health-related reasons in comparison to estimates of general social media use by parents (62% v 75%) [35]; however, existing work has primarily examined parents of infants, toddlers, or children under 18 years of age broadly [15]. Findings also indicate that less than half the parents of young children currently use mobile apps (38%) and other digital media (48%) to search for health-related information, advice, or support.

Additionally, there have been inconsistent findings regarding the relationship between family SES and parent HTU. For objective dimensions of SES, this is partly attributable to the underreporting of household income, household composition, and parent educational attainment in studies (40% did not report the education level of participants in a recent meta-analysis), as well as the recruitment of predominantly highly educated (over 50% to 75% with an academic degree) and higher-income parents among remaining studies [1]. Furthermore, to our knowledge, no studies to date have included subjective dimensions of SES in analyses (eg, perceived financial hardship, subjective social status), despite their distinct effects on parenting behavior and family health [36-41]. In contrast to studies observing higher rates of health-seeking behavior via the internet with increased parent educational attainment [1,42,43], our findings suggest no significant associations between parent educational attainment and engagement in or frequency of health-related technology use across sources. However, parents in lower-income households and those experiencing greater financial difficulty were significantly more likely to use social media (69% vs 56% for both) and mobile apps (45% vs 32% and 46% vs 32%, respectively) for health-related reasons. Parents who reported greater financial difficulty were also more likely to use other forms of digital media (58% vs 41%). Moreover, parents experiencing moderate to severe financial difficulty used social media less frequently than their peers. In terms of search content, both lower income and increased perceived financial difficulty (52% vs 39%) were associated with increased self-seeking behavior related to parent stress and stress management, and lower income was additionally associated with a decreased likelihood of parent engagement in proxy seeking related to their child's mental and physical health. Finally, parent perceptions of health-related technology use broadly did not vary by any objective dimensions of SES; however, parents experiencing moderate to severe financial difficulty were significantly more likely to perceive technology sources as the most comfortable (29% vs 15%), understanding (36% vs 16%), and likely to influence behavior change (33% vs 19%) compared to traditional sources. These findings support early research suggesting that SES indicators have differential impacts on health behavior and outcomes, providing a basis for further exploration of the underlying mechanisms contributing to outcomes in parent HTU.

Taken together, the results of this study underscore potential considerations for clinicians, researchers, and public health practitioners engaged in the design and dissemination of digital resources, programs, and interventions targeting family health and well-being. For instance, our findings suggest that digital health tools developed with greater attention to the types of technology sources parents prefer for health-related information, their frequency of engagement with these sources (eg, daily or weekly), and the availability of technology devices required to access these sources may yield increased uptake. Further, our results suggest practical considerations for efforts striving to optimize effectiveness (eg, which commercial devices and sources have the necessary features and functionality?), scalability (eg, what are the current estimates of, trends in, and barriers to adoption of these devices, especially in historically excluded communities?), and sustainability (eg, how acceptable and usable are both the devices and sources for the target population?). For example, digital resources, programs, and interventions requiring devices compatible solely with mobile operating systems (eg, mobile apps for Android, Apple iOS, and iPadOS) may call for a consideration of parent access to, familiarity with, and perceptions of smartphones and tablets, as well as their perceptions of mobile apps as a source for health-related information and support. Importantly, the success of these efforts hinges on broader attention to policies that address the structural information, infrastructure, and implementation barriers to diverse parents' safe and effective engagement in HTU, such as access to technology devices and reliable internet (eg, [44]) and threats to online safety (eg, health misinformation and disinformation [45-47]).

Limitations and Future Directions

Despite its strengths, this study has some limitations. First, primarily descriptive analyses were conducted to explore associations between sociodemographic factors and outcome variables. Second, inadequate representation of all racial and ethnic groups precluded our ability to examine how the diversity of social context and experiences across and within groups influence health-related technology use, which is a critical step in future research given the well-established disparities in digital adoption, health outcomes, and access to care among racially and ethnically minoritized children and their families [48-51]. Third, sources were grouped into technology (ie, search engines, mobile apps, social media, and other digital media) and traditional (ie, family, friends, mental health care providers, other health care providers, school professionals, and community leaders) categories for the analyses of perceptions of HTU, despite their potential interconnections in daily life (eg, use of social media to connect with family members about child-related

XSL•FO
RenderX

health concerns, use of telemedicine apps for remote health care services). Future research should explore these complex relationships, which are likely linked to other relevant individual (eg, parent and child psychosocial factors and attitudes) and environmental (eg, social support, discrimination) factors associated with engagement in HTU and outcomes (eg, specific parent behaviors, family health outcomes, subsequent HTU). Finally, survey data were collected in late 2018 (prior to the COVID-19 pandemic), spotlighting the importance of future work examining potentially evolving trends in technology adoption and parent HTU.

Conclusion

In summary, this study investigated engagement in support, advice, and information-seeking behavior among parents of young children across technology devices and sources. It also examined resource access and perceptions that may influence engagement and explored patterns across family SES. Overall, this study supports the growing body of evidence demonstrating the potential for digital technologies to disseminate health-related information, support, and resources to young children and families facing structural socioeconomic barriers. Furthermore, it may inform future research necessary to advance understanding on how to more optimally tailor and deliver supports that benefit the health and well-being of all children.

Acknowledgments

This research was supported by grants from the National Institutes of Health (NIH) R21MH113887 (NCT03597789) and the Robert Wood Johnson Foundation Health Policy Research Scholarship.

Conflicts of Interest

None declared.

References

1. Kubb C, Foran HM. Online health information seeking by parents for their children: systematic review and agenda for further research. J Med Internet Res 2020 Aug 25;22(8):e19985 [FREE Full text] [doi: 10.2196/19985] [Medline: 32840484]
2. Lee HS. Research trends in mothers' health information seeking behaviors: a review of the literature. Proc Assoc Info Sci Tech 2016 Dec 27;53(1):1-6. [doi: 10.1002/pra2.2016.14505301130]
3. Jiang S, Street RL. Pathway linking internet health information seeking to better health: a moderated mediation study. Health Commun 2017 Dec;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
4. Doty JL, Dworkin J. Online social support for parents: a critical review. Marriage Fam Rev 2014 Mar 03;50(2):174-198. [doi: 10.1080/01494929.2013.834027]
5. Wang X, Shi J, Kong H. Online health information seeking: a review and meta-analysis. Health Commun 2020 Apr 14:1-13. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
6. Walsh AM, Hamilton K, White KM, Hyde MK. Use of online health information to manage children's health care: a prospective study investigating parental decisions. BMC Health Serv Res 2015 Apr;15:131 [FREE Full text] [doi: 10.1186/s12913-015-0793-4] [Medline: 25889493]
7. Cree RA, Bitsko RH, Robinson LR, Holbrook JR, Danielson ML, Smith C, et al. Health care, family, and community factors associated with mental, behavioral, and developmental disorders and poverty among children aged 2-8 years - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Dec 21;67(50):1377-1383 [FREE Full text] [doi: 10.15585/mmwr.mm6750a1] [Medline: 30571671]
8. Powell D, Dunlap G, Fox L. Prevention and intervention for the challenging behaviors of toddlers and preschoolers. Infants Young Child 2006;19(1):25-35. [doi: 10.1097/00001163-200601000-00004]
9. Huaqing Qi C, Kaiser AP. Behavior problems of preschool children from low-income families. Topics Early Child Spec Educ 2016 Aug 17;23(4):188-216. [doi: 10.1177/02711214030230040201]
10. Brennan LM, Shaw DS, Dishion TJ, Wilson MN. The predictive utility of early childhood disruptive behaviors for school-age social functioning. J Abnorm Child Psychol 2015 Aug 20;43(6):1187-1199 [FREE Full text] [doi: 10.1007/s10802-014-9967-5] [Medline: 25526865]
11. Carter AS, Briggs-Gowan MJ, Davis NO. Assessment of young children's social-emotional development and psychopathology: recent advances and recommendations for practice. J Child Psychol Psychiatry 2004 Jan;45(1):109-134. [doi: 10.1046/j.0021-9630.2003.00316.x] [Medline: 14959805]
12. Shi L, Stevens GD. Vulnerability and unmet health care needs. The influence of multiple risk factors. J Gen Intern Med 2005 Feb;20(2):148-154 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40136.x] [Medline: 15836548]
13. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a health information-seeking population and the current digital divide: cross-sectional analysis of the 2015-2016 California Health Interview Surveyb. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
14. Powe BD. Health information seeking among rural African Americans, Caucasians, and Hispanics: it is built, did they come? Nurs Clin North Am 2015 Sep;50(3):531-543. [doi: 10.1016/j.cnur.2015.05.007] [Medline: 26333608]

XSL•FO
RenderX

15. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
16. Vogels E. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2022-01-10]
17. Vogels E. Digital divide persists even as Americans with lower incomes make gains in tech adoption. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/ [accessed 2022-01-10]
18. 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 2022-01-10]
19. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-01-10]
20. Chmielewski M, Kucker SC. An MTurk crisis? Shifts in data quality and the impact on study results. Soc Psychol Personal Sci 2019 Oct 10;11(4):464-473. [doi: 10.1177/1948550619875149]
21. Hauser DJ, Schwarz N. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behav Res Methods 2016 Mar 12;48(1):400-407. [doi: 10.3758/s13428-015-0578-z] [Medline: 25761395]
22. Kennedy R, Clifford S, Burleigh T, Waggoner PD, Jewell R, Winter NJG. The shape of and solutions to the MTurk quality crisis. PSRM 2020 Apr 24;8(4):614-629. [doi: 10.1017/psrm.2020.6]
23. Naftel RP, Safiano NA, Falola MI, Shannon CN, Wellons JC, Johnston JM. Technology preferences among caregivers of children with hydrocephalus. J Neurosurg Pediatr 2013 Jan;11(1):26-36. [doi: 10.3171/2012.9.PEDS12208] [Medline: 23092227]
24. AlSaadi MM. Evaluation of internet use for health information by parents of asthmatic children attending pediatric clinics in Riyadh, Saudi Arabia. Ann Saudi Med 2012 Nov;32(6):630-636 [FREE Full text] [doi: 10.5144/0256-4947.2012.630] [Medline: 23396028]
25. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
26. Jones DJ, Anton M, Zachary C, Pittman S, Turner P, Forehand R, et al. A review of the key considerations in mental health services research: a focus on low-income children and families. Couple Family Psychol 2016 Dec;5(4):240-257 [FREE Full text] [doi: 10.1037/cfp0000069] [Medline: 28503361]
27. Koball H, Jiang Y. Basic facts about low-income children: children under 9 years, 2016. Columbia University Libraries Academic Commons. 2018. URL: https://academiccommons.columbia.edu/doi/10.7916/d8-t97p-bf73 [accessed 2022-01-10]
28. Poverty rate of children higher than national rate, lower for older populations. United States Census Bureau. URL: https://www.census.gov/library/stories/2022/10/poverty-rate-varies-by-age-groups.html [accessed 2022-10-07]
29. State health facts: poverty rate by age. Kaiser Family Foundation. 2020 Oct 23. URL: https://www.kff.org/other/state-indicator/poverty-rate-by-age/ [accessed 2022-01-10]
30. America's families and living arrangements: 2019. United States Census Bureau. URL: https://www.census.gov/data/tables/2019/demo/families/cps-2019.html [accessed 2022-01-10]
31. Jensen-Doss A, Patel ZS, Casline E, Mora Ringle VA, Timpano KR. Using Mechanical Turk to study parents and children: an examination of data quality and representativeness. J Clin Child Adolesc Psychol 2021 Jan 15:1-15. [doi: 10.1080/15374416.2020.1815205] [Medline: 33448875]
32. Ghandour RM, Sherman LJ, Vladutiu CJ, Ali MM, Lynch SE, Bitsko RH, et al. Prevalence and treatment of depression, anxiety, and conduct problems in US children. J Pediatr 2019 Mar;206:256-267.e3 [FREE Full text] [doi: 10.1016/j.jpeds.2018.09.021] [Medline: 30322701]
33. 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]
34. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
35. Duggan M, Lenhart A, Lampe C, Ellison N. Parents and social media. Pew Research Center. 2015 Jul 16. URL: https://www.pewresearch.org/internet/2015/07/16/parents-and-social-media/ [accessed 2022-01-10]
36. Diemer MA, Mistry RS, Wadsworth ME, López I, Reimers F. Best practices in conceptualizing and measuring social class in psychological research. Anal Soc Issues Public Policy 2012 Dec 26;13(1):77-113. [doi: 10.1111/asap.12001]
37. Odgers CL. Income inequality and the developing child: Is it all relative? Am Psychol 2015 Nov;70(8):722-731 [FREE Full text] [doi: 10.1037/a0039836] [Medline: 26618957]
38. Operario D, Adler NE, Williams DR. Subjective social status: reliability and predictive utility for global health. Psychol Health 2007 Feb;19(2):237-246. [doi: 10.1080/08870440310001638098]
39. Michelson N, Riis JL, Johnson SB. Subjective social status and psychological distress in mothers of young children. Matern Child Health J 2016 Oct 20;20(10):2019-2029 [FREE Full text] [doi: 10.1007/s10995-016-2027-8] [Medline: 27323755]

XSL•FO
RenderX

40. Jarrin DC, McGrath JJ, Quon EC. Objective and subjective socioeconomic gradients exist for sleep in children and adolescents. Health Psychol 2014 Mar;33(3):301-305 [FREE Full text] [doi: 10.1037/a0032924] [Medline: 23730721]
41. Kim KW, Wallander JL, Peskin M, Cuccaro P, Elliott MN, Schuster MA. Associations between parental SES and children's health-related quality of life: the role of objective and subjective social status. J Pediatr Psychol 2018 Jun 01;43(5):534-542. [doi: 10.1093/jpepsy/jsx139] [Medline: 29155956]
42. Kummervold PE, Chronaki CE, Lausen B, Prokosch H, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008 Nov;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
43. Wangberg SC, Andreassen HK, Prokosch H, Santana SMV, Sørensen T, Chronaki CE. Relations between internet use, socio-economic status (SES), social support and subjective health. Health Promot Int 2008 Mar;23(1):70-77 [FREE Full text] [doi: 10.1093/heapro/dam039] [Medline: 18083686]
44. 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]
45. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
46. Coughler C, M Burke S, Cardy JO. Analysis of the quality of online resources for parents of children who are late to talk. Autism Dev Lang Impair 2020 Apr 06;5:239694152091794. [doi: 10.1177/2396941520917940]
47. Howard P, Neudert L, Prakash N, Vosloo S. Digital misinformation/disinformation and children. UNICEF. 2021. URL: https://www.unicef.org/globalinsight/media/2096/file/UNICEF-Global-Insight-Digital-Mis-Disinformation-and-Children-2021.pdf [accessed 2022-01-10]
48. Alegria M, Vallas M, Pumariega AJ. Racial and ethnic disparities in pediatric mental health. Child Adolesc Psychiatr Clin N Am 2010 Oct;19(4):759-774 [FREE Full text] [doi: 10.1016/j.chc.2010.07.001] [Medline: 21056345]
49. Bauer GR. Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity. Soc Sci Med 2014 Jun;110:10-17 [FREE Full text] [doi: 10.1016/j.socscimed.2014.03.022] [Medline: 24704889]
50. Flores G, Olson L, Tomany-Korman SC. Racial and ethnic disparities in early childhood health and health care. Pediatrics 2005 Feb;115(2):e183-e193. [doi: 10.1542/peds.2004-1474] [Medline: 15687426]
51. Perrin A, Atske S. 7% of Americans don't use the internet. Who are they? Pew Research Center. 2021 Apr 02. URL: https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/ [accessed 2022-01-10]

Abbreviations

FPL: federal poverty level
GED: General Education Diploma
HTU: health-related technology use
IRB: institutional review board
MTurk: Amazon Mechanical Turk
OHIS: online health information seeking
SES: socioeconomic status

Edited by G Eysenbach; submitted 21.02.22; peer-reviewed by C Urquhart, A Videira-Silva; comments to author 30.08.22; revised version received 18.10.22; accepted 20.10.22; published 30.11.22.

Please cite as:
McCall MP, Hineline MT, Anton MT, Highlander A, Jones DJ
The Socioeconomic Indicators Linked to Parent Health-Related Technology Use: Cross-sectional Survey
J Med Internet Res 2022;24(11):e37455
URL: https://www.jmir.org/2022/11/e37455
doi:10.2196/37455
PMID:36449346

©Madison P McCall, Megan T Hineline, Margaret T Anton, April Highlander, Deborah J Jones. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceptions of Quality of Care Among Users of a Web-Based Patient Portal: Cross-sectional Survey Analysis

Rachael Lear[1], RN, BA, PhD; Lisa Freise[1], BSc, MSc; Matthew Kybert[2], BEng(Hons); Ara Darzi[1], PC, KBE; Ana Luisa Neves[1], MD, MSc, PhD; Erik K Mayer[1], BSc, MBBS, MRCS, PhD

[1]National Institute for Health and Care Research Imperial Patient Safety Translational Research Centre, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[2]Imperial College Healthcare NHS Trust, London, United Kingdom

Corresponding Author:
Rachael Lear, RN, BA, PhD
National Institute for Health and Care Research Imperial Patient Safety Translational Research Centre
Institute of Global Health Innovation
Imperial College London
10F Queen Elizabeth Queen Mother Wing Praed Street
St Mary's Campus
London, W2 1NY
United Kingdom
Phone: 44 (0)20 7594 1419
Email: r.lear12@imperial.ac.uk

Abstract

Background: Web-based patient portals enable patients access to, and interaction with, their personal electronic health records. However, little is known about the impact of patient portals on quality of care. Users of patient portals can contribute important insights toward addressing this knowledge gap.

Objective: We aimed to describe perceived changes in the quality of care among users of a web-based patient portal and to identify the characteristics of patients who perceive the greatest benefit of portal use.

Methods: A cross-sectional web-based survey study was conducted to understand patients' experiences with the Care Information Exchange (CIE) portal. Patient sociodemographic data were collected, including age, sex, ethnicity, educational level, health status, geographic location, motivation to self-manage, and digital health literacy (measured by the eHealth Literacy Scale). Patients with experience using CIE, who specified both age and sex, were included in these analyses. Relevant survey items (closed-ended questions) were mapped to the Institute of Medicine's 6 domains of quality of care. Users' responses were examined to understand their perceptions of how portal use has changed the overall quality of their care, different aspects of care related to the 6 domains of care quality, and patient's satisfaction with care. Multinomial logistic regression analyses were performed to identify patient characteristics associated with perceived improvements in overall care quality and greater satisfaction with care.

Results: Of 445 CIE users, 38.7% (n=172) reported that the overall quality of their care was better; 3.2% (n=14) said their care was worse. In the patient centeredness domain, 61.2% (273/445) of patients felt more in control of their health care, and 53.9% (240/445) felt able to play a greater role in decision-making. Regarding timeliness, 40.2% (179/445) of patients reported they could access appointments, diagnoses, and treatment more quickly. Approximately 30% of CIE users reported better care related to the domains of effectiveness (123/445, 27.6%), safety (138/445, 31%), and efficiency (174/445, 28.6%). Regarding equity, patients self-reporting higher digital health literacy (odds ratio 2.40, 95% CI 1.07-5.42; *P*=.03) and those belonging to ethnic minority groups (odds ratio 2.27, 95% CI 1.26-3.73; *P*<.005) were more likely to perceive improvements in care quality. Across ethnic groups, Asian and British Asian patients perceived the greatest benefits. Increased frequency of CIE use also predicted perceived better care quality and greater satisfaction with care.

Conclusions: A large proportion of CIE users perceived better care quality and greater satisfaction with care, although many portal users reported no change. The most favorable perceived improvements related to the domain of patient centeredness. With national policy directed toward addressing health disparities, patient portals could be valuable in improving care quality for ethnic minority groups. Future research should test the causal relationship between patient portal use and care quality.

(J Med Internet Res 2022;24(11):e39973) doi:10.2196/39973

XSL•FO
RenderX

KEYWORDS

electronic health records; personal health records; patient participation; patient safety; care quality; digital health literacy

Introduction

Background

Web-based patient portals are thought to contribute to improvements in care quality by providing patients with access to their personal health information, empowering them to self-manage their health and become true partners in their own care [1]. As the trend toward patients being able to access their electronic health records accelerates [2], there is a pressing need to evaluate the impact of patient portals, understand their risks and benefits from both patient and provider perspectives, and generate evidence to inform future health policy [3].

Although care is traditionally delivered through face-to-face clinical consultations, patient-provider communication through patient portals is increasingly common [1]. The Care Information Exchange (CIE) is the largest shared personal health records program in the United Kingdom and provides patients with secure web-based access to their health and social care records. Patients can additionally use CIE in different ways: for example, to self-monitor their health by linking home health care devices (eg, activity tracker and blood pressure monitor) to the portal, to communicate with care providers through messaging and videoconferencing, and to check appointments and test results and be signposted to useful weblinks and resources by health and care professionals.

One of the most influential guides for evaluating health care initiatives is the Institute of Medicine's framework, which includes 6 domains of quality of care: effectiveness, safety, timeliness, efficiency, patient centeredness, and equity [4,5]. Effectiveness is about achieving optimal health outcomes by providing appropriate treatment to patients who could benefit and avoiding the underuse and misuse of health services [4,5]. Patient safety seeks to prevent patients from being harmed by the care that is intended to help them [4,5]. Timeliness is about reducing harmful waits and delays, whereas efficiency is about minimizing resource waste [4,5]. Patient centeredness respects patient preferences and needs and values and ensures these are incorporated into clinical decision-making [4,5]. Equity ensures that care does not vary in quality because of differences in patient characteristics such as ethnicity or geographic location [4,5].

Over the last decade, a considerable body of evidence has uncovered important barriers to portal use, enabling the development of portals in line with patient and health service need [6-8]. In contrast, relatively few studies have investigated the relationship between patient portals and quality of care. Some prior evidence demonstrates the beneficial effects of patient portal use, particularly in supporting preventive behaviors and disease control in people with chronic conditions [3,7,9]. A number of studies have documented positive associations between patient portals and patient safety [3,7,10-13], including improved adherence to medical regimens and reductions in medication discrepancies [3]. However, evidence for the impact of patient portals across other domains of quality is sparse, and where evidence does exist, findings have been mixed [3,7]. Among patients who use web-based portals, little is known about which sociodemographic groups perceive the greatest benefits of access to their personal health records. Furthermore, policy makers agree that more evidence is needed to understand the impact of tools that use digital technologies amidst concerns over a growing *digital divide* [14].

Objectives

The aims of this study were to describe perceptions of quality of care among users of a web-based patient portal and to identify the characteristics of portal users who perceive the greatest benefit of portal use.

Methods

Study Design, Participants, and Data Collection

A cross-sectional survey study was conducted to explore patients' views and experiences of using CIE. The questionnaire was administered via Qualtrics (web-based survey platform) and was open for completion between July 1, 2018, and July 1, 2019. At the time of the survey, CIE was deployed to the diverse 2.3 million patients treated in North West London, including patients residing in London and in other geographic locations across England. CIE held records from hospitals and general practitioners in North West London and from 15 hospitals outside of London, in Birmingham, Bristol, Liverpool, Manchester, Scotland, and Wales. All patients registered with the CIE at the time of the survey were invited via email to complete the questionnaire (n=27,411). The email explained the purpose of the study; informed consent was obtained. Patients accessed the questionnaire via a web link in the portal. Patients had to be aged at least 18 years to be registered with CIE. Not all patients registered with CIE were using the portal. With this data set, we have previously evaluated differences between users and nonusers of CIE with respect to their sociodemographic characteristics and demonstrated the importance of addressing educational aspects and digital literacy to ensure equitable and sustainable portal adoption [15]. Our further work has sought to evaluate the impact of web-based patient portals on safety and quality of care from the patient's perspective. Our recent study found that a large proportion of patients are able and willing to use patient portals to participate in identifying and rectifying errors in their personal health records [16]. This study builds on previous work to understand patients' perceptions of the impacts of CIE across 6 domains of care quality.

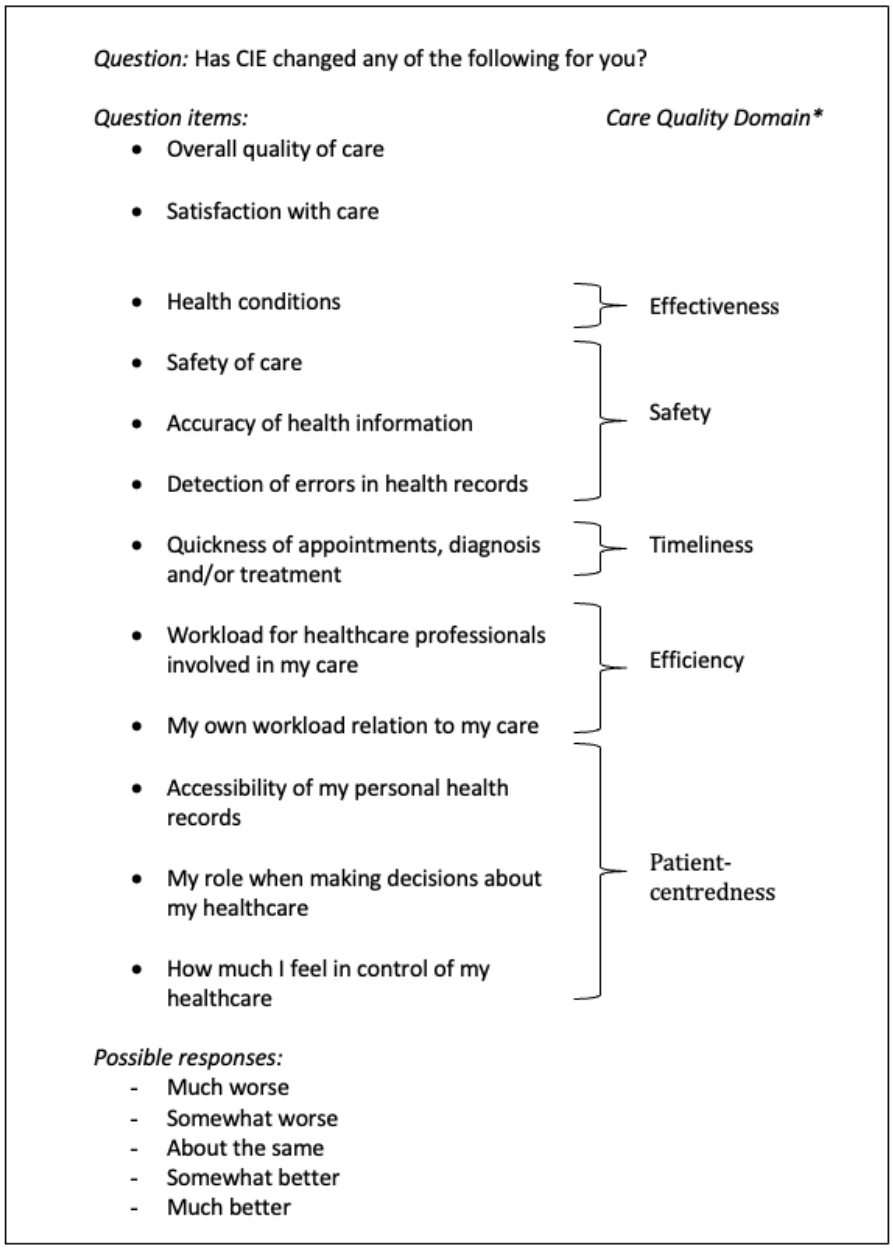
For these analyses, we included patients who had previously accessed and used the CIE portal. We excluded patients who did not provide basic demographics regarding age and sex. Considering this population, a CI of 95%, and a margin of error of 5%, the minimum sample size to ensure representativeness was calculated as 379 respondents. We mapped relevant survey items to the Institute of Medicine's domains of quality of care: effectiveness, safety, timeliness, efficiency, and patient

XSL•FO
RenderX

centeredness [5,17]. Patients' responses to 12 multiple-choice, closed-ended question items were analyzed. Figure 1 outlines the 12 question items, with mapping to care quality domains.

To evaluate equity, we conducted multivariable regression analyses to determine associations between patients' sociodemographic characteristics and perceptions of the impact of CIE on overall care quality and satisfaction with care. The following information was collected to input into multivariable analyses: age, sex, ethnicity, native language, education level, digital health literacy, motivation to be involved in own care, and health status. Respondents' level of motivation to be involved in their own care was assessed via a multiple-choice question ("In general, how motivated to be involved in your healthcare are you?" Possible responses: "A little," "A moderate amount," "A lot," and "Very much"). Digital health literacy was assessed using the eHealth Literacy Scale (eHEALS), developed and validated by Norman and Skinner [18]. The eHEALS tool is an 8-item measure of patients' combined knowledge, comfort, and perceived skills in finding, evaluating, and using internet health resources for health problems [18]. The 8 items are answered on a 5-point Likert scale (1, strongly disagree to 5, strongly agree); total eHEALS scores range from 8 to 40, with a higher score indicating higher digital literacy.

Figure 1. Questionnaire items mapped to care quality domains. The domain of equity was assessed using the methods described in this section. *As defined by the Institute of Medicine, 2001 [5]. CIE: Care Information Exchange.



Data Analysis

We used descriptive statistics to summarize respondent characteristics and patients' responses to question items. Counts and proportions were calculated for categorical variables; means and SDs were calculated for continuous variables. Age was categorized into bands (<30, 31-40, 41-50, 51-65, and ≥65 years). Owing to the small numbers of patients self-identifying to individual categories of ethnic minority background, ethnicity was categorized as "ethnic minorities" or "White."

XSL•FO
RenderX

We conducted multinomial regression analyses to identify sociodemographic characteristics that predict patient-perceived improvement in overall care quality and greater satisfaction with care. To overcome the issue of sparse counts in multivariable modeling (Tables S1 and S2 in Multimedia Appendix 1), "age," "motivation to be involved in own care," "digital health literacy," and "frequency of CIE use" were treated as dichotomous variables, and respondents reporting sex as "other" were excluded. Consistent with previous studies, we selected an eHEALS score ≥26 to indicate higher digital health literacy and <26 to indicate lower digital health literacy [19-23]. We also combined categories of the dependent variable (ie, "much worse" and "somewhat worse" were analyzed as a single category; equally, "somewhat better" and "much better" were combined into 1 category). We performed univariate multinomial logistic analyses to identify possible predictors to include in the multivariable model. We adopted the approach by Hosmer et al [24,25] for variable selection: (1) variables that demonstrated significance (P<.25) in the univariate analyses were entered into the preliminary multivariable model; (2) variables that were nonsignificant at P>.05 according to the likelihood ratio test were removed one at a time according to the variable with the highest P value (backward elimination); (3) to check for suppressor effects, variables excluded during backward selection were re-entered separately into the regression model (forward selection). Only variables that were significant at P<.05 (Likelihood Ratio Test) were retained in the final multinomial regression models. Model quality comparisons were conducted using the Akaike Information Criterion [26], and goodness-of-fit was assessed using the Pearson chi-square statistic [25]. Effect estimates are presented as odds ratios (ORs) with their 95% CIs.

To assess the effects of excluding patients with missing data regarding age and sex, we compared the sociodemographic characteristics of the missing data sample (n=78) and the analysis sample (n=445). We ran a Pearson chi-square test of homogeneity (χ^2) to compare the distribution of item responses between the analysis sample and the missing data sample for the perceived impact of CIE on the overall quality of care and satisfaction with care.

Analyses were conducted using Microsoft Excel (version 16.54) and SPSS software (version 27; IBM Corp).

Ethics Approval

The study was approved as a Service Evaluation at Imperial College Health care NHS Trust (registration number: 296/2018).

Reporting

We followed the reporting recommendations in the Strengthening the Reporting of Observational Studies in Epidemiology Statement (Multimedia Appendix 2). [27].

Results

Respondent Characteristics

Of 1083 patients who responded to the survey, 523 (48.29%) patients who were "CIE users" completed the questionnaire. CIE users who provided basic demographic details regarding age and sex were included in the analysis (445/523, 85.1%; +117% of the minimum target sample size); 14.9% (78/523) of respondents with missing data for age and sex were excluded.

Of 445 respondents, most (n=313, 70.3%) were aged >50 years and 276 (62%) were female. Approximately 1 in 5 (97/445, 21.8%) respondents belonged to an ethnic minority group. Most (292/445, 65.6%) respondents were educated to the degree level or higher, and the mean eHEALS score was 33.6 (SD 6.4, range 8-40); a score ≥26 indicates higher digital health literacy. Of 445 patients, 177 (39.8%) patients reported being in good health; 162 (36.4%) of patients reported that the status of their health was poor. Most (278/445, 62.5%) patients reported being very motivated in their own care. Most (284/445, 63.8%) patients said they used CIE at least once a month, and 93.2% (415/445) of patients said they found CIE useful. Patient characteristics are presented in Table 1.

XSL•FO
RenderX

Table 1. Respondent characteristics (N=445).

	Respondents
Sex, n (%)	
Male	167 (37.5)
Female	276 (62)
Other	2 (0.4)
No response	N/A[a]
Age group (years), n (%)	
<30	22 (4.9)
31-40	48 (10.8)
41-50	62 (13.9)
51-64	166 (37.3)
>65	147 (33)
No response	N/A
Ethnicity, n (%)	
Asian or British Asian	44 (9.9)
Black African or Black Caribbean or Black British	20 (4.5)
Mixed or multiple ethnic groups	11 (2.5)
Other	22 (4.9)
White	343 (77.1)
No response	5 (1.1)
Geographic location, n (%)	
London	284 (63.8)
Other locations in England	145 (32.6)
No response	16 (3.6)
Education, n (%)	
Secondary school or below	118 (26.5)
Undergraduate or professional degree	180 (40.4)
Postgraduate or higher	112 (25.2)
No response	35 (7.9)
Language, n (%)	
English	379 (85.2)
Non-English	58 (13.0)
No response	8 (1.8)
eHealth literacy (eHEALS[b] score), mean (SD; range)	33.6 (6.4; 8-40)
Overall health status, n (%)	
Good or very good	177 (39.8)
Neither good nor poor	106 (23.8)
Poor or very poor	162 (36.4)
No response	0 (0)
Motivation to be involved in own care, n (%)	
Not very much	6 (1.3)
A moderate amount	43 (9.7)

XSL•FO
RenderX

	Respondents
A lot	116 (26.1)
Very much	278 (62.5)
No response	2 (0.4)

[a]N/A: not applicable.

[b]eHEALS: eHealth Literacy Scale.

Patients' Perceptions of the Impact of CIE on the Overall Quality of Care

Patients were asked to consider how CIE has changed the overall quality of care they receive. Of 429 patients who answered this question, 172 (38.7%) reported that the quality of their care was better with CIE. A further 54.6% (243/445) said that their care was about the same, and 3.2% (14/445) of patients said their care was worse (Multimedia Appendix 3).

Patients' Perceptions of the Impact of CIE on Satisfaction With Care

When asked to consider how CIE has changed and how satisfied they are with their care, 43.6% (194/445) of patients said their care was better, 47.6% (212/445) said their care was the same, and 4.3% (19/445) said their care was worse. In addition, 4.5% (20/445) of patients did not respond to this question (Multimedia Appendix 3).

Patients' Perceptions of the Impact of CIE Across 6 Domains of Care Quality

Overview

Patients' responses to a further 10 survey items revealed their perceptions of how CIE use has changed the care they receive across the following domains of quality of care: effectiveness, safety, timeliness, efficiency, and patient centeredness (Table 2).

Table 2. Survey items and patients' responses, mapped to the Institute of Medicine's domains of health care quality (N=445).

Health care quality domain[a] and survey item: "Has CIE changed any of the following...?"	Missing data, n (%)	Much worse, n (%)	Somewhat worse, n (%)	About the same, n (%)	Somewhat better, n (%)	Much better, n (%)
Effective						
Health conditions	30 (6.7)	7 (1.6)	9 (2)	276 (62)	61 (13.7)	62 (13.9)
Safe						
Safety of care	33 (7.4)	7 (1.6)	7 (1.6)	260 (58.4)	68 (15.3)	70 (15.7)
Accuracy of health information	25 (5.6)	9 (2)	11 (2.5)	187 (42.0)	117 (26.3)	96 (21.6)
Detection of errors in health records	32 (7.2)	8 (1.8)	13 (2.9)	246 (55.3)	70 (15.7)	76 (17.1)
Timely						
Quickness of appointments, diagnosis, and/or treatment	29 (6.5)	13 (2.9)	12 (2.7)	212 (47.6)	77 (17.3)	102 (22.9)
Efficient						
Workload for health care professionals involved in my care	31 (7.0)	7 (1.6)	15 (3.4)	265 (59.6)	63 (14.2)	64 (14.4)
My own workload relating to my care	28 (6.3)	11 (2.5)	23 (5.2)	209 (47)	92 (20.7)	82 (18.4)
Patient centeredness						
Accessibility of my personal health records	16 (3.6)	6 (1.3)	9 (2.0)	72 (16.2)	112 (25.2)	230 (51.7)
My role when making decisions about my health care	6 (1.3)	11 (2.5)[b]	11 (2.5)[b]	188 (42.2)[c]	240 (53.9)[d]	240 (53.9)[d]
How much I feel in control of my health care	6 (1.3)	19 (4.3)[e]	19 (4.3)[e]	147 (33)[c]	273 (61.2)[f]	273 (61.2)[f]

[a]As defined by the Institute of Medicine [5].

[b]I feel I have less of a role.

[c]No change.

[d]I feel I have more of a role.

[e]I feel I have less control.

[f]I feel I have more control.

XSL•FO
RenderX

Effectiveness

Patients were asked whether CIE use had changed their health condition. Most (276/445, 62%) patients responded that their health condition was about the same; however, 27.6% (123/445) patients reported that their health condition had improved with CIE use. Only 3.6% (16/445) said their health condition was worse.

Safety

Although many (260/445, 58.4%) patients reported that the safety of the care was the same with CIE; 31% (138/445) felt that their care was safer. Approximately half (213/445, 47.9%) believed that CIE had led to improvements in the accuracy of their health information, and 32.8% (146/445) of patients felt CIE was associated with better detection of errors in the health record. Only 3.2% (14/445) of patients felt the safety of their care was worse with CIE.

Timeliness

Approximately 40% (179/445) of patients felt that the timeliness of their care (being able to access appointments, diagnoses, and treatment quickly) had improved with CIE. Only 5.6% (25/445) said the timeliness of their care was worse, and 47.6% (212/445) said the timeliness of their care was about the same.

Efficiency

Patients were asked whether CIE had changed the workload relating to their health, including both patients' own workload and the workload of health professionals involved in their care. Many (209/445, 47%) patients reported that their own workload was about the same; however, 28.6% (174/445) felt that their workload was better, and 7.7% (34/445) felt their workload was worse. Regarding the impact of CIE on the workload of health professionals, 39.1% (174/445) of patients perceived that this had improved, 59.6% (265/445) believed it to be about the same, and 5% (22/445) thought that it was worse.

Patient Centeredness

Most (342/445, 76.9%) patients reported that CIE had improved the accessibility of their personal health records. A few (72/445, 16.2%) patients felt that the accessibility of their records was about the same with CIE, whereas only 3.3% (15/445) said their records were less accessible. More than half (240/445, 53.9%) of the survey respondents reported that CIE had led to them having more of a role in decision-making, and 61.3% (273/445) feel they have more control of their health care. Only 2.5% (11/445) of patients reported feeling they have less of a role, and 4.3% (19/445) felt they have less control of their health care with CIE.

Equity

To identify the characteristics of CIE users who perceived better overall quality of care and greater satisfaction with care with portal use, patient characteristics and survey responses were entered into univariate and multivariable multinomial regression models.

For the survey item, "How has CIE changed the overall quality of care you have received?" the final multivariable multinomial regression model with 3 predictor variables (ethnicity, digital health literacy, and frequency of CIE use) predicted significantly better than the null (intercept) model ($P<.001$) and Pearson chi-square statistic indicated satisfactory model fit ($\chi^2_8=14.4$; $P=.07$). The results of the regression are presented in Table 3. Patients with higher digital health literacy (eHEALS score≥26) were more likely to report that the overall quality of their care was better with CIE use (OR 2.40, 95% CI 1.07-5.42; $P=.03$). Compared with their White counterparts, patients self-identifying to an ethnic minority group were also more likely to perceive improvements in care quality based on CIE use (OR 2.27, 95% CI 1.26-3.73; $P=.005$). Across ethnic groups, 68% (30/44) of Asian and British Asian patients reported better overall quality of care with CIE use, compared with 45% (9/20) of Black or African or Caribbean or Black British patients, 36.6% (120/328; missing data, n=15) of White patients, and 36% (4/11) of patients from mixed or multiple ethnic groups (Table S1 in Multimedia Appendix 4).

XSL•FO
RenderX

Table 3. Multinomial regression results of patient characteristics and perceived change in overall quality of care with Care Information Exchange use.

	Univariate[a]				Multivariable[a]			
	Worse care quality vs about the same		Better care quality vs about the same		Worse care quality vs about the same		Better care quality vs about the same	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Sex								
Female	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Male	0.47 (0.13-1.74)	.26	1.26 (0.84-1.88)	.26	N/A[b]	N/A	N/A	N/A
Age (years)								
≥65	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
≤64	1.35 (0.41-4.42)	.63	1.28 (0.84-1.94)	.26	N/A	N/A	N/A	N/A
Ethnicity								
White	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Ethnic minority	1.88 (0.49-7.18)	.36	2.27 (1.37-3.78)	.002	2.44 (0.61-9.80)	.21	2.27 (1.26-3.73)	.005
Native language								
English	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Non-English	2.56 (0.66-9.91)	.18	1.81 (1.02-3.21)	.04	—[c]	—	—	—
Education								
Secondary or below	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Undergraduate or professional	4.60 (0.55-38.23)	.16	0.85 (0.53-1.38)	.51	—	—	—	—
Postgraduate or higher	4.00 (0.44-36.76)	.22	0.73 (0.42-1.25)	.25	—	—	—	—
Digital literacy								
Lower digital health literacy	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Higher digital health literacy	1.57 (0.20-12.63)	.67	2.51 (1.15-5.45)	.02	1.51 (0.18-12.42)	.70	2.40 (1.07-5.42)	.03
Health status								
Neither good nor poor	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Poor	0.72 (0.20-2.60)	.62	1.29 (0.77-2.16)	.34	N/A	N/A	N/A	N/A
Good	0.52 (0.14-2.02)	.35	1.22 (0.73-2.03)	.45	N/A	N/A	N/A	N/A
Motivation to be involved in own care								
Not very much or a moderate amount	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
A lot or very much	1.92 (0.24-15.19)	.54	1.67 (0.86-3.24)	.13	—	—	—	—
Frequency of Care Information Exchange use								
Once a month or less	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Once a week or more	1.05 (0.32-3.45)	.94	2.40 (1.59-3.63)	<.001	0.92 (0.24-3.60)	.90	2.31 (1.49-3.58)	<.001

[a]Goodness-of-fit: χ^2_8=14.5; *P*=.07.

[b]N/A: not applicable; variable not entered into the multivariable analyses due to nonsignificance (*P*>.25) in univariate analyses.

[c]Variable excluded from the final multivariable model using a backward elimination approach.

Frequency of CIE Use

Patients using CIE at least once per week were more likely to perceive improved care quality compared with patients using CIE less frequently (OR 2.31, 95% CI 1.49-3.58; *P*<.001). Sensitivity analyses assessing the effects of including or excluding predictor variables that had demonstrated significance in univariate analyses did not alter the results of the multivariable regression.

XSL•FO
RenderX

For the survey item "How has CIE changed how satisfied you are with your care?" the final multivariable model with 3 predictor variables (ethnicity, digital health literacy, and frequency of CIE use) predicted significantly better than the null (intercept) model ($P<.001$) and Pearson chi-square statistic suggested that the model fit the data well ($\chi^2_8=5.6$; $P=.69$). Patients with higher digital health literacy (eHEALS score≥26) were more likely to report greater satisfaction with their care with CIE use, compared with those with lower digital health literacy (OR 2.35, 95% CI 1.09-5.04; $P=.03$; Table 4). CIE use was also associated with greater satisfaction with care among patients belonging to an ethnic minority group compared with White patients (OR 2.12, 95% CI 1.22-3.67; $P=.007$). Cross-tabulation of patients' ethnicity and perceived change in satisfaction with care revealed that 77% (34/44) of Asian or British Asian patients reported greater satisfaction with care with CIE use, compared with 55% (11/20) of Black or African or Caribbean or Black British patients, 36% (4/11) of patients from mixed or multiple ethnic groups, and 42.1% (137/325; missing data n=18) of White patients (Table S2 in Multimedia Appendix 4).

Patients using CIE at least once per week were more likely to report greater satisfaction with care compared with patients using CIE less frequently (OR 2.03, 95% CI 1.31-3.14; $P=.002$).

Sensitivity analyses assessing the effects of including or excluding predictor variables that had demonstrated significance in univariate analyses did not alter the results of the multivariable analyses.

XSL•FO
RenderX

Table 4. Multinomial regression results of patients' sociodemographic characteristics and impact of Care Information Exchange on patient's satisfaction with care.

	Univariate[a]				Multivariable[a]			
	Worse care quality vs about the same		Better care quality vs about the same		Worse care quality vs about the same		Better care quality vs about the same	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Sex								
Female	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Male	0.84 (0.31-2.31)	.74	1.32 (0.88-1.97)	.17	—[b]	—	—	—
Age (years)								
≥65	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
≤64	0.69 (0.27-1.80)	.45	1.215 (0.76-1.75)	.51	N/A[c]	N/A	N/A	N/A
Ethnicity								
White	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Ethnic minority	1.30 (0.35-4.78)	.70	2.32 (1.38-3.90)	.002	1.68 (0.44-6.41)	.45	2.12 (1.22-3.67)	.007
Native language								
English	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Non-English	1.74 (0.47-6.52)	.41	1.63 (0.91-2.89)	.10	—	—	—	—
Education								
Secondary or below	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Undergraduate or Professional	8.15 (1.03-64.80)	.05	1.14 (0.67-1.96)	.63	—	—	—	—
Postgraduate or higher	5.94 (0.67-52.47)	.11	1.11 (0.69-1.80)	.67	—	—	—	—
Digital literacy								
Lower digital health literacy	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Higher digital health literacy	2.29 (0.29-18.03)	.43	2.47 (1.19-5.13)	.02	2.17 (0.27-17.35)	.46	2.35 (1.09-5.04)	.03
Health status								
Neither good nor poor	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Poor	0.92 (0.27-3.16)	.89	1.34 (0.80-2.23)	.27	N/A	N/A	N/A	N/A
Good	1.04 (0.32-3.33)	.95	1.07 (0.65-1.78)	.78	N/A	N/A	N/A	N/A
Motivation to be involved in own care								
Not very much or a moderate amount	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
A lot or very much	3.00 (0.39-23.31)	.29	1.99 (1.04-3.82)	.04	X[d]	X	X	X
Frequency of CIE use								
Once a month or less	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Once a week or more	0.92 (0.32-2.67)	.88	2.13 (1.41-3.23)	<.001	0.90 (0.27-2.95)	.86	2.03 (1.31-3.13)	.002

[a]Goodness-of-fit: χ^2_8=5.6; *P*=.69.

[b]Variable excluded from the final multivariable model using a backward elimination approach.

[c]N/A: not applicable; variable not entered into the multivariable analyses due to nonsignificance (*P*>.25) in univariate analyses.

[d]Variable excluded from the final multivariable model due to 0 cell counts producing unstable estimates.

XSL•FO
RenderX

Missing Data Analysis

Of 523 survey respondents, 78 (14.9%), who had previously used CIE, had missing data regarding age and gender, and these respondents were excluded from our analyses. Meaningful comparisons of sociodemographic characteristics between the missing data sample and the analysis sample were not possible due to considerable additional missing data in the group of 78 respondents excluded from this analysis (Multimedia Appendix 5). There were no differences in the distribution of responses between the analysis sample and the missing data sample for the questionnaire item "How has CIE changed how satisfied you are with your care?" However, patients included in the analysis were more likely to view the impact of CIE on overall quality of care favorably, compared with those in the missing data sample (χ^2_4=10.3; *P*=.04; Multimedia Appendix 6).

Discussion

Principal Findings

Although many portal users perceived no change with CIE use, a large proportion reported better care quality and greater satisfaction with their care. Around 30% patients perceived their care to be safer, more effective, and more efficient with CIE, and approximately 40% reported that the timeliness of appointments, diagnoses, and treatments had improved. The most positive patient-perceived changes were in the domain of patient centeredness: more than half of patients using CIE felt more in control of their health care and able to play a greater role in decision-making. Patients from ethnic minority groups, those with higher digital health literacy, and those using CIE more frequently were more likely to perceive improvements in overall care quality and greater satisfaction with care. Across ethnic groups, patients of Asian or British Asian ethnicity reported the greatest benefits of portal use in terms of improving care quality and satisfaction with care received.

Comparison With Wider Literature

These reports from users of a web-based patient portal in the United Kingdom are consistent with the findings of other patient experience studies in finding that many patients perceive a range of benefits associated with portal use [28-37]. To our knowledge, this is the first empirical study to map patients' experiences against the 6 domains of quality of care to provide broad insight into the perceived effects of portal use from the patient perspective.

Regarding the domain of effectiveness, around 1 in 4 patients in our study believed that CIE use contributed to improving their overall health, and this finding echoes the results of other survey studies and meta-analyses of randomized trials [3,38]. We did not collect information about respondents' medical histories; however, prior studies have shown that portal use may be particularly effective in supporting people with long-term conditions to improve their health, including those with diabetes and hypertension [3,38].

Existing evidence links patient portals to increased medication safety through patients possessing greater knowledge about their medicines, improved medication adherence, and increased reporting of medicine discrepancies [3,39-41]. Our study has shown that patients perceive additional safety impacts of web-based portals including improved accuracy of personal health information and detection of health record errors. Our previous work, together with studies conducted in the United States, has demonstrated that around 1 in 5 patients who access their web-based personal health records can, and do, notice errors in their records, and most patients would like to play an active role in rectifying these discrepancies [16,42]. Moreover, Blease et al [40] have shown that enabling informal carers to access the electronic health records of vulnerable patients (eg, people with serious mental illness) can help to prevent medication errors, delayed diagnoses, and other patient safety risks.

Regarding the efficiency domain, more than one-third of patients in our study perceived their own workload relating to their health had changed for the better. In a previous survey study in Canada, patients reported that web-based portals save time when scheduling appointments, patients needed to repeat themselves less during appointments, and portal use meant that patients could avoid unnecessary clinic visits [43]. Similarly, a review of randomized trials found a reduction in health care use (or no change) when patients have access to their electronic health records [3]. No experimental trials have investigated the impact of web-based portals on the timeliness of care delivery [3]; however, approximately 40% of the patients in our study perceived that CIE enabled them to access appointments, diagnoses, and treatment more quickly.

A growing body of evidence suggests that patients who are engaged in their care are more likely to adhere to medication and treatment plans, take up screening opportunities and prevention practices, participate in the detection of errors and safety risks, and adopt effective management strategies for chronic conditions [28,44-47]. The findings of this and numerous other survey studies have consistently found that patients feel more in control of their health care and better able to play a role in decision-making with access to their personal health records [28,33,34,37,40].

Regarding equity, our findings are consistent with previous research demonstrating that patients experiencing barriers to accessing web-based portals (including low digital literacy), and those with low levels of engagement in technology-enabled care are less likely to report that portals improve their health [38,48]. Previous research has also demonstrated that portal uptake is lower among patients belonging to ethnic minority groups [38]. However, in line with survey studies of portal users in the United States [28,29], we found that CIE users self-identifying to an ethnic minority group were more likely to report better care quality and greater satisfaction with care. Gerard et al [29] found that, compared with White patients, patients of Asian ethnicity in the United States were twice as likely to report the benefits of portal use; our study echoes this finding in the United Kingdom. Sharing electronic health records with patients appears to increase transparency and trust and strengthens the relationship between patients and their providers [44]. These benefits may be particularly important for ethnic minority groups to feel satisfied with their care; however, further

XSL•FO
RenderX

qualitative research is needed to understand the mechanisms of portal adoption across different ethnic minority groups.

Of note, we found that patients who use CIE frequently were more likely to perceive improvements in overall care quality (and greater satisfaction with care). However, the direction of this effect is unclear. We suggest that this mechanism is likely to be circular, with initial portal use leading to perceived improvements in care quality, resulting in greater satisfaction with care, prompting increased portal use. In this way, the perception of quality of care could serve as a mechanism of sustained portal adoption. This theory is consistent with the Technology Acceptance Model, which suggests that use behavior (actual use) is partly predicated by the perceived benefits of using the technology [49]. In the study by Portz et al [50], which used the Technology Acceptance Model to explore portal use among older adults with chronic conditions, patient-perceived usefulness (communicating with care provider, saving time and money, addressing concerns without a clinic visit) was linked to frequent use of specific portal features, including the message center, pharmacy center, and viewing laboratory results. Further evaluation of CIE should include developing and testing a "Theory of Change" to determine how and why portal use leads to greater satisfaction with care in some patient groups [51].

Policy Implications and Future Research

This study confirms the importance of addressing "the digital divide" as a policy priority to ensure equitable access to the benefits of patient portals for all patients [14,52]. Crosscutting interventions with system impacts, including user-centric design of portal platforms that adhere to accessibility, legibility and readability standards, and a commitment to "safety net" strategies such as the provision of low-cost, Wi-Fi–enabled devices or patient outreach programs, could all help to ensure that traditionally underserved groups can benefit from portal use [40,53]. More work is required to understand the relative effectiveness of these interventions, such that equity of access and adoption can be achieved for all patients. However, beyond literacy and technology access, our findings suggest that there are other potential avenues for addressing health disparities by expanding patient portal use to underrepresented groups. That ethnic minority groups see greater benefits in accessing their personal health records is worthy of further careful inquiry. Further research using qualitative methods would help to elucidate the mechanisms of patient portal adoption among ethnic minority communities.

Strengths and Limitations

We mapped survey items to the Institute of Medicine's 6 domains to provide a broad overview of perceptions of care quality among CIE users. However, our questionnaire was not designed to evaluate the domains of care quality as multidimensional constructs. There is a need to develop instruments that can measure subjective accounts of care quality as seen through the patient lens; developing and validating such a questionnaire could be the focus of future work.

We recruited a diverse sample, with one-third of respondents residing outside London and 1 in 5 self-identifying to an ethnic minority group. However, the numbers of patients in subgroups of ethnic minority were small. As such, we combined categories of ethnicity for the multivariable regression. Research exploring issues of equity should disaggregate ethnic categories where possible so the experiences of different ethnic groups can be understood [54]. Although we ran cross-tabulations to explore differences between ethnic groups, the numbers were small and may not generalize to larger populations.

Our web-based recruitment strategy may have introduced selection bias because web-based survey studies may favor the inclusion of patients who are digitally literate and more able to fully engage with patient portals. Our sample only included users of a web-based portal, and our findings are based on patient self-reported and perceived changes in care quality based on portal use. As such, and due to the nature of the study design, we cannot make any causal claims about the impact of patient portals on the quality of care. Building on limited existing evidence from controlled trials [2,3], further experimental or quasi-experimental studies should test the relationship between patient portal use and care quality using validated end points.

Conclusions

A large proportion of CIE users perceived better overall quality of care and greater satisfaction with care, although many portal users reported no change. Perceived improvements were reported across all 6 domains of care quality, with the most favorable in the domain of patient centeredness. Patients from ethnic minority backgrounds (particularly Asian or British Asian) and those with higher digital health literacy perceived the greatest benefits of CIE use. With national policy directed toward addressing health disparities, patient portals could be valuable in improving care quality for patients in underrepresented groups, providing the needs of digitally disempowered patients are addressed. Further research should test the relationship between patient portal use and validated measures of the domains of care quality.

Acknowledgments

The research was funded by the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre and supported by the NIHR Imperial Biomedical Research Centre. The research was enabled by the Imperial Clinical Analytics Research and Evaluation environment and used the Imperial Clinical Analytics Research and Evaluation team and data resources. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR or the Department of Health and Social Care.

XSL•FO
RenderX

Authors' Contributions

ALN, LF, MK, and EKM designed the study. ALN and LF administered the survey. RL conducted the analyses. RL drafted the manuscript. All authors contributed to the revision, editing, and approval of the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Cross-tabulation of patients' sociodemographic characteristics and their perceptions of the impact of Care Information Exchange on satisfaction with care.
[DOCX File , 42 KB - jmir_v24i11e39973_app1.docx]

Multimedia Appendix 2
The Strengthening the Reporting of Observational Studies in Epidemiology Statement—Checklist of Items That Should Be Addressed in Reports of Observational Studies.
[DOCX File , 328 KB - jmir_v24i11e39973_app2.docx]

Multimedia Appendix 3
Patients' perceptions of the impact of Care Information Exchange on (1) overall quality of care and (2) satisfaction with care.
[DOCX File , 20 KB - jmir_v24i11e39973_app3.docx]

Multimedia Appendix 4
Cross-tabulation of patients' ethnicity and perceived change in overall quality of care with Care Information Exchange use.
[DOCX File , 24 KB - jmir_v24i11e39973_app4.docx]

Multimedia Appendix 5
Sociodemographic characteristics of patients in the missing data sample and in the analysis sample.
[DOCX File , 25 KB - jmir_v24i11e39973_app5.docx]

Multimedia Appendix 6
Missing data analysis for questionnaire items.
[DOCX File , 25 KB - jmir_v24i11e39973_app6.docx]

References

1. Glöggler M, Ammenwerth E. Development and validation of a useful taxonomy of patient portals based on characteristics of patient engagement. Methods Inf Med 2021 Jun;60(S 01):e44-e55 [FREE Full text] [doi: 10.1055/s-0041-1730284] [Medline: 34243191]
2. Accelerating patient access to their record. NHS Digital. URL: https://digital.nhs.uk/services/nhs-app/nhs-app-guidance-for-gp-practices/accelerating-patient-access-to-their-record [accessed 2022-01-13]
3. Neves AL, Freise L, Laranjo L, Carter AW, Darzi A, Mayer E. Impact of providing patients access to electronic health records on quality and safety of care: a systematic review and meta-analysis. BMJ Qual Saf 2020 Dec;29(12):1019-1032 [FREE Full text] [doi: 10.1136/bmjqs-2019-010581] [Medline: 32532814]
4. Six domains of health care quality. Agency for Healthcare Research and Quality. 2018 Nov. URL: https://www.ahrq.gov/talkingquality/measures/six-domains.html [accessed 2022-04-05]
5. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm A New Health System for the 21st Century. Washington, D.C., United States: National Academies Press; 2001.
6. 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]
7. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021 Sep 08;23(9):e26189 [FREE Full text] [doi: 10.2196/26189] [Medline: 34494966]
8. 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]
9. 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]

XSL•FO
RenderX

10. Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract 2015 Mar;65(632):e141-e151 [FREE Full text] [doi: 10.3399/bjgp15X683941] [Medline: 25733435]
11. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Sep 08;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]
12. 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]
13. 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]
14. Maguire D, Honeyman M, Fenney D, Jabbal J. Shaping the future of digital technology in health and social care. The King's Fund. 2021 Apr 7. URL: https://www.kingsfund.org.uk/publications/future-digital-technology-health-social-care [accessed 2022-05-30]
15. Neves AL, Smalley KR, Freise L, Harrison P, Darzi A, Mayer EK. Determinants of use of the care information exchange portal: cross-sectional study. J Med Internet Res 2021 Nov 11;23(11):e23481 [FREE Full text] [doi: 10.2196/23481] [Medline: 34762063]
16. Lear R, Freise L, Kybert M, Darzi A, Neves A, Mayer E. Patients' willingness and ability to identify and respond to errors in their personal health records: mixed methods analysis of cross-sectional survey data. J Med Internet Res 2022 Jul 08;24(7):e37226 [FREE Full text] [doi: 10.2196/37226] [Medline: 35802397]
17. Institute of Medicine. Health IT and Patient Safety: Building Safer Systems for Better Care. Washington, DC: The National Academies Press; 2012.
18. 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]
19. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
20. Van der Vaart R, Van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
21. Shiferaw KB, Mehari EA. Internet use and eHealth literacy among health-care professionals in a resource limited setting: a cross-sectional survey. Adv Med Educ Pract 2019 Jul;Volume 10:563-570. [doi: 10.2147/amep.s205414]
22. Milne RA, Puts MT, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of high eHealth literacy in primary lung cancer survivors. J Cancer Educ 2015 Dec;30(4):685-692. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]
23. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. [Developing Japanese version of the eHealth Literacy Scale (eHEALS)]. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
24. Hosmer D, Lemeshow S, Sturdivant R. Applied Logistic Regression, Third Edition. New York: John Wiley & Sons; 2013.
25. Zhang Z. Model building strategy for logistic regression: purposeful selection. Ann Transl Med 2016 Mar;4(6):111 [FREE Full text] [doi: 10.21037/atm.2016.02.15] [Medline: 27127764]
26. Snipes M, Taylor D. Model selection and Akaike information criteria: an example from wine ratings and prices. Wine Econ Policy 2014 Jun;3(1):3-9. [doi: 10.1016/j.wep.2014.03.001]
27. Vandenbroucke JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. PLoS Med 2007 Oct 16;4(10):e297 [FREE Full text] [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
28. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
29. 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]
30. Peck P, Torous J, Shanahan M, Fossa A, Greenberg W. Patient access to electronic psychiatric records: a pilot study. Health Policy Technol 2017 Sep;6(3):309-315. [doi: 10.1016/j.hlpt.2017.06.003]
31. Kayastha N, Pollak K, LeBlanc T. Open oncology notes: a qualitative study of oncology patients' experiences reading their cancer care notes. J Oncol Pract 2018 Apr;14(4):e251-e258. [doi: 10.1200/jop.2017.028605]
32. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]

XSL•FO
RenderX

33. Oster NV, Jackson SL, Dhanireddy S, Mejilla R, Ralston JD, Leveille S, et al. Patient access to online visit notes: perceptions of doctors and patients at an urban HIV/AIDS clinic. J Int Assoc Provid AIDS Care 2015;14(4):306-312 [FREE Full text] [doi: 10.1177/2325957414526783] [Medline: 24729072]
34. Weinert C. Giving doctors' daily progress notes to hospitalized patients and families to improve patient experience. Am J Med Qual 2017;32(1):58-65. [doi: 10.1177/1062860615610424] [Medline: 26453323]
35. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with OpenNotes. Appl Clin Inform 2019 Jan;10(1):10-18 [FREE Full text] [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
36. Chu D, Lessard D, Laymouna MA, Engler K, Schuster T, Ma Y, et al. Understanding the risks and benefits of a patient portal configured for HIV care: patient and healthcare professional perspectives. J Pers Med 2022 Feb 19;12(2):314 [FREE Full text] [doi: 10.3390/jpm12020314] [Medline: 35207803]
37. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
38. 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]
39. Chimowitz H, Gerard M, Fossa A, Bourgeois F, Bell SK. Empowering informal caregivers with health information: OpenNotes as a safety strategy. Jt Comm J Qual Patient Saf 2018 Mar;44(3):130-136. [doi: 10.1016/j.jcjq.2017.09.004] [Medline: 29499809]
40. Blease C, Dong Z, Torous J, Walker J, Hägglund M, DesRoches CM. Association of patients reading clinical notes with perception of medication adherence among persons with serious mental illness. JAMA Netw Open 2021 Mar 01;4(3):e212823 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2823] [Medline: 33760088]
41. DesRoches CM, Bell SK, Dong Z, Elmore J, Fernandez L, Fitzgerald P, et al. Patients managing medications and reading their visit notes: a survey of opennotes participants. Ann Intern Med 2019 Jul 02;171(1):69-71. [doi: 10.7326/M18-3197] [Medline: 31132794]
42. Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. JAMA Netw Open 2020 Jun 01;3(6):e205867 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5867] [Medline: 32515797]
43. Graham TA, Ali S, Avdagovska M, Ballermann M. Effects of a web-based patient portal on patient satisfaction and missed appointment rates: survey study. J Med Internet Res 2020 May 19;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
44. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2017 Apr;26(4):262-270 [FREE Full text] [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
45. Bell SK, Folcarelli PH, Anselmo MK, Crotty BH, Flier LA, Walker J. Connecting patients and clinicians: the anticipated effects of open notes on patient safety and quality of care. Jt Comm J Qual Patient Saf 2015 Aug;41(8):378-384. [doi: 10.1016/s1553-7250(15)41049-9] [Medline: 26215527]
46. Náfrádi L, Nakamoto K, Schulz PJ. Is patient empowerment the key to promote adherence? A systematic review of the relationship between self-efficacy, health locus of control and medication adherence. PLoS One 2017;12(10):e0186458 [FREE Full text] [doi: 10.1371/journal.pone.0186458] [Medline: 29040335]
47. Yeom H, Lee J. Impact of autonomy support on the association between personal control, healthy behaviors, and psychological well-being among patients with hypertension and cardiovascular comorbidities. Int J Environ Res Public Health 2022 Mar 31;19(7):4132 [FREE Full text] [doi: 10.3390/ijerph19074132] [Medline: 35409814]
48. Cross DA, Levin Z, Raj M. Patient portal use, perceptions of electronic health record value, and self-rated primary care quality among older adults: cross-sectional survey. J Med Internet Res 2021 May 10;23(5):e22549 [FREE Full text] [doi: 10.2196/22549] [Medline: 33970111]
49. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
50. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
51. De Silva MJ, Breuer E, Lee L, Asher L, Chowdhary N, Lund C, et al. Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. Trials 2014 Jul 05;15:267 [FREE Full text] [doi: 10.1186/1745-6215-15-267] [Medline: 24996765]
52. Why digital inclusion matters to health and social care. NHS Digital. URL: https://digital.nhs.uk/about-nhs-digital/our-work/digital-inclusion/digital-inclusion-in-health-and-social-care#document-content [accessed 2022-05-30]

XSL•FO
RenderX

53. Grossman LV, Masterson Creber RM, Benda NC, 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]
54. Kapadia D, Zhang J, Salway S, Nazroo J, Booth A, Villarroel-Williams N, et al. Ethnic inequalities in healthcare: a rapid evidence review. NHS Race & Health Observatory. 2002 Feb. URL: https://www.nhsrho.org/wp-content/uploads/2022/02/RHO-Rapid-Review-Final-Report_v.7.pdf [accessed 2022-05-30]

Abbreviations

CIE: Care Information Exchange
eHEALS: eHealth Literacy Scale
OR: odds ratio

Edited by R Kukafka; submitted 30.05.22; peer-reviewed by J Lee, C Sullivan, B Iott; comments to author 26.06.22; revised version received 15.08.22; accepted 03.11.22; published 17.11.22.

Please cite as:
Lear R, Freise L, Kybert M, Darzi A, Neves AL, Mayer EK
Perceptions of Quality of Care Among Users of a Web-Based Patient Portal: Cross-sectional Survey Analysis
J Med Internet Res 2022;24(11):e39973
URL: https://www.jmir.org/2022/11/e39973
doi:10.2196/39973
PMID:36394922

©Rachael Lear, Lisa Freise, Matthew Kybert, Ara Darzi, Ana Luisa Neves, Erik K Mayer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Usage of a Web-Based eHealth Intervention for Women With Stress Urinary Incontinence: Mixed Methods Study

Lotte Firet[1], MSc; Theodora Alberta Maria Teunissen[1], MD, PhD; Rudolf Bertijn Kool[2], MD, PhD; Kim Josephina Bernadette Notten[3], MD, PhD; Antoinette Leonarda Maria Lagro-Janssen[1], MD, PhD; Huub van der Vaart[4], MD, PhD; Willem Jan Jozef Assendelft[1], MD, PhD

[1]Department of Primary and Community Care, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[2]IQ Healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[3]Department of Obstetrics and Gynecology, Radboud University Medical Center, Nijmegen, Netherlands
[4]Department of Gynecology, University Medical Center Utrecht, Utrecht, Netherlands

Corresponding Author:
Lotte Firet, MSc
Department of Primary and Community Care
Radboud Institute for Health Sciences
Radboud University Medical Center
Geert Grooteplein 21
Nijmegen, Postbox 9101
Netherlands
Phone: 31 243618181
Email: lotte.firet@radboudumc.nl

Abstract

Background: Stress urinary incontinence (SUI) is highly prevalent among women and has an impact on physical and mental well-being. eHealth with pelvic floor muscle training (PFMT) has shown to be effective in reducing complaints. The usage and nonusage attrition of eHealth for SUI is unknown, but knowledge about users and their usage patterns is crucial for implementation purposes.

Objective: This study aimed to evaluate how an eHealth intervention for SUI was used and by whom, explore reasons for nonusage attrition, and determine what factors are associated with usage.

Methods: In this observational, mixed methods study, women with SUI independently registered to a web-based eHealth intervention, *Baas over je blaas*, a translation of the Swedish internet program *Tät-treatment of stress urinary incontinence*. Log-in data were collected during 3-month access to the website, and surveys were sent at baseline. Participants were divided into three user groups (low, intermediate, and high) and were compared based on sociodemographic and incontinence-related characteristics. Nominal logistic regression analysis was used to study factors associated with eHealth usage. Qualitative content analysis was used for open-ended questions about nonusage attrition and about facilitators of and barriers to eHealth usage.

Results: Participants (n=561) had a mean age of 50.3 (SD 12.1) years, and most of them (340/553, 61.5%) had never visited a health care professional for SUI before. Most users were low users (295/515, 57.3%), followed by intermediate users (133/515, 25.8%) and high users (87/515, 16.9%). User groups differed significantly in age (48.3, SD 12 years; 52.1, SD 11.6 years; and 55.3, SD 10.9 years; P<.001) and in their expected ability to train the pelvic floor muscles (7.5, SD 1.4; 7.7, SD 1.4; and 8.1, SD 1.5 for low, intermediate, and high users, respectively; P=.006). Nonusage attrition was mainly caused by problems in integrating PFMT into everyday life. High age (>50 years), previous PFMT, and high expected ability to train the pelvic floor muscles are associated with high usage. Facilitators for eHealth usage were the clear explanation of exercises and the possibility of self-management. Barriers were its noncommittal character and the absence of personal contact.

Conclusions: eHealth fulfills a need for women with SUI who have never received treatment. Those who discontinued prematurely did so mainly because it was difficult to integrate the training schedule into their everyday lives. High eHealth usage was more likely for women aged >50 years, with previous PFMT, and with high expectations about their ability to train the pelvic floor muscles. Knowledge of these user characteristics can guide clinicians and correct their misunderstandings about the suitable target population for this intervention. Furthermore, strategies for reinforcing expectations and self-efficacy are important to upscale eHealth usage, together with paying attention to people's need for personal contact.

XSL•FO
RenderX

International Registered Report Identifier (IRRID): RR2-10.2196/13164

(J Med Internet Res 2022;24(11):e38255) doi:10.2196/38255

KEYWORDS

eHealth; urinary incontinence; women; usage; nonattrition; adherence; implementation science; pelvic floor muscle training; mixed methods design

Introduction

Background

Urinary incontinence is a common condition with a serious impact on quality of life and well-being [1,2]. Stress urinary incontinence (SUI) is a prevalent subtype, which is defined as the complaint of any involuntary urinary leakage on effort, exertion, sneezing, or coughing [1]. SUI affects 1 in 4 middle-aged adult women [3,4]. It can lead to psychological problems, such as fear of producing malodors that can be detected by others, shame, or even depression [2,5,6]. Furthermore, it hampers physical mobility by interfering with daily activities, causing the affected individuals to avoid work duties, sport and exercise, or social activities as these may provoke urinary leakage [6].

Pelvic floor muscle training (PFMT) is an evidence-based treatment option for SUI and is recommended as first-line treatment by a general practitioner (GP), nurse, or physiotherapist [7,8]. However, only a minority of women with SUI receive treatment because GPs tend to underdiagnose urinary incontinence and women are not likely to consult a health care professional with this problem [9]. These women feel ashamed, do not prioritize this problem, or believe that no suitable therapy is available [10-12]. Digital treatment options for SUI are promising and upcoming because they gain a broad reach by lowering the threshold to seek help [8]. Various randomized controlled trials (RCTs) and a cohort study showed that web-based (eHealth) and app-based (mobile health) self-management interventions with PFMT are effective in reducing or stopping incontinence [13-17]. Women are highly satisfied with these interventions as they give them the opportunity to deal with the problem themselves and thus promote independence [18-20].

eHealth as a self-management intervention requires users to engage with it for the recommended training period and it requires them to guide themselves throughout the program. Our previous study showed that women need to possess self-efficacy to adopt eHealth for SUI [21]. However, these requirements can be challenging, which is reflected by the high rate of nonusage in eHealth interventions [22-24]. Eysenbach [24] uses the term *nonusage attrition* to refer to the phenomenon of people prematurely discontinuing eHealth usage. Various factors could contribute to this, such as demographic factors, absence of personal or face-to-face contact, push factors (such as reminders), or external events [24]. A systematic review showed that predictors for adherence to web-based psychological interventions are being female, having high expectations, and having therapist support (eg, email support) [25].

Objectives

Extended information on real-life usage and nonusage attrition is lacking in the case of eHealth for urinary incontinence. Knowledge of how eHealth is actually used and when and why people stop using it could lead to further improvement of the design of the intervention, and such improvements could increase usage on a large scale and thus contribute to wide implementation. Furthermore, knowledge of factors associated with usage can guide clinicians in understanding the target audience for whom eHealth would be a suitable treatment option. As part of an implementation project, this study had a three-fold aim: (1) to evaluate how an eHealth intervention for SUI is used and by whom, (2) to explore reasons for nonusage attrition, and (3) to determine what factors are associated with usage. As adherence to regular PFMT is hard to maintain, we hypothesized that, as with other eHealth interventions, nonusage attrition rates may be high [26].

Methods

Design

We used a mixed methods design to study the usage of an eHealth intervention for women with SUI. The quantitative strand consisted of technical log-in statistics and data from web-based questionnaires. The qualitative strand consisted of data from open-ended web-based survey questions. A detailed description of the study has been published previously [27]. The study was registered in the Dutch Trial Registry (NTR6956) prospectively, which is now included in the International Clinical Trial Registry Platform. The CONSORT-eHealth (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) criteria that are applicable to this study will be reported [28].

Participants

Dutch women were recruited through news items in local newspapers, in magazines, on websites, or on social media channels between July 2018 and March 2019. Google AdWords (Google LLC) was used to make our website more retrievable. GPs in the University's network were asked to place leaflets or posters in their waiting rooms. Women who were interested could register on the web-based eHealth intervention, and after providing informed consent, they received a short questionnaire that enabled the researcher to check their eligibility. Eligibility criteria are published in detail elsewhere [27], but, in short, women were included when aged >18 years and when having SUI or mixed urinary incontinence, which is a combination of SUI and urgency urinary incontinence [27], meaning involuntary leakage accompanied by or immediately preceded by urgency [1]. Diagnosis was based on self-assessment questions from the Questionnaire for Female Urinary Incontinence Diagnosis [29].

XSL•FO
RenderX

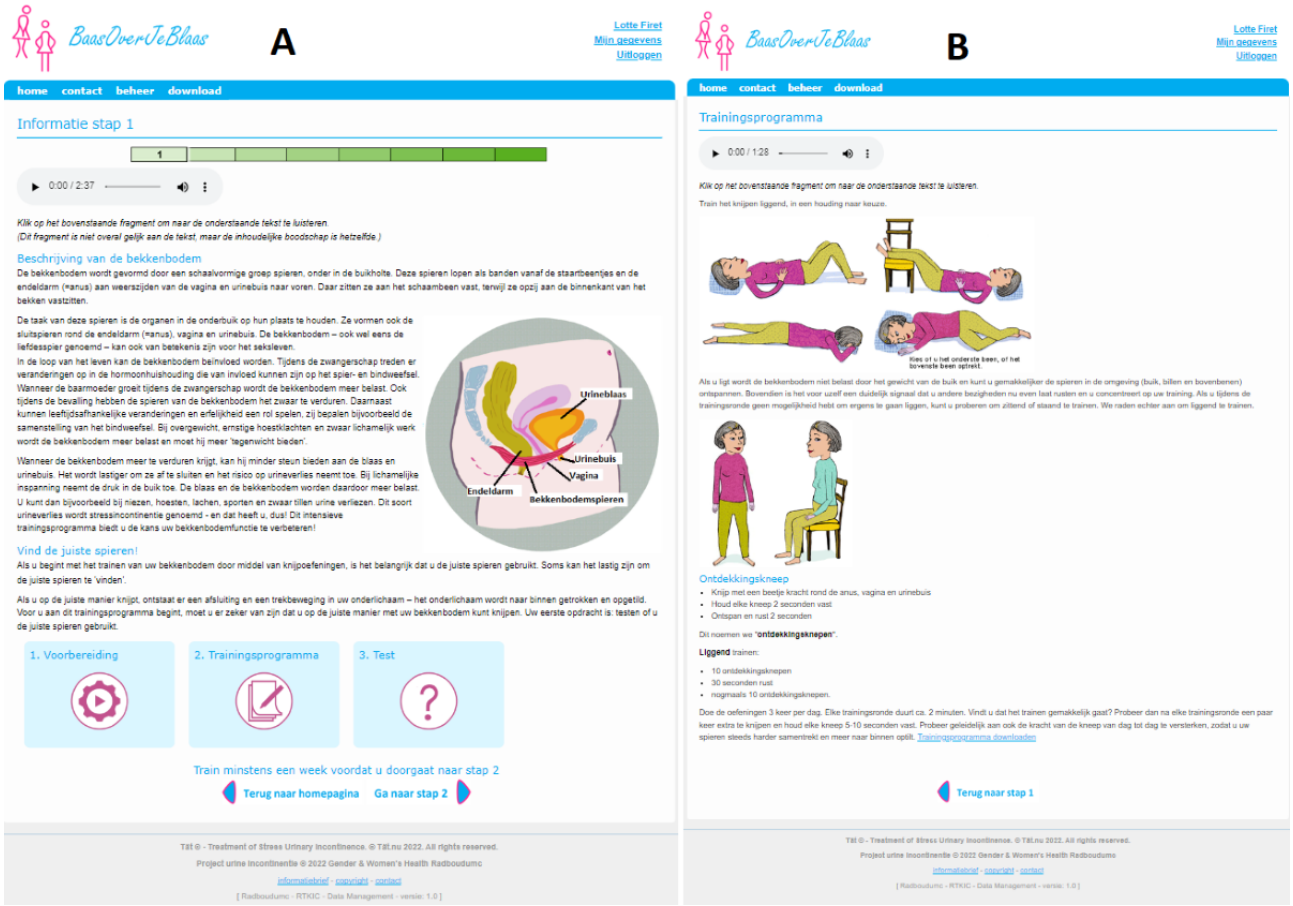
A woman was considered to have SUI if she replied positive to the following question: “Do you lose urine during quick moments such as coughing, sneezing, jumping, or lifting something up?” The researcher checked the eligibility criteria and then sent the web-based baseline questionnaire, which provided access to the website after completion. To use the eHealth intervention or to participate in this study, participants neither had to pay nor were they reimbursed.

Intervention

The eHealth intervention, *Baas over je blaas*, is a translation of the Swedish eHealth intervention, *Tät-treatment of stress urinary incontinence* [30], the effectiveness of which was shown in an RCT [16,31]. It is a web-based password-protected intervention addressing PFMT. The developers of the intervention are assembled in the eContinence group from Umeå University, Sweden. They translated the program into Dutch and gave their permission to use it for research purposes through a noncommercial license agreement. The copyright of the program, Tät-treatment of stress urinary incontinence, belongs to the eContinence group at Umeå University, and the trademark is registered by the Swedish Patent and Registration office for eContinence AB, a Swedish eHealth company founded in July 2021 with the aim of maintaining, distributing, commercializing, and further developing the programs created within the research project [32]. The website was secured via HTTP Secure and hosted on the Apache web server that belonged to the Radboud University Medical Center. Data were saved in a MySQL database on the password-protected webserver. Before the study, the test version of the intervention was pilot-tested with women who varied in age, education level, and profession. Technical issues were resolved, and a new video with explanations about the program was recorded and uploaded on the log-in page. After the test phase, version 1.0 remained *frozen* during the entire study period.

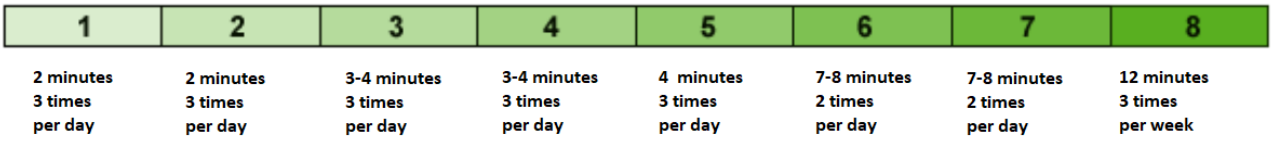
The core content of the intervention consisted of 4 different pelvic floor muscle exercises that were addressed in 8 escalating modules with increasing intensity and complexity. The modules contained background information about incontinence and pelvic floor muscles, a training program, and a test exercise (Figure 1). Cognitive behavioral assignments were included to induce lifestyle changes. Information was provided via text, illustrations, and audio fragments and could be downloaded as PDF file. Women were advised to perform training for at least one week per module (Figure 2), and after that, the test exercise enabled them to check whether they had gained the skills required for continuation. Access to the next module was gained after the woman had completed a training report at the end of each module, which contained 2 questions about the frequency and time they spent on that module. Women were advised to consult their GP in case of no progression or if they were unable to perform the exercises.

Figure 1. Screenshot of module number 1 (A), screenshot of training program 1 (B).



XSL•FO
RenderX

Figure 2. Training schedule per module.



The intervention was accessible for 3 months, and women could do the training at their own pace. A total of 3 months after the first log-in, women had 2 weeks to download the exercises of all 8 modules, and thereafter, access was closed. This restriction of access to the eHealth intervention was chosen to have a clear cutoff point for the collection of log-in data.

There was no face-to-face contact during the entire study, but the researcher was available for both content-related and technology-related questions through email (asynchronous communication). To stimulate usage, email reminders were sent if participants did not log in for 1 week, with a maximum of 2 reminders per module. Women could unsubscribe for reminders via email.

Outcomes—Quantitative

Demographic and Incontinence-Related Variables

The baseline survey contained sociodemographic items (age, education level, and recruitment method) and incontinence-related items (type of incontinence, burden, duration, incontinence aid usage, previous contact with a health care professional, previous PFMT, expected ability to train the pelvic floor muscles, and expected training result). This questionnaire has been extensively described in the protocol [27]. Education level was divided into two levels: low (primary and lower secondary education) versus high (from upper secondary level to doctoral equivalent level). The expected ability to train the pelvic floor muscles was assessed on a 10-point scale ranging from 1 (very low expectations) to 10 (very high expectations) [31]. Severity of incontinence was assessed by the International Consultation on Incontinence Questionnaire–Urinary Incontinence Short Form (ICIQ-UI SF), which, based on their total score (0-21), allows women to be divided into severity categories: slight (1-5), moderate (6-12), severe (13-18), or very severe (19-21). Quality of life was assessed using the disease-specific International Consultation on Incontinence Questionnaire for Lower Urinary Tract Symptoms–Quality of Life (ICIQ LUTS-QoL), resulting in a score ranging between 19 and 76, with high score implying great impact on quality of life.

Usage

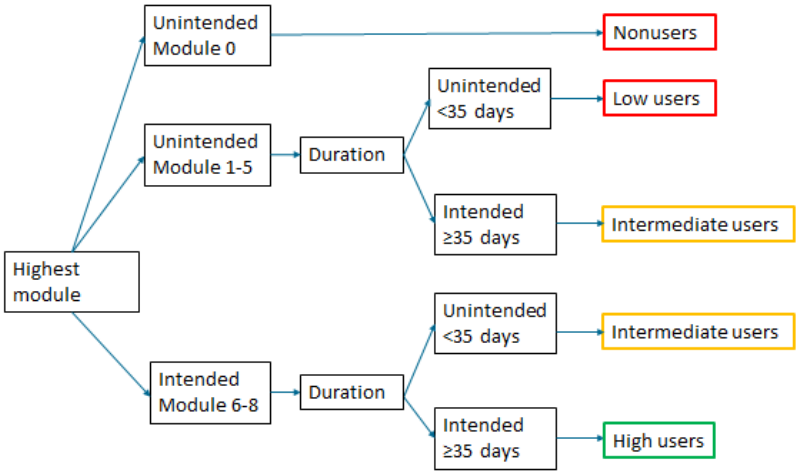
Log data were collected during the 3 months when participants had access to the eHealth intervention, of which the following three user parameters were defined: module number, frequency, and duration. Module number was the module that a participant had reached when access to the website was closed. Frequency was the total number of log-ins. Duration was the total number of days between the first log-in and the date on which the last training report was completed. We chose the date of the last training report instead of the last log-in date because it is a better reflection of the period in which women actively used the intervention. The last log-in date could possibly reflect women who had stopped using the intervention and logged in shortly before access to the intervention closed, to download the training program, for example. The date of the last training report could not be used for women who dropped out in module 1 because they did not complete any training report; therefore, duration for this group was the number of days between first log-in and last log-in. We corrected for false prolonged duration by choosing the previous last log-in date when women in module 1 had a duration of >90 days (12/515, 2.3%).

User Groups

For the construction of user groups, the concept of *intended use* was applied, which is defined as “the extent to which individuals should experience the content (of the intervention) to derive maximum benefit from the intervention, as defined or implied by its creators” [33]. In this study, *intended use* was defined for two usage parameters: module number and duration. A module number was defined as *intended* if women completed at least module 5 because all exercises would have been addressed after 5 modules. Duration was defined as *intended* if it comprised at least 35 days, which is the multiplication of 5 modules with the recommended training duration of at least one week. On the basis of these usage parameters, we divided participants into the following user groups: non, low, intermediate, and high users (Figure 3). Nonusers completed the baseline questionnaire, but never logged in and, therefore, received no further questionnaires. Low users reached an unintended module number and had an unintended duration. Intermediate users had a combination of unintended module number with intended duration or vice versa. High users reached an intended module number and had an intended duration.

XSL•FO
RenderX

Figure 3. Flowchart of user groups based on module number and duration.



Adherence

Exercise adherence was defined as the percentage of time spent on PFMT out of expected time spent on PFMT and was measured using the training reports. Adherence was categorized into three levels: high (>80%), moderate (20%-80%), and low (<20%) adherence [34]. The expected time spent on PFMT was based on the prespecified training schedule (Figure 2).

Outcomes—Qualitative: Facilitators of and Barriers to eHealth Usage and Reasons of Nonusage Attrition

After the intervention, which was 3 months after the first log-in, all women who ever logged in to the website (515/515, 100%) received a web-based survey. Nonresponders were approached via email first and then via telephone to try to collect data of all user groups. This survey contained 2 open-ended questions about facilitators of and barriers to eHealth usage ("What did you like/dislike on the program?"). These factors were studied to explore if there were additional factors associated with eHealth usage. Nonusage attrition (reasons to stop) was asked by another open-ended question to a subgroup of women who responded that they had dropped out during the intervention. The exact phrasing of this question was the following: "What was the reason for stopping prematurely OR never starting with 'Baas over je blaas'?"

Data Analysis

Quantitative Data

Data were analyzed using SPSS (version 25; IBM Corp). Descriptive statistics were calculated for usage parameters and characteristics of all user groups. As module number, frequency, and duration were not normally distributed, median and percentiles were calculated. Continuous variables were assessed using a 2-tailed, independent sample *t* test for 2 groups (nonusers vs users) and a 1-way ANOVA for 3 groups (user groups). A Pearson chi-square test was used for categorical variables. Statistical significance was determined at $P<.05$ (2-sided). Nominal logistic regression analysis was used to study factors associated with eHealth usage. Univariate analyses were performed, and variables with a significance level of $P<.20$ were included in the multivariate model. Variables were excluded stepwise in order of the highest *P* value until only statistically significant ($P<.05$) variables remained.

Qualitative Data

Qualitative results from 2 open-ended survey questions were analyzed through conventional content analysis [35]. First, open coding was applied to the open-ended responses. Overall, two researchers coded independently (LF and TT) and compared their codes. In case of disagreement, a third researcher (AL) gave her opinion. Codes were clustered into categories and were discussed by the research team. Data saturation was reached for all open-ended questions. Quotes are used to illustrate the findings, and they are followed by identifier number, age, and module number. The words "most," "many," "several," and "a few" indicate that >50%, 20% to 50%, 10% to 20%, or <10% of respondents, respectively, shared an opinion. Microsoft Excel 2016 was the most convenient software to code the data because data were exported from SPSS.

Ethics Approval

Ethics approval was granted by the research ethics committee of the Radboud University Medical Center, Nijmegen, the Netherlands (file number 2016-2721). The committee declared that the risks for participation in this study were negligible. The study was conducted in accordance with rules applicable in the Netherlands and the Medical Research Involving Human Subjects Act.

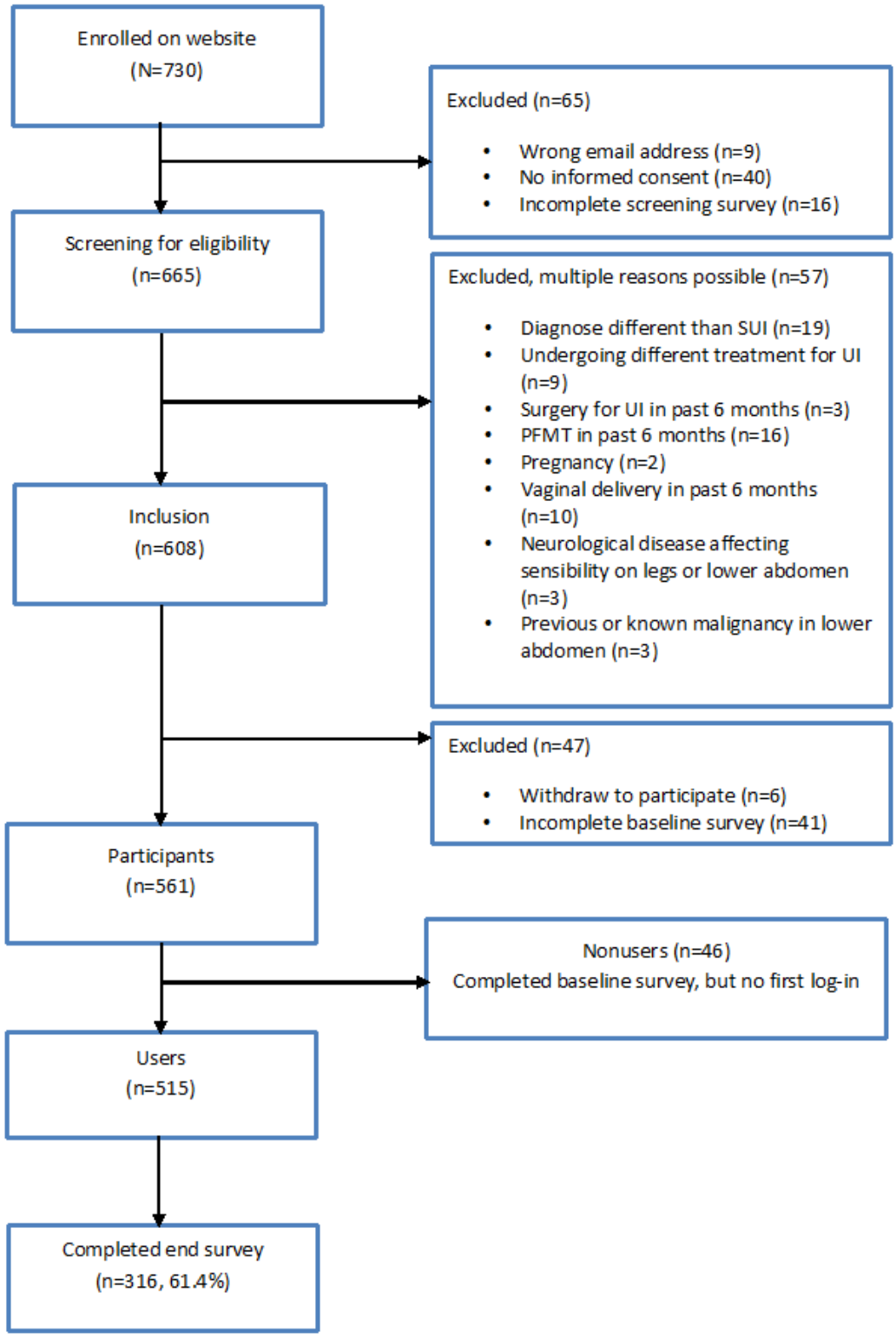
Results

Overview

A total of 730 women enrolled on the website, 608 (83.3%) of whom were included (Figure 4). Overall, 7.8% (57/730) of the women were excluded for the following main reasons: diagnosis other than SUI (mostly urgency urinary incontinence), following regular PFMT in the past 6 months, or vaginal delivery in the past 6 months. Participants (561/608, 92.3%) were women who ever logged in to the eHealth intervention, 8.2% (46/561) of whom were nonusers. Participants reached the website mostly via news items on websites or Google (244/561, 43.5%); newspapers, magazines, or newsletters (170/561, 30.3%); or social media (66/561, 11.8%).

XSL•FO
RenderX

Figure 4. Flowchart of users and nonusers of eHealth intervention. PFMT: pelvic floor muscle training; SUI: stress urinary incontinence; UI: urinary incontinence.



Baseline Characteristics

Participants (n=561) had a mean age of 50.3 (SD 12.1) years, and a minority of them had a low educational attainment level and a moderate to severe degree of incontinence (Table 1). Overall, two-thirds of them (340/553, 61.5%) had never visited a health care professional for their incontinence, and (399/557, 71.6%) had never taken previous PFMT. Most participants expected major improvement of incontinence (402/560, 71.8%) or even cure (62/560, 11.1%). There were no significant differences in demographic and incontinence-related characteristics between users (515/561, 91.8%) and nonusers (46/561, 8.2%). The burden of incontinence, duration, and incontinence aid usage between these groups (data not shown) did not differ.

XSL•FO
RenderX

Table 1. Characteristics of all participants (n=561).

Characteristics	Values
Demographics	
Age (years), mean (SD)	50.3 (12.1)
Educational attainment level, n (%)	
Low	46 (8.2)
High	515 (91.8)
Related to incontinence	
Type, n (%)	
SUI[a]	459 (81.8)
MUI[b]	102 (18.2)
Severity (ICIQ-UI SF[c]), n (%)	
Slight	42 (7.5)
Moderate	391 (69.7)
Severe	128 (22.8)
Very severe	0 (0)
Quality of life (ICIQ LUTS-QoL[d]), mean (SD)	32 (6.9)
Previous contact with health care professional, n (%)[e]	213 (38.5)
Previous PFMT[f] for incontinence, n (%)[g]	158 (28.4)
Expected ability to train pelvic floor muscles, mean (SD)	7.61 (1.5)
Expected treatment results, n (%)[h]	
Slight improvement	96 (17.1)
Major improvement	402 (71.8)
Cure	62 (11.1)

[a]SUI: stress urinary incontinence.

[b]MUI: mixed urinary incontinence.

[c]ICIQ-UI SF: International Consultation on Incontinence Questionnaire–Urinary Incontinence Short Form.

[d]ICIQ LUTS-QoL: ICIQ for Lower Urinary Tract Symptoms–Quality of Life.

[e]Missing values were removed (8/561, 1.4%); sample size, n=553.

[f]PFMT: pelvic floor muscle training.

[g]Missing values were removed (4/561, 0.7%); sample size, n=557.

[h]Missing values were removed (1/561, 0.2%); sample size, n=560.

Usage

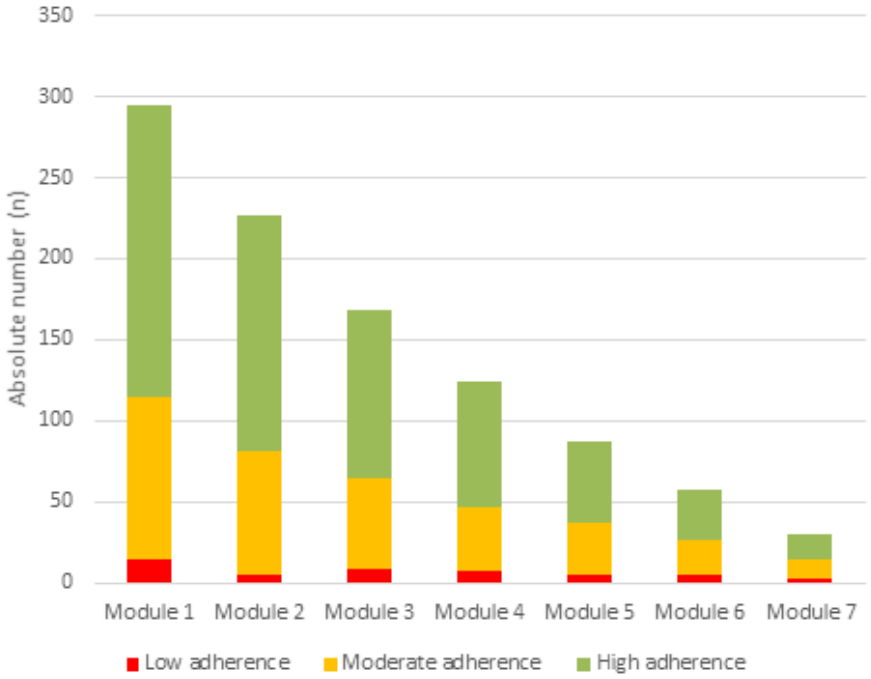
Most participants (220/561, 39.2%) dropped out in module 1, and (87/561, 15.5%) reached the intended module number ≥6. The median of the log-in frequency was 4 (range 2-9), and the median of duration was 26 days (range 4-62 days).

Adherence

For all modules, approximately 60% of participants had a high exercise adherence rate, as shown in the training reports (Figure 5).

XSL•FO
RenderX

Figure 5. Adherence to pelvic floor muscle training, as reported in the training reports for each module.



User Groups

On the basis of the intended module number and intended duration, users were divided into 3 user groups, most of whom were low users (295/515, 57.3%), followed by intermediate users (133/515, 25.8%) and high users (87/515, 16.9%). User groups differed significantly in age: the mean age for low, intermediate, and high users was 48.3 (SD 12) years, 52.1 (SD 11.6) years, and 55.3 (SD 10.9) years, respectively; *P*<.001. User groups also differed significantly in previous PFMT, with rates of 29.1% (85/292), 21.2% (28/132), and 37% (32/87) for low, intermediate, and high users, respectively; *P*=.04. The expected ability to train the pelvic floor muscles also differed significantly between user groups, with an increase in expectations from low to high users (mean 7.5, SD 1.4; mean 7.7, SD 1.4; mean 8.1, SD 1.5 for low, intermediate, and high users, respectively; *P*=.006; Table 2).

XSL•FO
RenderX

Table 2. Comparison between user groups in demographics and incontinence-related variables (n=515).

Variables	Low users (n=295, 57.3%)	Intermediate users (n=133, 25.8%)	High users (n=87, 16.9%)	Comparison between user groups (*P* value)
Demographics				
Age (years), mean (SD)	48.3 (12)	52.1 (11.6)	55.3 (10.9)	<.001
Educational attainment level, n (%)				.95
Low	27 (9.2)	11 (8.3)	8 (9.2)	
High	268 (90.8)	122 (91.7)	79 (90.8)	
Related to incontinence				
Severity (ICIQ-UI SF[a]), n (%)				.36
Slight	22 (7.5)	12 (9)	5 (5.7)	
Moderate	196 (66.4)	96 (72.2)	65 (74.7)	
Severe	77 (26.1)	25 (18.8)	17 (19.5)	
Quality of life (ICIQ LUTS-QoL[b]), mean (SD)	32.3 (7.1)	31.5 (7)	31.4 (5.8)	.38
Previous PFMT for incontinence, n (%)	85 (29.1)[c]	28 (21.2)[d]	32 (36.8)	.04
Expected ability to train pelvic floor muscles, mean (SD)	7.5 (1.4)	7.7 (1.4)	8.1 (1.5)	.006
Expected treatment results, n (%)				.09
Slight improvement	52 (17.7)[e]	24 (18)	9 (10.3)	
Major improvement	218 (74.1)[e]	93 (69.9)	63 (72.4)	
Cure	24 (8.2)[e]	16 (12)	15 (17.2)	

[a]ICIQ-UI SF: International Consultation on Incontinence Questionnaire–Urinary Incontinence Short Form.

[b]ICIQ LUTS-QoL: ICIQ for Lower Urinary Tract Symptoms–Quality of Life.

[c]Missing values were removed (3/295, 1%); sample size, n=292.

[d]Missing values were removed (1/133, 0.8%); sample size, n=132.

[e]Missing values were removed (1/295, 0.3%); sample size, n=294.

Reasons for Nonusage Attrition

Reasons for nonusage attrition were reported in the survey, which had a response rate of 61.4% (316/515). The response rate differed significantly per user group, with a completion rate of 44.1% (130/295), 81.9% (109/133), and 90% (78/87) for low, intermediate, and high users, respectively (*P*<.001). Overall, 68% (215/316) of the participants reported that they dropped out during the intervention.

There were five categories for terminating the intervention: *mismatch between everyday life and performing PFMT*, *motivational difficulties*, *problems with execution*, *guidance needed*, and *usage of eHealth*. The most common category was *mismatch between everyday life and performing PFMT*, which was mentioned by approximately half the respondents. A mismatch was mostly caused by being very busy, forgetting to practice, (new) comorbidities or illnesses, being a caretaker for a relative, or change in daily routine (such as holidays):

> *I had several other physical problems that were consuming my attention, so my impatience prevented me from focusing even more on myself.* [ID141; aged 62 years; module 1]

A second category was that women mentioned they had *motivational difficulties* and experienced loss of motivation or lack of self-discipline. Another reason for motivational loss was either that the training had no effect or that there was an effect and the decrease in symptoms made training less urgent:

> *I got good results really quickly and then it all went downhill.* [ID242; aged 52 years; module 4]

Problems with execution was a third category for nonusage attrition. Several women found the exercises hard to perform or struggled with the increasing complexity of the training:

> *...It was pretty difficulty to do the exercises standing up. It took a while before I got the hang of this exercise. I didn't feel the muscle, as if it wasn't there.* [ID10308; aged 52 years; module 5]

A few participants mentioned that the exercises gave them bodily discomfort, such as pain in the legs or back. A minor category was *guidance needed*, with reasons such as the absence of guidance through personal contact with a caregiver and wishes for more frequent reminders:

> *At first I got emails to remind me, but later they stopped coming. That also made me forget where I'd got to in the program.* [ID10099; aged 38 years; module 2]

XSL•FO
RenderX

The last category was *usage of eHealth*, in which a few women said that they had terminated the intervention owing to technical issues or because the intervention was not user-friendly.

Factors Associated With Usage

On the basis of univariate analysis, the following four variables were selected for the multivariate model: age (*P*<.001), previous PFMT (*P*=.04), expected ability to train the pelvic floor muscles (*P*=.005), and expected treatment results (*P*=.03). After multivariate analysis had been performed, 3 factors were significantly associated with usage (Table 3). The likelihood ratio test showed that age (*P*<.001), previous PFMT (*P*=.03), and expected ability to train the pelvic floor muscles (*P*=.008) significantly contributed to the model. Age was lower for low users than for intermediate and high users (odds ratio [OR] 0.97, 95% CI 0.96-0.99; *P*=.002 [not shown in Table 3] and OR 0.95, 95% CI 0.93-0.97; *P*<.001, respectively). The ORs for previous PFMT were lower for intermediate users than for high users (OR 0.45, 95% CI 0.24-0.83; *P*=.01). The expected ability to train the pelvic floor muscles was lower for low users than for high users (OR 0.75, 95% CI 0.62-0.91; *P*=.003).

Table 3. Multivariate estimates from the multinominal logistic regression model[a].

Groups	Low users vs high users		Intermediate users vs high users	
Variables	Coefficients (SE)	Odds ratio (95% CI)	Coefficients (SE)	Odds ratio (95% CI)
Age	−0.05 (0.01)[b]	0.95 (0.93-0.97)	−0.02 (0.01)	0.98 (0.96-1)
Previous PFMT[c]	−0.35 (0.27)	0.70 (0.42-1.19)	−0.80 (0.31)	0.45 (0.24-0.83)[d]
Expected ability to train the pelvic floor muscles	−0.29 (0.10)[e]	0.75 (0.62-0.91)	−0.20 (0.11)	0.82 (0.67-1.01)

[a]Pseudo R^2 (Nagelkerke)=0.093.

[b]*P*<.001.

[c]PFMT: pelvic floor muscle training.

[d]*P*=.01.

[e]*P*=.003.

Facilitators and Barriers

The response rate on the end survey was 61.4% (316/515), of which 84.8% (268/316) and 55.1% (174/316) reported facilitators and barriers, respectively. In total, four categories emerged, including codes from both facilitators and barriers: *training instructions*, *self-management*, *usage and content*, and *effects*.

Training Instructions

More than half the respondents appreciated the clear explanation of the exercises, and the stepwise setup was also highly valued. Women were able to perform the exercises and felt that the instructions guided them to find the right muscles to exercise:

> *Clear explanation of how to squeeze and, when you can't do it right away, reassurance that things will improve if you keep trying. That was correct in my case.* [ID10326; aged 71 years; module 8]

However, a few women said that the training instructions should be explained in more detail or had problems when the complexity of the exercises increased. Several women mentioned the training frequency and training duration as barriers. These women thought that the training frequency was very high and that training would be more feasible if it was lowered from 3 times to once or twice a day. The 3-month training duration was regarded as very short, and the time lock caused pressure as women who were sufficiently motivated were unable to complete all modules.

Self-management

Engaging in self-management treatment through eHealth was valued because women said that it provided them the flexibility to practice in their own place and in their own time and to practice by themselves without interference from a health care professional. A few women said that they liked being able to practice via the web because it provided privacy. In contrast, half of the respondents mentioned that they found eHealth to be very noncommittal and that they missed personal contact to obtain feedback on their performance or to stay motivated:

> *For me personally, I need more encouragement to do the exercises, working with a therapist, for instance.* [ID297; aged 49 years; module 5]

Usage and Content

Email reminders were highly valued because they made women feel guided and supported and provided them the opportunity to ask questions via email. Other content-related facilitators were the images, videos, audio fragments, download option, and possibility to write down one's personal goals. Barriers to eHealth usage were named by a few participants: the lack of overview of the website and the lack of overview of all exercises. Others had problems in navigating through the website, reading the website on their mobile phones, or logging in repeatedly into their computers. An app would be more accessible, according to several women. Other barriers were the presence of a lot of text or redundancy and the need for more visual support.

Effects

Women were encouraged to continue when they noticed a positive effect on their incontinence symptoms; their ability to contract the pelvic floor muscles; or other pelvic floor symptoms, such as prolapse or urge incontinence. They also said that they had gained knowledge, become more aware of

XSL•FO
RenderX

their pelvic floor, and gained confidence. Having no improvement was reported as a barrier by a few participants:

> *I gained self-confidence. It's very nice to feel that I'm partly back in control.* [ID120; aged 61 years; module 8]

Overall, women appreciated that such a program was available, urinary incontinence was getting attention, and the problem was normalized; however, a woman said that using a website for this problem had the effect of keeping it a taboo.

Discussion

Principal Findings

This study shows that eHealth for SUI was mainly used by women who had never visited a health care professional for PFMT or never performed PFMT before. Although adherence to the exercises was high for all modules, most participants (295/561, 52.6%) were low users. User groups differed in age and their expected ability to train the pelvic floor muscles. Reasons for nonusage attrition were problems with scheduling and prioritizing PFMT in everyday life and with execution of the exercises. Factors that were associated with high eHealth usage were high age, previous PFMT, and high expectation of being able to train the pelvic floor muscles. Further facilitating factors for eHealth usage were clear explanation, stepwise setup, guidance by email reminders, and self-management opportunities. In contrast, its noncommittal character, absence of personal contact, and high training frequency were hindering factors.

Comparison With Previous Studies

This study indicates that eHealth fulfills a need for women who would not turn elsewhere to deal with this problem, as reflected by two-thirds of the participants (340/553, 61.5%) who had never contacted a health care professional before. Reasons for not seeking help for this problem are well studied and often related to shame and to not recognizing SUI as a treatable problem [10-12]. eHealth fills this gap by supplying a self-management tool that increases access to care. It is known that some women prefer eHealth because it allows them to take a first step before seeking help and that most eHealth participants have not had treatment before they start [13,21].

In accordance with eHealth for other conditions, this study reports a high nonusage attrition rate, with an initial rapid decline and a remaining group of steady users [24,36]. Some women in this study reported that they stopped using the intervention because the training has an early positive effect on their symptoms. Other reasons for nonusage attrition in this study have also been reported as barriers by other studies evaluating the usage of eHealth for urinary incontinence [18-21]. These barriers are adherence challenges and problems in integrating PFMT into everyday life. The current version of the intervention addresses this by providing suggestions on how to fit the exercises into daily life, such as setting an alarm or incorporating the exercises into daily routines. Several women said that a mobile app would be more accessible and push messages on a telephone could increase usage and adherence. Another reason for abandoning the program consisted of problems with executing and insecurity about correct training, which matches our findings suggesting that previous experience with PFMT facilitates eHealth usage. As most participants (399/557, 71.6%) did not perform PFMT before, they were unable to depend on previous experiences, which may have contributed to the high nonusage attrition figure. These women may profit from more intensive (digital) contact with a health care professional throughout the eHealth program. This contact can be achieved either through digital or physical consultations on several, time fixed moments. Finally, it could be that some women start out of curiosity and lose their interest early [17]. Future studies could investigate log-in data on the views per webpage to provide insight into attractive eHealth ingredients, which could lead to further improvements.

A remarkable finding is that high age (>50 years) is associated with high usage of eHealth for urinary incontinence. This shows that women with SUI who are aged >50 years are better candidates for using an eHealth program. This is not consistent with previous findings, in which clinicians were concerned about the suitability of eHealth for older people owing to their low access to technology [37,38]. GPs thought that eHealth was more beneficial for young women with SUI because the burdens of time-consuming jobs and childcare would prevent them from visiting a health care professional [38]. Nevertheless, a recent study confirmed our findings by showing that women who were recruited through social and conventional media, for participation in an app-based intervention for urinary incontinence, were old compared with those who were recruited through their GP [39]. This indicates that concerns by GPs are incorrect because eHealth is suitable for older women also and that attention is needed before clinicians exclude women from participation. We hypothesize that women aged >50 years have few conflicts in everyday life that prevent them from continuing treatment. Another explanation could be that *older women* prefer eHealth treatment to be delivered via a website instead of a mobile app, which was also found by others [40,41]. This eHealth intervention was not mobile friendly, which was also mentioned by several women who said that it was hard to view the website on their mobile phone. If readability on a mobile phone is improved, young women may be encouraged to continue usage.

Finally, this study underlines the importance of paying attention to people's expectations about their ability to execute the exercises. Having high expectations is associated with high eHealth usage. Although the question about the expected ability to perform PFMT is not a validated question for assessing self-efficacy, it approximates it. Self-efficacy is defined as "people's beliefs about their capabilities to produce effects" [42]. It is known that self-efficacy expectations and attitudes to exercise are determinants of adherence to PFMT [26]. Our previous study already showed that women need a certain degree of self-efficacy to adopt the eHealth intervention for SUI [21]. In this study, one can argue whether the assessed difference in expectations about PFMT between low and high users is clinically relevant. However, studies on eHealth for urinary incontinence confirm our finding by showing that high expectations about treatment and self-rated ability to perform PFMT are determinants for treatment success [31,43]. Therefore,

XSL•FO
RenderX

it is important for women to have a sense of confidence before the start of the eHealth intervention. This can be achieved either by health care professionals or by improvements in the eHealth intervention itself. Currently, the intervention starts with plain information about the effectiveness of the intervention and about goal setting options, and encouraging phrases in email reminders are used to stimulate adherence. Improving self-efficacy can be effectively achieved by including positive suggestions about eHealth before and during participation [44]. When GPs refer to eHealth, they need to enhance the women's confidence in performing PFMT with web-based training and emphasize their ability as part of motivational interviewing [26].

Strengths and Limitations

A major strength of this study is that it focuses on eHealth intervention usage for urinary incontinence in a real-world setting rather than in a trial setting. It is known that adherence and nonusage attrition differ for users of open access websites versus users of websites in an RCT [36]. Participants in this study had to register and complete at least the baseline survey but did not have to follow a strict research protocol, which simulated real-life usage more accurately. Another strength is that we categorized user groups based on 2 log parameters and used the term *intended use*, both highly recommended [33,45]. Using a mixed methods design enabled us to seek explanations for findings from quantitative analyses from the qualitative data, such as reasons for nonusage attrition.

A limitation of this study is that its generalizability may be restricted owing to the low proportion of participants with low education in contrast to the general Dutch population (9% vs 29%) [46]. Previous studies on eHealth for urinary incontinence showed that most participants were highly educated [13,15,16], possibly because eHealth users are generally more literate [41]. Another limitation is that adherence to PFMT exercises was assessed using self-reported training reports, which may have affected data validity because participants had to complete them before they gained access to the next module. However, the adherence to PFMT for every module was approximately 60%, and this is consistent with adherence rates to regular PFMT, which is estimated to be 64% [26]. Finally, the log-in data could have been affected by 2 aspects. First, some women may have trained offline because it was possible to download the training. Second, for research purposes, access to the intervention was restricted to 3 months. If no download option were included or more time were provided, numbers in the intermediate or high user groups may have increased.

Conclusions

This study shows that eHealth fulfills a need for women with SUI who have never received treatment before. Although adherence to PFMT was high for every module, most participants stopped prematurely because it was difficult for them to integrate training into their everyday lives. High usage is more likely among women aged >50 years, those who received previous PFMT, and those with high expected ability to train the pelvic floor muscles. Knowledge about these user characteristics can guide clinicians and correct possible misunderstandings about the suitable target population for this intervention. Furthermore, strategies for reinforcing expectations and self-efficacy are important to upscale eHealth usage. Paying attention to people's need for personal contact is also important; including digital methods for communicating with a health care professional or implementing eHealth into primary care (*blended care*) can enhance personal contact.

Acknowledgments

The authors thank the eContinence group from the Umeå University for their permission to use their eHealth intervention for research purposes in the Netherlands. The authors are grateful to the women who invested their time to participate in this study.

Conflicts of Interest

The logo Tät is registered as trademark by EQUIPO for eContinence AB, a Swedish e-health company founded in July 2021. The authors of this paper declare no conflicts of interest, and they have no involvement in the Swedish eContinence group.

References

1. Abrams P, Andersson KE, Birder L, Brubaker L, Cardozo L, Chapple C, Members of Committees, Fourth International Consultation on Incontinence. Fourth International Consultation on Incontinence Recommendations of the International Scientific Committee: evaluation and treatment of urinary incontinence, pelvic organ prolapse, and fecal incontinence. Neurourol Urodyn 2010;29(1):213-240. [doi: 10.1002/nau.20870] [Medline: 20025020]
2. Bartoli S, Aguzzi G, Tarricone R. Impact on quality of life of urinary incontinence and overactive bladder: a systematic literature review. Urology 2010 Mar;75(3):491-500. [doi: 10.1016/j.urology.2009.07.1325] [Medline: 19962738]
3. Schreiber Pedersen L, Lose G, Høybye MT, Elsner S, Waldmann A, Rudnicki M. Prevalence of urinary incontinence among women and analysis of potential risk factors in Germany and Denmark. Acta Obstet Gynecol Scand 2017 Aug;96(8):939-948 [FREE Full text] [doi: 10.1111/aogs.13149] [Medline: 28401541]
4. Linde JM, Nijman RJ, Trzpis M, Broens PM. Urinary incontinence in the Netherlands: prevalence and associated risk factors in adults. Neurourol Urodyn 2017 Aug;36(6):1519-1528. [doi: 10.1002/nau.23121] [Medline: 27701779]
5. Gonzalez G, Vaculik K, Khalil C, Zektser Y, Arnold C, Almario CV, et al. Women's experience with stress urinary incontinence: insights from social media analytics. J Urol 2020 May;203(5):962-968 [FREE Full text] [doi: 10.1097/JU.0000000000000706] [Medline: 31855097]

XSL•FO
RenderX

6. Monz B, Pons ME, Hampel C, Hunskaar S, Quail D, Samsioe G, et al. Patient-reported impact of urinary incontinence--results from treatment seeking women in 14 European countries. Maturitas 2005 Nov 30;52 Suppl 2:S24-S34. [doi: 10.1016/j.maturitas.2005.09.005] [Medline: 16297579]
7. Teunissen D, Dekker JH, Lagro-Janssen LA, Berghmans LC, Uijen JH, Mientjes GH, et al. Practice guideline 'Urinary incontinence in women' from the Dutch College of General Practitioners. Nederlands Huisartsen Genootschap. URL: https://richtlijnen.nhg.org/standaarden/incontinentie-voor-urine-bij-vrouwen#volledige-tekst [accessed 2022-04-01]
8. Dumoulin C, Adewuyi T, Booth J, Bradley C, Burgio K, Hagen S, et al. Adult conservative management. In: Abrams P, Cardozo L, Wagg A, Wein A, editors. Incontinence: 6th International Consultation on Incontinence. Bristol, UK: International Continence Society; 2016:1443-1628.
9. Nguyen K, Hunter KF, Wagg A. Knowledge and understanding of urinary incontinence: survey of family practitioners in northern Alberta. Can Fam Physician 2013 Jul;59(7):e330-e337 [FREE Full text] [Medline: 23851561]
10. Shaw C, Das Gupta R, Williams KS, Assassa RP, McGrother C. A survey of help-seeking and treatment provision in women with stress urinary incontinence. BJU Int 2006 Apr;97(4):752-757. [doi: 10.1111/j.1464-410X.2006.06071.x] [Medline: 16536767]
11. Schreiber Pedersen L, Lose G, Høybye MT, Jürgensen M, Waldmann A, Rudnicki M. Predictors and reasons for help-seeking behavior among women with urinary incontinence. Int Urogynecol J 2018 Apr;29(4):521-530. [doi: 10.1007/s00192-017-3434-0] [Medline: 28780650]
12. Lamerton TJ, Mielke GI, Brown WJ. Urinary incontinence in young women: risk factors, management strategies, help-seeking behavior, and perceptions about bladder control. Neurourol Urodyn 2020 Nov;39(8):2284-2292. [doi: 10.1002/nau.24483] [Medline: 32805080]
13. Loohuis AM, Wessels NJ, Dekker JH, van Merode NA, Slieker-Ten Hove MC, Kollen BJ, et al. App-based treatment in primary care for urinary incontinence: a pragmatic, randomized controlled trial. Ann Fam Med 2021;19(2):102-109 [FREE Full text] [doi: 10.1370/afm.2585] [Medline: 33685871]
14. Wu C, Newman DK, Palmer MH. Unsupervised behavioral and pelvic floor muscle training programs for storage lower urinary tract symptoms in women: a systematic review. Int Urogynecol J 2020 Dec;31(12):2485-2497. [doi: 10.1007/s00192-020-04498-9] [Medline: 32870339]
15. Asklund I, Nyström E, Sjöström M, Umefjord G, Stenlund H, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2017 Jun;36(5):1369-1376. [doi: 10.1002/nau.23116] [Medline: 27611958]
16. Sjöström M, Umefjord G, Stenlund H, Carlbring P, Andersson G, Samuelsson E. Internet-based treatment of stress urinary incontinence: a randomised controlled study with focus on pelvic floor muscle training. BJU Int 2013 Aug;112(3):362-372 [FREE Full text] [doi: 10.1111/j.1464-410X.2012.11713.x] [Medline: 23350826]
17. Rygh P, Asklund I, Samuelsson E. Real-world effectiveness of app-based treatment for urinary incontinence: a cohort study. BMJ Open 2021 Jan 04;11(1):e040819 [FREE Full text] [doi: 10.1136/bmjopen-2020-040819] [Medline: 33397664]
18. Asklund I, Samuelsson E, Hamberg K, Umefjord G, Sjöström M. User experience of an app-based treatment for stress urinary incontinence: qualitative interview study. J Med Internet Res 2019 Mar 14;21(3):e11296 [FREE Full text] [doi: 10.2196/11296] [Medline: 30869644]
19. Björk AB, Sjöström M, Johansson EE, Samuelsson E, Umefjord G. Women's experiences of Internet-based or postal treatment for stress urinary incontinence. Qual Health Res 2014 Apr;24(4):484-493 [FREE Full text] [doi: 10.1177/1049732314524486] [Medline: 24598777]
20. Wessels NJ, Hulshof L, Loohuis AM, van Gemert-Pijnen L, Jellema P, van der Worp H, et al. User experiences and preferences regarding an app for the treatment of urinary incontinence in adult women: qualitative study. JMIR Mhealth Uhealth 2020 Jun 12;8(6):e17114 [FREE Full text] [doi: 10.2196/17114] [Medline: 32530431]
21. Firet L, Teunissen TA, Kool RB, van Doorn L, Aourag M, Lagro-Janssen AL, et al. Women's adoption of a web-based intervention for stress urinary incontinence: a qualitative study. BMC Health Serv Res 2021 Jun 12;21(1):574 [FREE Full text] [doi: 10.1186/s12913-021-06585-z] [Medline: 34118900]
22. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
23. Verhoeks C, Teunissen D, van der Stelt-Steenbergen A, Lagro-Janssen A. Women's expectations and experiences regarding e-health treatment: a systematic review. Health Informatics J 2019 Sep;25(3):771-787 [FREE Full text] [doi: 10.1177/1460458217720394] [Medline: 28764600]
24. 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]
25. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
26. Dumoulin C, Hay-Smith J, Frawley H, McClurg D, Alewijnse D, Bo K, International Continence Society. 2014 consensus statement on improving pelvic floor muscle training adherence: International Continence Society 2011 State-of-the-Science Seminar. Neurourol Urodyn 2015 Sep;34(7):600-605. [doi: 10.1002/nau.22796] [Medline: 25998603]

XSL•FO
RenderX

27. Firet L, Teunissen TA, van der Vaart H, Assendelft WJ, Notten KJ, Kool RB, et al. An electronic health intervention for Dutch women with stress urinary incontinence: protocol for a mixed methods study. JMIR Res Protoc 2019 Jul 11;8(7):e13164 [FREE Full text] [doi: 10.2196/13164] [Medline: 31298222]
28. 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]
29. Farrell SA, Bent A, Amir-Khalkhali B, Rittenberg D, Zilbert A, Farrell KD, et al. Women's ability to assess their urinary incontinence type using the QUID as an educational tool. Int Urogynecol J 2013 May;24(5):759-762. [doi: 10.1007/s00192-012-1925-6] [Medline: 22940842]
30. Dutch translation of Tät® - Treatment of Stress Urinary Incontinence. Baas over je blaas. 2018. URL: https://baasoverjeblaas.nl/ [accessed 2018-12-30] [WebCite Cache ID 73Hpmu8V2]
31. Lindh A, Sjöström M, Stenlund H, Samuelsson E. Non-face-to-face treatment of stress urinary incontinence: predictors of success after 1 year. Int Urogynecol J 2016 Dec;27(12):1857-1865 [FREE Full text] [doi: 10.1007/s00192-016-3050-4] [Medline: 27260323]
32. Tät.nu. URL: https://xn--tt-via.nu/?page_id=84&lang=en [accessed 2022-11-04]
33. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
34. Hung HC, Chih SY, Lin HH, Tsauo JY. Exercise adherence to pelvic floor muscle strengthening is not a significant predictor of symptom reduction for women with urinary incontinence. Arch Phys Med Rehabil 2012 Oct;93(10):1795-1800. [doi: 10.1016/j.apmr.2012.03.010] [Medline: 22446154]
35. 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]
36. Wanner M, Martin-Diener E, Bauer G, Braun-Fahrländer C, Martin BW. Comparison of trial participants and open access users of a Web-based physical activity intervention regarding adherence, attrition, and repeated participation. J Med Internet Res 2010 Feb 10;12(1):e3 [FREE Full text] [doi: 10.2196/jmir.1361] [Medline: 20147006]
37. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
38. Firet L, de Bree C, Verhoeks CM, Teunissen DA, Lagro-Janssen AL. Mixed feelings: general practitioners' attitudes towards eHealth for stress urinary incontinence - a qualitative study. BMC Fam Pract 2019 Jan 26;20(1):21 [FREE Full text] [doi: 10.1186/s12875-019-0907-x] [Medline: 30684962]
39. van der Worp H, Loohuis AM, Flohil IL, Kollen BJ, Wessels NJ, Blanker MH. Recruitment through media and general practitioners resulted in comparable samples in an RCT on incontinence. J Clin Epidemiol 2020 Mar;119:85-91. [doi: 10.1016/j.jclinepi.2019.12.001] [Medline: 31811893]
40. Morrison LG, Geraghty AW, Lloyd S, Goodman N, Michaelides DT, Hargood C, et al. Comparing usage of a Web and app stress management intervention: an observational study. Internet Interv 2018 Jun;12:74-82 [FREE Full text] [doi: 10.1016/j.invent.2018.03.006] [Medline: 30135771]
41. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
42. Bandura A. Self-efficacy. In: Ramachaudran VS, editor. Encyclopedia of Human Behavior. New York, NY, USA: Academic Press; 1994:71-81.
43. Nyström E, Asklund I, Sjöström M, Stenlund H, Samuelsson E. Treatment of stress urinary incontinence with a mobile app: factors associated with success. Int Urogynecol J 2018 Sep;29(9):1325-1333 [FREE Full text] [doi: 10.1007/s00192-017-3514-1] [Medline: 29222718]
44. Skvortsova A, Cohen Rodrigues T, de Buisonjé D, Kowatsch T, Santhanam P, Veldhuijzen DS, et al. Increasing the effectiveness of a physical activity smartphone intervention with positive suggestions: randomized controlled trial. J Med Internet Res 2022 Mar 01;24(3):e32130 [FREE Full text] [doi: 10.2196/32130] [Medline: 35230245]
45. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
46. Trends in Nederland. Centraal Bureau voor de Statistiek (CBS). 2018. URL: https://longreads.cbs.nl/trends18/maatschappij/cijfers/onderwijs/ [accessed 2022-02-01]

Abbreviations

CONSORT-eHealth: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

XSL•FO
RenderX

GP: general practitioner
ICIQ LUTS-QoL: International Consultation on Incontinence Questionnaire for Lower Urinary Tract Symptoms–Quality of Life
ICIQ-UI SF: International Consultation on Incontinence Questionnaire–Urinary Incontinence Short Form
OR: odds ratio
PFMT: pelvic floor muscle training
RCT: randomized controlled trial
SUI: stress urinary incontinence

Edited by R Kukafka; submitted 04.04.22; peer-reviewed by M Blanker, L Morrison; comments to author 11.05.22; revised version received 07.06.22; accepted 26.09.22; published 17.11.22.

Please cite as:
Firet L, Teunissen TAM, Kool RB, Notten KJB, Lagro-Janssen ALM, van der Vaart H, Assendelft WJJ
Usage of a Web-Based eHealth Intervention for Women With Stress Urinary Incontinence: Mixed Methods Study
J Med Internet Res 2022;24(11):e38255
URL: https://www.jmir.org/2022/11/e38255
doi:10.2196/38255
PMID:36394923

©Lotte Firet, Theodora Alberta Maria Teunissen, Rudolf Bertijn Kool, Kim Josephina Bernadette Notten, Antoinette Leonarda Maria Lagro-Janssen, Huub van der Vaart, Willem Jan Jozef Assendelft. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Estimating the Economic Value of Automated Virtual Reality Cognitive Therapy for Treating Agoraphobic Avoidance in Patients With Psychosis: Findings From the gameChange Randomized Controlled Clinical Trial

James Altunkaya[1], MSc; Michael Craven[2,3,4], PhD; Sinéad Lambe[5,6], DClinPsy; Ariane Beckley[5], MSc; Laina Rosebrock[5,6], PhD; Robert Dudley[7], PhD; Kate Chapman[8], DClinPsy; Anthony Morrison[9,10], DClinPsy; Eileen O'Regan[11], MB, MRCPsych; Jenna Grabey[12], MSc; Aislinn Bergin[2,4,13], PhD; Thomas Kabir[14], PhD; Felicity Waite[5,6], DClinPsy; Daniel Freeman[5,6], PhD; José Leal[1], DPhil

[1]Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

[2]National Institute for Health and Care Research MindTech Med-Tech Co-operative, Nottingham, United Kingdom

[3]Human Factors Research Group, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom

[4]Mental Health & Technology Theme, National Institute for Health and Care Research Nottingham Biomedical Research Centre, Nottingham, United Kingdom

[5]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[6]Oxford Health NHS Foundation Trust, Oxford, United Kingdom

[7]Cumbria, Northumberland, Tyne, and Wear NHS Foundation Trust, Newcastle upon Tyne, United Kingdom

[8]Avon and Wiltshire Mental Health Partnership (AWP) NHS Trust, Bath, United Kingdom

[9]Greater Manchester Mental Health Foundation Trust, Manchester, United Kingdom

[10]Division of Psychology and Mental Health, University of Manchester, Manchester, United Kingdom

[11]Nottinghamshire Healthcare NHS Foundation Trust, Nottingham, United Kingdom

[12]Oxford Primary Care Clinical Trials Unit, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[13]Mental Health & Clinical Neurosciences, School of Medicine, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[14]McPin Foundation, London, United Kingdom

Corresponding Author:
José Leal, DPhil
Health Economics Research Centre
Nuffield Department of Population Health
University of Oxford
Richard Doll Building
Old Road Campus
Oxford, OX3 7LF
United Kingdom
Phone: 44 1865 289263
Email: jose.leal@dph.ox.ac.uk

Abstract

Background: An automated virtual reality cognitive therapy (gameChange) has demonstrated its effectiveness to treat agoraphobia in patients with psychosis, especially for high or severe anxious avoidance. Its economic value to the health care system is not yet established.

Objective: In this study, we aimed to estimate the potential economic value of gameChange for the UK National Health Service (NHS) and establish the maximum cost-effective price per patient.

Methods: Using data from a randomized controlled trial with 346 patients with psychosis (ISRCTN17308399), we estimated differences in health-related quality of life, health and social care costs, and wider societal costs for patients receiving virtual reality therapy in addition to treatment as usual compared with treatment as usual alone. The maximum cost-effective prices of

XSL•FO
RenderX

gameChange were calculated based on UK cost-effectiveness thresholds. The sensitivity of the results to analytical assumptions was tested.

Results: Patients allocated to gameChange reported higher quality-adjusted life years (0.008 QALYs, 95% CI –0.010 to 0.026) and lower NHS and social care costs (–£105, 95% CI –£1135 to £924) compared with treatment as usual (£1=US $1.28); however, these differences were not statistically significant. gameChange was estimated to be worth up to £341 per patient from an NHS and social care (NHS and personal social services) perspective or £1967 per patient from a wider societal perspective. In patients with high or severe anxious avoidance, maximum cost-effective prices rose to £877 and £3073 per patient from an NHS and personal social services perspective and societal perspective, respectively.

Conclusions: gameChange is a promising, cost-effective intervention for the UK NHS and is particularly valuable for patients with high or severe anxious avoidance. This presents an opportunity to expand cost-effective psychological treatment coverage for a population with significant health needs.

Trial Registration: ISRCTN Registry ISRCTN17308399; https://www.isrctn.com/ISRCTN17308399

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-031606

(J Med Internet Res 2022;24(11):e39248) doi:10.2196/39248

KEYWORDS

gameChange; virtual reality; National Health Service; NHS; cost-effectiveness; economic evaluation; maximum price

Introduction

Background

There are recognized challenges for health services providing evidence-based psychological therapies such as cognitive behavioral therapy (CBT) to patients diagnosed with psychosis [1]. There is a shortage of trained therapists in these approaches, and there are also issues of adherence and competence in therapists' delivery of current evidence-based approaches [2,3]. The latest evidence indicates that psychosis affects approximately 0.7% of the people in the United Kingdom [4].

An automated virtual reality (VR) therapy (gameChange) was therefore recently trialed to help patients diagnosed with psychosis to re-engage with everyday situations avoided because of anxiety [5,6]. A digital coach guides the patient through the 6-session program. The automated therapy is supported by peer support workers, assistant psychologists, or clinical psychologists, enabling a much wider workforce to deliver the intervention than CBT therapists alone. The gameChange intervention led to significant reductions in anxious avoidance of, and distress in, everyday situations compared with treatment as usual (TAU) alone [6]. The largest treatment benefits were seen for patients with high or severe agoraphobic avoidance, with a corresponding clinical recommendation that these patients are prioritized within any future implementation of gameChange [6].

Nevertheless, for an intervention to be implemented in resource-constrained health systems such as the UK National Health Service (NHS), evidence of clinical benefit must be supplemented by evidence supporting a new intervention's cost-effectiveness. This requires an intervention to be assessed using a generalizable set of economic methods that can compare the relative value of interventions across different clinical areas. This determines whether a new intervention is a cost-effective investment for the health system compared with all other potential uses of its resources.

In the United Kingdom, before a new intervention's implementation in the NHS, the intervention's cost-effectiveness is assessed by the National Institute for Health and Care Excellence (NICE) for England and Northern Ireland, the Scottish Medicines Consortium, and Health Technology Wales. Interventions are typically considered to be cost-effective (and therefore eligible for implementation) if they are shown to improve length of life and health-related quality of life at a cost lower than £20,000 to £30,000 (£1=US $1.28) per quality-adjusted life year (QALY) gained compared with usual care [7,8].

Although several mobile, internet, and VR CBT approaches have been previously developed and shown to be effective to varying degrees [9,10], the cost-effectiveness of only a handful of these therapies has been tested [11-15]. The economic value of only 1 VR CBT intervention has previously been tested in a population with psychosis [13], a nonautomated therapy delivered by psychologists with CBT training. By contrast, gameChange has the potential to reduce demand on clinical psychologists' time, with its automated delivery by nonspecialist staff likely to reduce pressure on treatment waiting lists and increase rates of psychological treatment.

Objectives

Our study examined the potential economic value of providing gameChange to patients with psychosis in the UK NHS. We used randomized controlled trial data to estimate the difference in health-related quality of life and use of NHS and wider societal resources for patients who received gameChange in addition to TAU compared with TAU alone. On the basis of the QALY thresholds of £20,000 to £30,000, we estimated a maximum cost-effective price for the gameChange intervention, especially for (largely housebound) patients with substantial agoraphobic avoidance who had been shown to experience the greatest clinical benefits [6].

XSL•FO
RenderX

Methods

Estimating Economic Value

We used UK national cost-effectiveness thresholds of £20,000 to £30,000 per QALY [8] to estimate the maximum cost-effective price for the gameChange intervention per patient treated for the UK health system. We compared differences in QALYs, health and social care costs, and wider societal costs between patients in the treatment and control arms of the gameChange trial to estimate the maximum price the UK health care system would be willing to pay for the gameChange intervention [16,17]. As gameChange is a new intervention, its pricing structure is currently unknown. This paper therefore presents a range of scenarios where gameChange may be implemented and the maximum cost-effective price that could be paid for an individual's treatment with gameChange, given the health benefits observed in the trial. We present maximum cost-effective prices of the intervention at the lower bound and upper bound of the UK agencies' willingness to pay for health at £20,000 per QALY and £30,000 per QALY, respectively.

Data and Analysis

We estimate the costs and health outcomes of patients randomized to receive gameChange plus TAU compared with TAU alone. The gameChange trial recruited patients of NHS services with self-reported difficulties going outside because of anxiety who also had a clinical diagnosis of schizophrenia spectrum disorder or an affective diagnosis with psychotic symptoms.

Between July 25, 2019, and May 7, 2021, a total of 346 trial participants with a mean age of 37.2 (SD 12.5) years across 5 trial sites were randomized to receive the gameChange intervention (n=174, 50.3%) or TAU (n=172, 49.7%). Of the 346 patients, 232 (67.1%) were male and were recruited from 3 types of psychiatric services: early intervention (n=133, 38.4%), community mental health (n=209, 60.4%), and inpatient services (n=4, 1.2%). TAU typically comprised a prescription of antipsychotic medications, regular visits from a community mental health worker, and occasional psychiatric outpatient appointments. The gameChange intervention consisted of an automated VR cognitive therapy delivered in approximately 6 sessions of 30 minutes each over 6 weeks. The therapy aims for participants to relearn safety by undertaking repeated behavioral experiments in one of six VR social situations: a café, a general practitioner (GP) waiting room, a pub, a bus, the front door of a home, and a small local shop. Participants selected virtual tasks to complete of a graded level of social difficulty, with participants progressing through tasks with different levels of difficulty during their therapy. Full demographic information of trial participants, details of the intervention, and the outcomes collected are reported in depth elsewhere [5,6].

Health Economic Data Collection

Resource use data were collected for each participant in the trial using questionnaires at baseline, at 6 weeks (end of VR therapy for those allocated), and at 6 months after randomization.

Costs

Health and Social Service Contacts

Participants recorded the frequency of their use of health and social care services using a client service receipt inventory self-report questionnaire adapted from previous psychiatric research at their baseline, 6-week, and 6-month interviews with trial research assistants [18]. Recorded service use consisted of GP contacts; contacts with psychiatrists, therapists, or community mental health teams; hospitalizations, including accident and emergency department visits or outpatient appointments; and use of paid help from NHS or social care services. Resource use was multiplied by unit costs (Table S1 in Multimedia Appendix 1) to estimate costs at each follow-up period. Hospital admissions were converted into an appropriate health care resource group, conditional on reason for admission, and valued using 2019-20 NHS reference costs [19]. Any length of stay beyond the health care resource group trim point was costed as excess bed days reported in 2017-18 NHS reference costs [20], inflated to the 2019 price level using the NHS pay-and-price index [21]. Admissions to mental health inpatient wards were documented by the research team from medical records and costed using NHS reference costs (Table S1 in Multimedia Appendix 1).

Medications and Therapies

Information on participants' psychotropic medications and therapies was obtained by the trial team from medical record data checks at each follow-up period. The cost of all psychotropic medication prescribed during follow-up was obtained by matching reported drugs and their dose to 2019 British National Formulary prices [22]. Recorded mental health therapies were costed using a national database (Table S1 in Multimedia Appendix 1) [21].

Criminal Justice Services and Informal Care

Participant questionnaires captured contacts with criminal justice services (police contacts, nights spent in a prison cell or prison, psychiatric assessments received while in custody, and criminal or civil court appearances) and unpaid care received from family or friends (employment status of carers, how often they received help, and hours of care received). Table S2 in Multimedia Appendix 1 reports the unit costs used to value contacts with informal care and criminal justice services.

Health-Related Quality of Life

Participants completed 2 health-related quality-of-life questionnaires—EQ-5D-5L and Recovering Quality of Life, 20-item version (ReQoL-20)—at baseline, 6 weeks, and 6 months. The EQ-5D-5L determines the self-reported health status of each individual across 5 domains: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. Respondents were asked to choose one of five possible levels for each domain that reflected their *own health state today*, representing (1) *no problems*, (2) *slight problems*, (3) *moderate problems*, (4) *severe problems*, or (5) *extreme problems*. Further information on the EQ-5D-5L and its test-retest reliability is provided elsewhere [23,24]. The descriptions of respondents' health states were converted into EQ-5D utility scores [23] using the Van Hout UK crosswalk approach to the EQ-5D-3L [25].

XSL•FO
RenderX

An alternative EQ-5D-3L crosswalk [26] was investigated in sensitivity analysis. The utility scores are truncated at 1 (full health), with 0 representing death and negative values representing states worse than death (where a person would prefer to be dead than experience a given health state).

The ReQoL-20 questionnaire consists of 20 mental health questions and 1 physical health question aimed at capturing health status for mental health conditions [27]. Further information on the ReQoL-20 questionnaire and its test-retest reliability is provided elsewhere [28]. The responses to the ReQoL-20 questionnaire were converted into ReQoL-20 utility scores [29], with the same interpretation as the EQ-5D scores: *1* represents perfect health, *0* represents death, and negative values represent states worse than death.

QALYs were calculated from health state utilities using the area under the curve approach, with utilities averaged among the time points from the baseline, 6-week, and 6-month questionnaires [30]. We produced 2 sets of QALYs: one informed by EQ-5D utilities (our base case) and a second one informed by ReQoL-20 utility scores.

Missing Data

We followed best practice methods for addressing missing data in cost-effectiveness studies [31]. Missing data on participant characteristics at baseline were imputed using unconditional mean imputation, and we used multiple imputation by chained equations to impute missing data on EQ-5D scores, ReQoL-20 scores, and cost components at each follow-up time point. Full details of the missing data strategy are presented in Multimedia Appendix 1 (refer to the Missing Data and Complete Case Analysis section).

Analysis

Overview

Our analysis took a 6-month time horizon in line with the duration of the trial, meaning that health outcomes and costs were calculated for a duration of 6 months. Costs and health outcomes were therefore not discounted, given the short time horizon of the trial. The analysis follows intention-to-treat principles and takes 2 perspectives: first, an NHS and personal social services (PSS) perspective and, second, a societal perspective incorporating NHS and PSS costs and wider costs of criminal justice contacts, informal caregiving, and private health care expenditure. NICE's preferred base case for estimating cost-effectiveness requires an NHS and PSS perspective and uses QALYs calculated using the EQ-5D measure [8]. Costs were analyzed at a 2019 price level, representing the price level at the start of the trial. All analyses were conducted using Stata (version 17.0; StataCorp LLC).

After multiple imputation, differences in costs between the treatment and control arms at 6 weeks and 6 months were calculated using multilevel mixed effects models, adjusted for a patients' recruitment site and psychiatric service at randomization. The model included a time×treatment interaction, where the follow-up time point was indicated as a categorical variable. Cost differences between the arms over the whole trial period were calculated from linear regression models, including adjustment for patients' recruitment site and psychiatric service at randomization. Differences in QALYs between the arms were calculated using the same methods, while additionally adjusting for patients' reported baseline utility scores. Estimates derived from each imputed data set were combined using Rubin's rules [32].

Calculation of Maximum Cost-effective Price

We estimated the joint uncertainty around incremental total costs and QALYs (ie, the difference between gameChange plus TAU and TAU) by bootstrapping 1000 times from each of the 22 imputed data sets (creating at least 22,000 bootstraps), running the estimation model on each bootstrapped data set, and extracting the estimated treatment effects. The maximum cost-effective price for the gameChange intervention was estimated using the net monetary benefit framework at cost-effectiveness thresholds of £20,000 per QALY and £30,000 per QALY. The net monetary benefit is the product of the mean difference in QALYs and the threshold value (representing the monetary value placed on QALY health gain by the UK health system) minus the mean difference in costs.

Investigating the Value of Targeting Therapy

Patients with more severe anxious avoidance and distress, as defined by their baseline Oxford Agoraphobic Avoidance Scale (OAS) scores [33], were found to have the greatest improvement in anxious avoidance of, and distress in, everyday situations between the trial arms [6]. Hence, we estimated the maximum cost-effective price for the gameChange intervention when targeted to patients with high or severe anxious avoidance or distress, as defined by their baseline OAS scores.

As avoidance and distress scores were calculated separately, we investigated 4 scenarios using each baseline measure individually, and in combination, with the intention to establish the maximum cost-effective price of gameChange in those target populations who stand the most chance to benefit from the intervention.

Sensitivity Analyses

We examined the impact of excluding the patients recruited from mental health inpatient services (4/346, 1.2%) who had significantly higher health care costs than the wider patient population. Furthermore, we examined the impact of using an alternative EQ-5D crosswalk function—Hernández Alava et al [26]—to estimate health utilities from patient responses to the EQ-5D-5L survey, as opposed to the Van Hout crosswalk [25] used in the main analysis. Finally, we repeated our main analysis using only complete cases (ie, individuals with cost and outcome data available in all time periods, without multiple imputation for missing data).

Ethics Approval

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human participants or patients were approved by the NHS Research Ethics Committee (NHS South Central-Oxford B Research Ethics Committee; 19/SC/0075).

XSL•FO
RenderX

The trial was registered prospectively (ISRCTN17308399), and the trial protocol has been published [5]. Written informed consent was obtained from all participants.

Results

Missing Data

The percentages of missing data in each trial arm for resource use, EQ-5D utilities, and ReQoL-20 utilities at each follow-up point are presented in Table S3 in Multimedia Appendix 1. The overall percentage of missing data across all items was 22%. The levels of missing data were similar between the trial arms, and no baseline characteristics were associated with the probability of data being missing. Previous lagged utility values were also not significantly associated with the probability of utility or resource use data being missing. These results suggest that it may be plausible to assume that data are missing completely at random. As a sensitivity analysis, we therefore present a complete case analysis without imputation for missing data (Tables S4 and S5; Figures S1 and S2), replicating the results of the main analysis that were derived from 22 multiply imputed data sets.

Costs and Health-Related Quality of Life at Each Follow-up Time Point

Table 1 presents multiply imputed costs for each category of NHS and PSS costs by treatment allocation and follow-up period, as well as adjusted mean differences, whereas Table 2 presents EQ-5D utility scores by treatment allocation at each follow-up time point, and Table 3 presents ReQoL-20 utility scores by treatment allocation at each follow-up time point. Table S6 in Multimedia Appendix 1 presents available data without imputation on reported resource use by treatment allocation and follow-up period. Table S7 in Multimedia Appendix 1 presents available data for utility values, and Table S8 in Multimedia Appendix 1 presents available data for costs. Tables S9 and S10 in Multimedia Appendix 1 present available response-level data for the EQ-5D survey and ReQoL-20 survey, respectively.

Table 1. Period costs by treatment allocation at each follow-up time point after multiple imputation (£1=US $1.28; N=346).[a]

	Baseline to 6 weeks			6 weeks to 6 months		
	gC[b]+TAU[c] (n=174), mean (SE), £	TAU (n=172), mean (SE), £	Adjusted mean difference (95% CI), £	gC+TAU (n=174), mean (SE), £	TAU (n=172), mean (SE), £	Adjusted mean difference (95% CI), £
Total NHS[d] and PSS[e] costs	860 (205)	638 (117)	–81 (–734 to 572)	1831 (435)	1564 (414)	–36 (–694 to 622)
Mental health inpatient stays	265 (157)	105 (105)	–116 (–696 to 463)	703 (388)	551 (384)	–124 (–703 to 455)
Physical health inpatient stays	91 (91)	23 (23)	74 (–137 to 285)	147 (120)	11 (9)	141 (–70 to 352)
Medication costs	51 (7)	37 (5)	7 (–31 to 44)	168 (24)	123 (14)	37 (0 to 74)
GP[f] visits	28 (5)	34 (5)	–9 (–24 to 7)	43 (8)	44 (5)	–3 (–19 to 13)
Psychiatrist visits	166 (35)	207 (35)	–54 (–176 to 69)	260 (35)	383 (78)	–136 (–285 to 13)
Therapist visits	21 (7)	17 (5)	2 (–43 to 48)	89 (21)	121 (26)	–34 (–80 to 12)
Community mental health	186 (25)	162 (14)	16 (–63 to 96)	361 (53)	229 (22)	124 (36 to 212)
A&E[g] visits	5 (3)	16 (7)	–11 (–28 to 6)	14 (4)	28 (10)	–14 (–33 to 4)
Outpatient care	27 (7)	36 (8)	–9 (–33 to 16)	35 (9)	51 (12)	–16 (–42 to 10)
Paid help at home	19 (10)	1 (1)	18 (–6 to 42)	10 (8)	21 (12)	–11 (–36 to 13)

[a]Cost differences between the treatment arms were obtained from multilevel mixed effects models, adjusted for treatment allocation, randomized service, and site.

[b]gC: gameChange.

[c]TAU: treatment as usual.

[d]NHS: National Health Service.

[e]PSS: personal social services.

[f]GP: general practitioner.

[g]A&E: accident and emergency.

XSL•FO
RenderX

Table 2. EQ-5D utility scores by treatment allocation at each follow-up time point after multiple imputation (N=346).[a]

EQ-5D data	gC[b]+TAU[c] (n=174), mean (SE)	TAU (n=172), mean (SE)	Adjusted mean difference (95% CI)	*P* value
Baseline	0.538 (0.021)	0.545 (0.020)	N/A[d]	N/A
6 weeks	0.608 (0.021)	0.588 (0.022)	0.026 (–0.023 to 0.075)	.30
6 months	0.570 (0.023)	0.568 (0.022)	0.007 (–0.043 to 0.057)	.78

[a]Utility differences between the treatment arms were obtained from multilevel mixed effects models, adjusted for treatment allocation, baseline utility, randomized service, and site.

[b]gC: gameChange.

[c]TAU: treatment as usual.

[d]N/A: not applicable.

Table 3. Recovering Quality of Life, 20-item version (ReQoL-20), utility scores by treatment allocation at each follow-up time point after multiple imputation (N=346).[a]

ReQoL-20 data	gC[b]+TAU[c] (n=174), mean (SE)	TAU (n=172), mean (SE)	Adjusted mean difference (95% CI)	*P* value
Baseline	0.733 (0.016)	0.746 (0.015)	N/A[d]	N/A
6 weeks	0.779 (0.016)	0.765 (0.017)	0.021 (–0.017 to 0.058)	.28
6 months	0.774 (0.017)	0.792 (0.014)	–0.012 (–0.051 to 0.028)	.56

[a]A time×treatment interaction was included in both models, where the follow-up time point was used as a categorical variable.

[b]gC: gameChange.

[c]TAU: treatment as usual.

[d]N/A: not applicable.

There were no significant differences in multiply imputed NHS and PSS costs between the arms at 6 weeks or at 6 months. The cost of mental health inpatient stays was the largest driver of total NHS and PSS costs across both time periods. Participants receiving gameChange plus TAU reported higher health utility values measured using both the ReQoL-20 and the EQ-5D questionnaires at 6 weeks, although this did not reach statistical significance. However, at 6 months, the EQ-5D utility score indicated little difference between the arms, and the adjusted mean difference in the ReQoL-20 utility score indicated slightly higher utility in the control arm.

Costs and Health Outcomes Over the Whole Trial Period

Table 4 presents QALYs calculated using EQ-5D and ReQoL-20 utilities alongside NHS and PSS as well as societal costs for the whole 6-month trial period.

There were small nonsignificant improvements in incremental QALYs for the intervention arm compared with the control arm, as measured by both health-related quality-of-life instruments. Incremental QALYs calculated using EQ-5D scores indicated a larger improvement of 0.008 QALYs (95% CI –0.010 to 0.026) for those allocated to gameChange plus TAU compared with using ReQoL-20 scores, which resulted in a gain of 0.003 QALYs (95% CI –0.011 to 0.017) compared with TAU.

No significant differences in costs were detected between the trial arms. The adjusted mean difference indicated slightly lower total NHS and PSS costs for the treatment arm compared with the control arm (–£105, 95% CI –£1135 to £924). This reduction was driven by the adjusted mean difference in the cost of NHS mental health inpatient care (–£240, 95% CI –£1098 to £617), with all other NHS and PSS cost categories indicating slightly increased mean incremental costs for the treatment arm.

From a societal perspective, cost differences between the trial arms were much larger, with an adjusted cost reduction of –£1731 (95% CI –£3886 to £424) in the intervention arm. This was driven by substantially lower costs of informal care for the treatment arm compared with the control arm.

XSL•FO
RenderX

Table 4. Quality-adjusted life years (QALYs) and health care and societal costs at 6 months after multiple imputation (£1=US $1.28).a

	gC[b]+TAU[c], mean (SE)	TAU only, mean (SE)	Adjusted difference between arms, mean (95% CI)
QALYs, EQ-5D	0.293 (0.010)	0.288 (0.009)	0.008 (–0.010 to 0.026)
QALYs, ReQoL-20[d]	0.387 (0.007)	0.388 (0.006)	0.003 (–0.011 to 0.017)
Cost of mental health admissions (NHS[e]), £	969 (517)	657 (484)	–240 (–1098 to 617)
Medication costs (NHS), £	220 (30)	160 (17)	44 (–20 to 107)
General health care costs (NHS), £	1476 (237)	1356 (117)	84 (–435 to 603)
Paid help costs (NHS and PSS[f]), £	31 (15)	22 (12)	7 (–30 to 45)
Total NHS and PSS costs, £	2695 (619)	2194 (515)	–105 (–1135 to 924)
Criminal justice costs, £	42 (20)	2 (2)	38 (–0 to 77)
Unpaid caregiving (societal costs), £	2839 (400)	4403 (860)	–1576 (–3432 to 280)
Total private health care costs, £	58 (15)	135 (30)	–88 (–149 to –26)
Total societal costs, £	5634 (763)	6733 (993)	–1731 (–3886 to 424)

[a]Cost differences between the treatment arms were obtained from a linear regression model, adjusted for treatment allocation, randomized service, and site. Differences in quality-adjusted life years between the treatment arms were obtained from a linear regression model, adjusted for treatment allocation, baseline utility, randomized service, and site.

[b]gC: gameChange.

[c]TAU: treatment as usual.

[d]ReQoL-20: Recovering Quality of Life, 20-item version.

[e]NHS: National Health Service.

[f]PSS: personal social services.

Calculation of Maximum Cost-effective Price

Table 5 reports the maximum cost-effective price for the gameChange intervention for the whole trial population. Taking an NHS and PSS perspective and using QALYs calculated using the EQ-5D, the gameChange intervention could cost up to £341 per patient at the upper NICE willingness-to-pay threshold of £30,000 per QALY. Using ReQoL-20 utilities to calculate QALYs, the gameChange intervention could cost up to £193 per patient from the NHS and PSS perspective at the same £30,000 per QALY cost-effectiveness threshold. At the lower national cost-effectiveness threshold of £20,000 per QALY, the maximum value of gameChange was £262 per patient calculated using EQ-5D QALYs or £164 per patient calculated using ReQoL-20 QALYs.

When considering the intervention's impact on wider societal costs beyond the NHS and social care system, the maximum cost-effective price of gameChange was greater. At the upper cost-effectiveness threshold of £30,000 per QALY, gameChange could cost up to £1967 per patient using QALYs calculated using EQ-5D utilities or up to £1819 per patient using QALYs calculated using ReQoL-20 utilities. Using a threshold of £20,000 per QALY, the maximum cost was £1888 per patient using EQ-5D QALYs or £1790 per patient using ReQoL-20 QALYs.

The probability that gameChange is cost-effective at a range of prices is presented in Figure 1, whereas uncertainty surrounding the calculations of mean costs and effects is presented in Figure S3 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 5. Maximum cost-effective prices of gameChange using EQ-5D and ReQoL-20 utilities after multiple imputation (£1=US $1.28).[a]

Costing perspective and utilities used	Maximum cost-effective price threshold: £20,000 per QALY[b]	Maximum cost-effective price threshold: £30,000 per QALY
Total NHS[c] and PSS[d] costs (per NICE[e] guidance), £, using EQ-5D utilities (per NICE guidance)	262	341
Total NHS and PSS costs (per NICE guidance), £, using ReQoL-20 utilities	164	193
Total societal costs, £, using EQ-5D utilities (per NICE guidance)	1888	1967
Total societal costs, £, using ReQoL-20 utilities	1790	1819

[a]The maximum cost-effective price of the gameChange intervention is estimated at the lower bound and upper bound of UK willingness to pay for health interventions at £20,000 per quality-adjusted life year and £30,000 per quality-adjusted life year, respectively. This represents the maximum price that can be charged for a patient's gameChange treatment that remains cost-effective at the lower and upper bounds of the cost-effectiveness threshold.

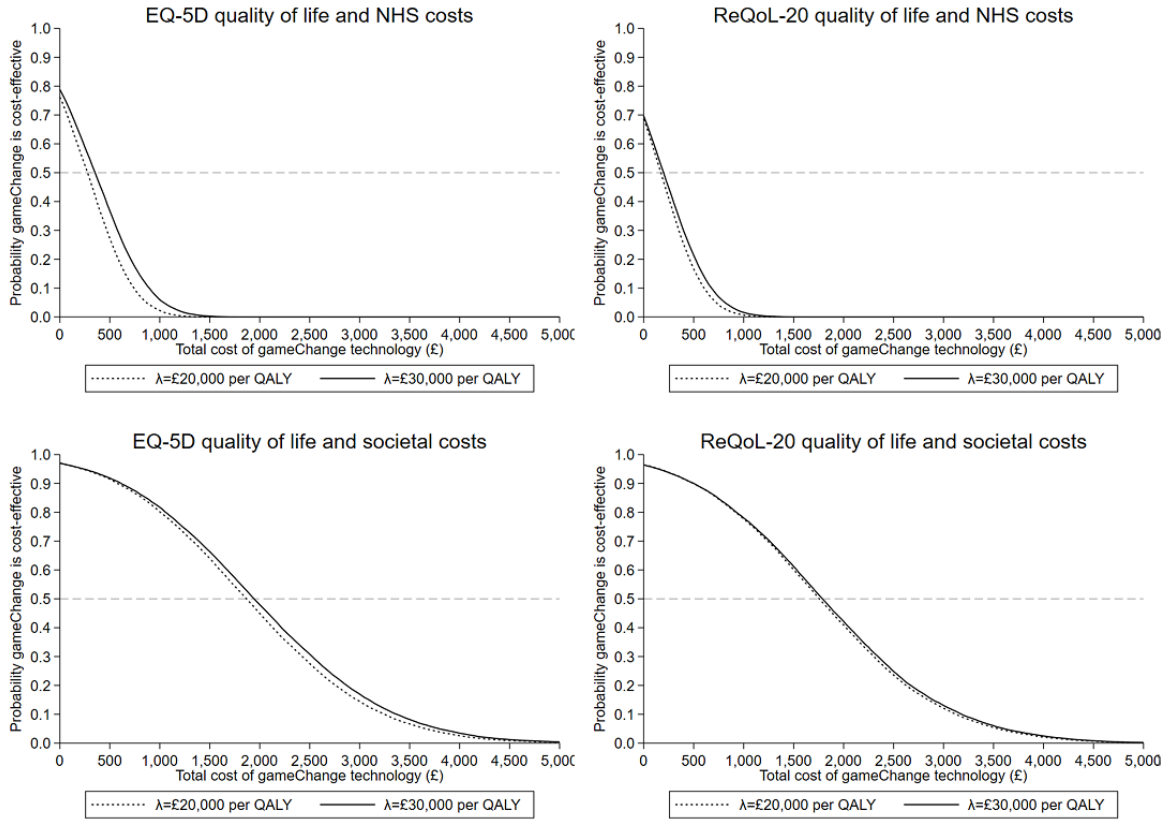
[b]QALY: quality-adjusted life year.

[c]NHS: National Health Service.

[d]PSS: personal social services.

[e]NICE: National Institute for Health and Care Excellence.

Figure 1. Uncertainty surrounding the maximum cost-effective price for the gameChange intervention after multiple imputation. Lines represent the maximum cost-effective price of gameChange for the UK National Health Service (NHS) at the National Institute for Health and Care Excellence cost-effectiveness threshold (between £20,000 and £30,000 per quality-adjusted life year [QALY]; £1=US $1.28). The dotted line represents the maximum cost-effective price of the gameChange intervention at the lower bound of the cost-effectiveness threshold (λ=£20,000 per QALY), whereas the solid line represents the maximum cost-effective price of the gameChange intervention at the upper bound of the cost-effectiveness threshold (λ=£30,000 per QALY). ReQoL-20: Recovering Quality of Life, 20-item version.



Investigating the Value of Targeting Therapy

Tables 6 and 7 show that more than half of the trial population were identified as having high or severe anxious avoidance (189/346, 54.6%) or distress (223/346, 64.5%) at baseline, measured using the OAS. Targeting gameChange to this patient group substantially increased the intervention's economic value. The maximum cost-effective price of gameChange was highest for patients with high or severe agoraphobic avoidance. At a cost-effectiveness threshold of £30,000 per QALY, gameChange could cost up to £877 per patient from an NHS and PSS perspective using EQ-5D utilities or £670 per patient using ReQoL-20 utilities. At this threshold, from a societal perspective, gameChange could cost up to £3073 per patient

XSL•FO
RenderX

using EQ-5D QALYs or £2866 per patient using ReQoL-20 QALYs. The probability that gameChange is cost-effective at a range of prices is shown for each stratified scenario in Figure S4 in Multimedia Appendix 1.

Table 6. Maximum cost-effective price of gameChange (gC) in subgroups stratified by Oxford Agoraphobic Avoidance Scale scores after multiple imputation using the EQ-5D quality-of-life measure (£1=US $1.28; N=346).[a]

	gC+TAU[b], n (%)	TAU, n (%)	Incremental QALY[c] (95% CI)	Incremental cost (95% CI), £	Maximum cost-effective price, £	
					λ=20,000 per QALY	λ=30,000 per QALY
NHS[d] and PSS[e] perspective						
Overall sample	174 (50.3)	172 (49.7)	0.008 (–0.010 to 0.026)	–105 (–1135 to 924)	262	341
High or severe avoidance	90 (51.7)	99 (57.6)	0.021 (–0.004 to 0.046)	–235 (–1986 to 1515)	663	877
High or severe distress	106 (60.9)	117 (68)	0.010 (–0.015 to 0.035)	–178 (–1683 to 1327)	374	472
High or severe avoidance or distress	124 (71.3)	126 (73.3)	0.014 (–0.008 to 0.037)	–98 (–1450 to 1255)	387	532
High or severe avoidance and distress	72 (41.4)	90 (52.3)	0.016 (–0.012 to 0.044)	–365 (–2399 to 1670)	684	844
Societal perspective						
Overall sample	174 (50.3)	172 (49.7)	0.008 (–0.010 to 0.026)	–1731 (–3886 to 424)	1888	1967
High or severe avoidance	90 (51.7)	99 (57.6)	0.021 (–0.004 to 0.046)	–2431 (–6005 to 1142)	2859	3073
High or severe distress	106 (60.9)	117 (68)	0.010 (–0.015 to 0.035)	–2101 (–5247 to 1045)	2297	2395
High or severe avoidance or distress	124 (71.3)	126 (73.3)	0.014 (–0.008 to 0.037)	–2137 (–5014 to 741)	2426	2571
High or severe avoidance and distress	72 (41.4)	90 (52.3)	0.016 (–0.012 to 0.044)	–2314 (–6398 to 1769)	2634	2794

[a]The maximum cost-effective price of the gameChange intervention is estimated at the lower (£20,000 per quality-adjusted life year) and upper (£30,000 per quality-adjusted life year) bounds of the UK cost-effectiveness threshold (λ), representing the willingness to pay for health interventions. The maximum cost-effective price represents the maximum price that can be charged for a patient's virtual reality therapy that remains cost-effective at the lower and upper bounds of the cost-effectiveness threshold.

[b]TAU: treatment as usual.

[c]QALY: quality-adjusted life year.

[d]NHS: National Health Service.

[e]PSS: personal social services.

XSL•FO
RenderX

Table 7. Maximum cost-effective price of gameChange (gC) in subgroups stratified by Oxford Agoraphobic Avoidance Scale scores after multiple imputation using the Recovering Quality of Life, 20-item version, measure (£1=US $1.28; N=346).a

	gC+TAU[b], n (%)	TAU, n (%)	Incremental QALY[c] (95% CI)	Incremental cost (95% CI), £	Maximum cost-effective price, £	
					λ=20,000 per QALY	λ=30,000 per QALY
NHS[d] and PSS[e] perspective						
Overall sample	174 (50.3)	172 (49.7)	0.003 (–0.011 to 0.017)	–105 (–1135 to 924)	164	193
High or severe avoidance	90 (51.7)	99 (57.6)	0.014 (–0.007 to 0.036)	–235 (–1986 to 1515)	525	670
High or severe distress	106 (60.9)	117 (68)	0.003 (–0.016 to 0.022)	–178 (–1683 to 1327)	240	271
High or severe avoidance or distress	124 (71.3)	126 (73.3)	0.008 (–0.010 to 0.026)	–98 (–1450 to 1255)	257	336
High or severe avoidance and distress	72 (41.4)	90 (52.3)	0.009 (–0.014 to 0.031)	–365 (–2399 to 1670)	535	620
Societal perspective						
Overall sample	174 (50.3)	172 (49.7)	0.003 (–0.011 to 0.017)	–1731 (–3886 to 424)	1790	1819
High or severe avoidance	90 (51.7)	99 (57.6)	0.014 (–0.007 to 0.036)	–2431 (–6005 to 1142)	2721	2866
High or severe distress	106 (60.9)	117 (68)	0.003 (–0.016 to 0.022)	–2101 (–5247 to 1045)	2163	2194
High or severe avoidance or distress	124 (71.3)	126 (73.3)	0.008 (–0.010 to 0.026)	–2137 (–5014 to 741)	2296	2375
High or severe avoidance and distress	72 (41.4)	90 (52.3)	0.009 (–0.014 to 0.031)	–2314 (–6398 to 1769)	2485	2570

[a]The maximum cost-effective price of the gameChange intervention is estimated at the lower (£20,000 per quality-adjusted life year) and upper (£30,000 per quality-adjusted life year) bounds of the UK cost-effectiveness threshold (λ), representing the willingness to pay for health interventions. The maximum cost-effective price represents the maximum price that can be charged for a patient's virtual reality therapy that remains cost-effective at the lower and upper bounds of the cost-effectiveness threshold.

[b]TAU: treatment as usual.

[c]QALY: quality-adjusted life year.

[d]NHS: National Health Service.

[e]PSS: personal social services.

Sensitivity Analyses

The results of a sensitivity analysis excluding the participants randomized from inpatient services (4/346, 1.2%) from the analysis sample are presented in Tables S11 to S13 in Multimedia Appendix 1. Table S11 in Multimedia Appendix 1 shows that these patients have notably higher average costs. Excluding these participants from the study sample had a substantial impact, reducing the maximum cost-effective price that could be charged for gameChange across all scenarios. Table S13 and Figure S5 in Multimedia Appendix 1 show, from an NHS perspective, that gameChange was cost-effective only when targeted to those with a high or severe OAS avoidance score. From a societal perspective, gameChange was nevertheless cost-effective for all patient groups.

A separate sensitivity analysis changed the EQ-5D mapping used to estimate EQ-5D utilities, which had a small impact on the results (Table S14 in Multimedia Appendix 1 [26]; Figure S6 in Multimedia Appendix 1 [26]), slightly reducing the maximum cost-effective price that could be charged for gameChange across all scenarios considered.

Tables S4 and S5 in Multimedia Appendix 1 and Figures S1 and S2 in Multimedia Appendix 1 present results from a complete case analysis, without imputation for missing data. The results from the complete case analysis indicate slightly larger QALY differences between the arms in favor of gameChange compared with the multiply imputed analysis; however, there were also slightly lower reductions in costs. Nevertheless, the maximum cost-effective prices of gameChange are similar to the main findings from the multiply imputed analysis. The complete case analysis showed slightly less differentiation between the maximum cost-effective prices for the whole sample and the maximum cost-effective prices for patients with high or severe OAS avoidance or distress scores.

Discussion

Principal Findings

The gameChange intervention is likely to be of economic value to the health system, particularly when the intervention is targeted to those with high or severe anxious avoidance. This is the first study to assess the economic value of an automated VR therapy and demonstrates its potential value to a resource-constrained health system. This corroborates the findings of the single prior published randomized study that established the short-term economic value of VR CBT in a

XSL•FO
RenderX

population with psychosis in the Netherlands [13]. The prior research investigated use of a nonautomated therapy that nevertheless required substantially greater therapist involvement over 16 supervised sessions, each lasting for 1 hour, compared with the automated therapy trialed in gameChange, where participants undertook a maximum of 6 VR sessions, each lasting for 30 minutes. Future research is required to establish the long-term value of both therapies within each health system because both trials only investigated quality-of-life and cost impacts of their VR intervention over a total follow-up of 6 months. It is critical to understand whether patients in receipt of VR therapies may also have improved quality of life or reduced health care costs beyond the time period of either trial, which would increase the long-term value of these interventions to the health system.

Over the 6-month trial period, the gameChange intervention is of particular value when considering a societal perspective, accounting for wider benefits beyond the health and social care system. A societal perspective is highly relevant to this patient population, who can struggle with everyday social functioning, which affects their ability to work, buy groceries, or speak with those outside their direct network. This patient group often relies on trusted friends and family to fulfill their needs, sometimes requiring significant support, which in turn affects the ability of their unpaid carers to fully contribute to society. This justifies consideration of a societal perspective when establishing the potential economic value of the gameChange intervention.

This study is based on the largest randomized trial of VR therapy to date, which collected data on many economic aspects relevant to the mental health condition studied. Our study was able to collect high-quality economic data from validated survey instruments that provide a robust assessment of changes in quality-of-life and patient costs over the 6-month trial follow-up. In particular, the diligence of the gameChange trial's research assistant interviewers helped to minimize reporting bias in participant self-reported surveys of health and social care use, and information on medication use and mental health inpatient stays was directly collected from medical records and reviewed alongside clinical colleagues to ensure the accuracy of our cost data.

In our analysis, we calculate the potential economic value of gameChange, given its impact on health-related quality of life and in offsetting costs of health and social care use, as well as wider societal costs. Therefore, the maximum cost-effective prices we present represent the maximum potential cost per patient to fully implement the gameChange intervention that would remain a cost-effective use of health care resources. Total implementation costs will include elements such as the staff time required to administer therapy, hardware costs to purchase VR devices, and software costs of obtaining user licenses to access gameChange. These implementation costs will vary, depending on how a local provider chooses to procure and deliver the gameChange intervention. As such, local providers are best able to determine whether their expected implementation costs are below the maximum cost-effective price we project for their intended population. One model of delivery used in the gameChange trial involved an NHS band 4 staff member delivering the VR therapy predominantly at participants' homes. We estimated that using this implementation model to deliver the entire trial caseload would cost approximately £184 per patient. Nevertheless, we anticipate that, with the decreasing cost and complexity of VR headsets, many patients could be provided a device for a period of time at home without requiring a staff member to be present at each therapy session, with patients instead supported by regular check-ins with mental health staff at lower cost.

Regarding hardware costs, the VR headsets used in the gameChange trial (HTC Vive Pro) have since been considerably superseded, with the hardware cost of VR headsets continuing to rapidly decline. The cost of a VR headset in the United Kingdom now stands at <£400, with a single device potentially being used to treat multiple patients over its life span and *per patient* prices therefore depending on expected local throughput. Software license costs to grant use of gameChange are currently locally determined, with costs subject to individual negotiation with the manufacturer. The maximum cost-effective prices we show in this analysis therefore allow individual health care providers to calculate whether the delivery model they intend to roll out for their local population at current local costs would represent a cost-effective use of health care resources in line with national UK cost-effectiveness guidelines. We believe that this approach maximizes transparency for health care providers looking to understand the value of implementing gameChange in their local population.

Limitations

Our study includes a number of limitations. First, data on trial participants' employment status could not be used because this data collection ceased in March 2020 at the start of the COVID-19 pandemic, which prevented us from capturing any further societal benefits of the intervention. Second, we could not control for the uneven impact of the COVID-19 pandemic on our trial outcomes because of the small numbers of participants who had completed 6-month follow-up before the pandemic. We assume that the pandemic would increase background levels of avoidance among all participants, while also affecting routine service delivery; for example, fewer GP appointments and mental health admissions being expected particularly during the first COVID-19 wave when public anxiety regarding COVID-19 and the potential infection risk from their use of health services was running high. Third, the results of a sensitivity analysis excluding trial participants randomized from psychiatric inpatient services indicate that it remains important to examine the value of gameChange for patients across different psychiatric services in future research. The small number of participants randomized from inpatient services was a major driver of the potential economic value of gameChange, with maximum cost-effective prices per patient being reduced when excluding these participants from the target population. This demonstrates the importance of adjusting for psychiatric service at randomization in the main analysis because substantially different cost and QALY profiles for participants randomized from psychiatric inpatient services likely caused skew to the mean costs and QALYs across all services. Qualitative investigation of gameChange's implementation in inpatient services is currently underway [34]; however, the small number of participants randomized from psychiatric inpatient

XSL•FO
RenderX

services limited further analytical investigation from an economic standpoint beyond the sensitivity analysis presented.

This study therefore gives health policy makers the health economic information they require to consider whether brief automated VR interventions such as gameChange should be implemented in the UK NHS. Although gameChange is shown to be economically valuable to society, its greatest value is likely to be in its ability to be implemented widely. Cost-effective interventions such as CBT and family therapy are already recommended for use in the gameChange trial population; however, only 4% (14/346) of the gameChange trial participants reported having received ≥8 sessions of either therapy at time of randomization into the trial. Because of the lack of requirement for specialist therapists, gameChange is advantaged in its potential to be cost-effective and implementable within the UK NHS, particularly when delivered by a wider supply of trained band 4 staff, which was a proven delivery method used in the gameChange trial. The gameChange intervention thus has the potential to reduce some resource pressures on the limited supply of clinical psychologists and CBT therapists, which makes gameChange potentially highly valuable to the NHS, expanding the pool of cost-effective therapies available for patients diagnosed with psychosis.

Acknowledgments

This research was funded by the National Health Service National Institute for Health and Care Research (NIHR) Invention for Innovation program (II-C7-0117-20001). This work was also supported by NIHR Oxford Health Biomedical Research Centre funding (BRC-1215-20005). The views expressed are those of the authors and not necessarily those of the National Health Service, NIHR, or the Department of Health and Social Care in England. AB and MC are supported by the NIHR Nottingham Biomedical Research Centre and NIHR MindTech MedTech Co-operative. FW is funded by a Wellcome Trust Clinical Doctoral Fellowship (102176/B/13/Z). DF is an NIHR Senior Investigator.

Data Availability

Deidentified participant data will be available in anonymized form from the team on reasonable request (including a study outline), subject to a review and contract with Oxford Health NHS Foundation Trust, after the publication of the results.

Authors' Contributions

JA and JL conducted the health economic analysis, and JA wrote the first draft of the manuscript. DF led the design of the therapy and was the trial chief investigator. FW, MC, TK, and JL contributed to the design of the trial. SL, TK, and FW contributed to the design of the therapy. MC and AB contributed to the study of therapy implementation. AB, SL, and LR contributed to the coordination of the trial. RD, KC, AM, EOR, FW, and DF led trial sites. JG and AB led the data management processes. RD, KC, SL, and LR contributed to delivery of the therapy. SL, LR, FW, and DF provided treatment supervision across study sites. JA and JL had access to all economic data used in the study and take responsibility for the integrity of the data and the accuracy of the economic analysis. All authors contributed to critical review and editing of the manuscript.

Conflicts of Interest

DF is a founder and a nonexecutive director of Oxford VR, a University of Oxford spinout company, which will commercialize the treatment. DF holds equity in Oxford VR and receives personal payments. DF holds a contract for his university team to advise Oxford VR on treatment development. SL reports consultancy work and fees from Oxford VR. RD reports grants from the National Institute for Health and Care Research, is a clinician working in the National Health Service delivering cognitive behavioral therapy (CBT), and receives payments for CBT workshops and royalties from books on CBT. The University of Oxford, Oxford Health NHS Foundation Trust, McPin Foundation, and NIHR MindTech MedTech Co-operative have received a share of the licensing fee from Oxford VR for the gameChange software.

Multimedia Appendix 1
Supplementary appendix.
[DOCX File , 730 KB - jmir_v24i11e39248_app1.docx]

References

1. Psychological Therapies for Severe Mental Health Problems (PT-SMHP). NHS England & NHS Improvement. 2022 Feb 04. URL: https://ppn.nhs.uk/resources/approved-national-pt-smhp-resources/40-psychological-therapies-for-severe-mental-health-problems-implementation-guidance/file [accessed 2022-03-30]
2. Ince P, Haddock G, Tai S. A systematic review of the implementation of recommended psychological interventions for schizophrenia: rates, barriers, and improvement strategies. Psychol Psychother 2016 Sep;89(3):324-350. [doi: 10.1111/papt.12084] [Medline: 26537838]
3. Xanidis N, Gumley A. Exploring the implementation of cognitive behaviour therapy for psychosis using the Normalization Process Theory framework. Psychol Psychother 2020 Jun;93(2):241-257. [doi: 10.1111/papt.12217] [Medline: 30672074]

XSL•FO
RenderX

4. Describing variation in numbers of people with psychosis and their access to care in England. Public Health England. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/774680/Psychosis_data_report.pdf [accessed 2022-03-30]
5. Freeman D, Yu LM, Kabir T, Martin J, Craven M, Leal J, et al. Automated virtual reality (VR) cognitive therapy for patients with psychosis: study protocol for a single-blind parallel group randomised controlled trial (gameChange). BMJ Open 2019 Aug 27;9(8):e031606 [FREE Full text] [doi: 10.1136/bmjopen-2019-031606] [Medline: 31462489]
6. Freeman D, Lambe S, Kabir T, Petit A, Rosebrock L, Yu LM, gameChange Trial Group. Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with mediation and moderation analyses. Lancet Psychiatry 2022 May;9(5):375-388 [FREE Full text] [doi: 10.1016/S2215-0366(22)00060-8] [Medline: 35395204]
7. Sanders GD, Maciejewski ML, Basu A. Overview of cost-effectiveness analysis. JAMA 2019 Apr 09;321(14):1400-1401. [doi: 10.1001/jama.2019.1265] [Medline: 30855638]
8. NICE health technology evaluations: the manual (PMG36). National Institute for Health and Care Excellence. 2022 Jan 31. URL: https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741 [accessed 2022-03-30]
9. Donker T, Cornelisz I, van Klaveren C, van Straten A, Carlbring P, Cuijpers P, et al. Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. JAMA Psychiatry 2019 Jul 01;76(7):682-690 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0219] [Medline: 30892564]
10. Wu J, Sun Y, Zhang G, Zhou Z, Ren Z. Virtual reality-assisted cognitive behavioral therapy for anxiety disorders: a systematic review and meta-analysis. Front Psychiatry 2021 Jul 23;12:575094 [FREE Full text] [doi: 10.3389/fpsyt.2021.575094] [Medline: 34366904]
11. Holst A, Björkelund C, Metsini A, Madsen JH, Hange D, Petersson EL, et al. Cost-effectiveness analysis of Internet-mediated cognitive behavioural therapy for depression in the primary care setting: results based on a controlled trial. BMJ Open 2018 Jun 14;8(6):e019716 [FREE Full text] [doi: 10.1136/bmjopen-2017-019716] [Medline: 29903785]
12. Baumann M, Stargardt T, Frey S. Cost-utility of Internet-based cognitive behavioral therapy in unipolar depression: a Markov model simulation. Appl Health Econ Health Policy 2020 Aug;18(4):567-578 [FREE Full text] [doi: 10.1007/s40258-019-00551-x] [Medline: 32060822]
13. Pot-Kolder R, Veling W, Geraets C, Lokkerbol J, Smit F, Jongeneel A, et al. Cost-effectiveness of virtual reality cognitive behavioral therapy for psychosis: health-economic evaluation within a randomized controlled trial. J Med Internet Res 2020 May 05;22(5):e17098 [FREE Full text] [doi: 10.2196/17098] [Medline: 32369036]
14. Aspvall K, Sampaio F, Lenhard F, Melin K, Norlin L, Serlachius E, et al. Cost-effectiveness of Internet-delivered vs in-person cognitive behavioral therapy for children and adolescents with obsessive-compulsive disorder. JAMA Netw Open 2021 Jul 01;4(7):e2118516 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.18516] [Medline: 34328501]
15. You JH, Luk SW, Chow DY, Jiang X, Mak AD, Mak WW. Cost-effectiveness of Internet-supported cognitive behavioral therapy for university students with anxiety symptoms: a Markov-model analysis. PLoS One 2022 May 5;17(5):e0268061 [FREE Full text] [doi: 10.1371/journal.pone.0268061] [Medline: 35511888]
16. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making 1998;18(2 Suppl):S68-S80. [doi: 10.1177/0272989X98018002S09] [Medline: 9566468]
17. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ 2002 Jul;11(5):415-430. [doi: 10.1002/hec.678] [Medline: 12112491]
18. Beecham J, Knapp M. Costing psychiatric interventions. In: Thornicroft G, Brewin C, Wing J, editors. Measuring Mental Health Needs. London, UK: Gaskell/Royal College of Psychiatrists; 1992:163-183.
19. 2019/20 National Cost Collection Data Publication. National Health Service England. 2021 Jun 22. URL: https://www.england.nhs.uk/publication/2019-20-national-cost-collection-data-publication/ [accessed 2022-11-08]
20. 2017/18 Reference Costs. National Health Service England. 2019. URL: https://webarchive.nationalarchives.gov.uk/ukgwa/20200501111106/https://improvement.nhs.uk/resources/reference-costs/ [accessed 2022-11-08]
21. Curtis LA, Burns A. Unit Costs of Health and Social Care 2020. PSSRU, University of Kent. 2020. URL: https://kar.kent.ac.uk/84818/ [accessed 2022-03-30]
22. Joint Formulary Committee. British National Formulary. BMJ Group and Pharmaceutical Press. 2019. URL: https://bnf.nice.org.uk/ [accessed 2022-03-30]
23. EQ-5D-5L. Euroqol foundation. URL: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/ [accessed 2022-03-30]
24. Long D, Polinder S, Bonsel GJ, Haagsma JA. Test-retest reliability of the EQ-5D-5L and the reworded QOLIBRI-OS in the general population of Italy, the Netherlands, and the United Kingdom. Qual Life Res 2021 Oct;30(10):2961-2971 [FREE Full text] [doi: 10.1007/s11136-021-02893-3] [Medline: 34075530]
25. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-715 [FREE Full text] [doi: 10.1016/j.jval.2012.02.008] [Medline: 22867780]

XSL•FO
RenderX

26. Alava MH, Pudney S, Wailoo A. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English Population Study. The University of Sheffield. 2020 Sep 30. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/estimating-the-relationship-betweenE-Q-5D-5L-and-EQ-5D-3L.pdf [accessed 2022-03-30]
27. Recovering Quality of Life (ReQoL). University of Sheffield. URL: https://www.reqol.org.uk/ [accessed 2022-03-30]
28. Keetharuth AD, Brazier J, Connell J, Bjorner JB, Carlton J, Taylor Buck E, et al. Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. Br J Psychiatry 2018 Jan;212(1):42-49 [FREE Full text] [doi: 10.1192/bjp.2017.10] [Medline: 29433611]
29. Keetharuth AD, Rowen D, Bjorner JB, Brazier J. Estimating a preference-based index for mental health from the recovering quality of life measure: valuation of recovering quality of life utility index. Value Health 2021 Feb;24(2):281-290 [FREE Full text] [doi: 10.1016/j.jval.2020.10.012] [Medline: 33518035]
30. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull 2010;96:5-21. [doi: 10.1093/bmb/ldq033] [Medline: 21037243]
31. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics 2014 Dec;32(12):1157-1170 [FREE Full text] [doi: 10.1007/s40273-014-0193-3] [Medline: 25069632]
32. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY, USA: Wiley Online Library; 1987.
33. Lambe S, Bird JC, Loe BS, Rosebrock L, Kabir T, Petit A, et al. The Oxford Agoraphobic Avoidance scale. Psychol Med 2021 Aug 23:1-11. [doi: 10.1017/S0033291721002713]
34. Brown P, Waite F, Lambe S, Rosebrock L, Freeman D. Virtual reality cognitive therapy in inpatient psychiatric wards: protocol for a qualitative investigation of staff and patient views across multiple national health service sites. JMIR Res Protoc 2020 Aug 20;9(8):e20300 [FREE Full text] [doi: 10.2196/20300] [Medline: 32667897]

Abbreviations

CBT: cognitive behavioral therapy
GP: general practitioner
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
OAS: Oxford Agoraphobic Avoidance Scale
PSS: personal social services
QALY: quality-adjusted life year
ReQoL-20: Recovering Quality of Life, 20-item version
TAU: treatment as usual
VR: virtual reality

Edited by G Eysenbach; submitted 05.05.22; peer-reviewed by B Bagarić, M Mudri, K Gliske; comments to author 14.07.22; revised version received 01.09.22; accepted 17.10.22; published 18.11.22.

Please cite as:
Altunkaya J, Craven M, Lambe S, Beckley A, Rosebrock L, Dudley R, Chapman K, Morrison A, O'Regan E, Grabey J, Bergin A, Kabir T, Waite F, Freeman D, Leal J
Estimating the Economic Value of Automated Virtual Reality Cognitive Therapy for Treating Agoraphobic Avoidance in Patients With Psychosis: Findings From the gameChange Randomized Controlled Clinical Trial
J Med Internet Res 2022;24(11):e39248
URL: https://www.jmir.org/2022/11/e39248
doi:10.2196/39248
PMID:36399379

©James Altunkaya, Michael Craven, Sinéad Lambe, Ariane Beckley, Laina Rosebrock, Robert Dudley, Kate Chapman, Anthony Morrison, Eileen O'Regan, Jenna Grabey, Aislinn Bergin, Thomas Kabir, Felicity Waite, Daniel Freeman, José Leal. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Online Health Information Seeking Among Patients With Chronic Conditions: Integrating the Health Belief Model and Social Support Theory

Yuxiang Chris Zhao[1], PhD; Mengyuan Zhao[2], MSc; Shijie Song[3,4], PhD

[1]School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

[2]Fudan University Library, Shanghai, China

[3]Business School, Hohai University, Nanjing, China

[4]School of Information Management, Wuhan University, Wuhan, China

Corresponding Author:
Shijie Song, PhD
Business School
Hohai University
Fo-Cheng West Rd 8
Nanjing, 211100
China
Phone: 86 15951973800
Email: ssong@hhu.edu.cn

Abstract

Background: Chronic diseases are the leading causes of death and disability. With the growing patient population and climbing health care expenditures, researchers and policy makers are seeking new approaches to improve the accessibility of health information on chronic diseases while lowering costs. Online health information sources can play a substantial role in effective patient education and health communication. However, some contradictory evidence suggests that patients with chronic conditions may not necessarily seek online health information.

Objective: This study aims to integrate 2 theories (ie, the health belief model and social support theory) and a critical health literacy perspective to understand online health information seeking (OHIS) among patients with chronic conditions.

Methods: We used the survey method to collect data from online chronic disease communities and groups on social media platforms. Eligible participants were consumers with at least 1 chronic condition and those who have experience with OHIS. A total of 390 valid questionnaires were collected. The partial least squares approach to structural equation modeling was employed to analyze the data.

Results: The results suggested that perceived risk (t=3.989, P<.001) and perceived benefits (t=3.632, P<.001) significantly affected patients' OHIS. Perceived susceptibility (t=7.743, P<.001) and perceived severity (t=8.852, P<.001) were found to influence the perceived risk of chronic diseases significantly. Informational support (t=5.761, P<.001) and emotional support (t=5.748, P<.001) also impacted the perceived benefits of online sources for patients. In addition, moderation analysis showed that critical health literacy significantly moderated the link between perceived risk and OHIS (t=3.097, P=.002) but not the relationship between perceived benefits and OHIS (t=0.288, P=.774).

Conclusions: This study shows that the health belief model, when combined with social support theory, can predict patients' OHIS. The perceived susceptibility and severity can effectively explain perceived risk, further predicting patients' OHIS. Informational support and emotional support can contribute to perceived benefits, thereby positively affecting patients' OHIS. This study also demonstrated the important negative moderating effects of critical health literacy on the association between perceived risk and OHIS.

(J Med Internet Res 2022;24(11):e42447) doi:10.2196/42447

KEYWORDS

health information seeking; patients with chronic conditions; health belief model, social support; critical health literacy

XSL•FO
RenderX

Introduction

Background

Chronic diseases are the leading global causes of death and disability. In the United States, 6 in 10 adults have 1 chronic disease, and 4 in 10 adults live with 2 or more chronic conditions [1]. According to the US Centers for Disease Control and Prevention (CDC), chronic diseases account for 3.8 trillion dollars in annual health care expenses in the United States [1]. In China, 3 chronic diseases (ie, cardiovascular diseases, cancer, and chronic respiratory diseases) were responsible for 80.7% of total deaths in 2019 [2]. Despite causing huge burdens, chronic diseases are influenced by several risk factors (eg, poor diet, physical inactivity, hyperlipidemia, and uncontrolled high blood pressure) that are generally preventable and manageable [3]. However, people living with chronic diseases often reported limited knowledge of the causes and consequences of their conditions [4]. Studies revealed that better informed patients are more likely to manage their chronic conditions, prevent exacerbations, and lower costs [5]. Due to the growing patient population and climbing health care expenditures, researchers and policy makers are seeking new approaches to improve the accessibility of health information on chronic diseases while lowering costs. Online health information sources can play a substantial role in facilitating effective patient education and health communication.

It is widely assumed that online health information seeking (OHIS) plays a significant role in the health management of patients with chronic diseases. Some evidence accords with this notion. For example, Madrigal and Escoffery [6] found that patients with chronic diseases are more likely to perform OHIS than those who are healthy and that patients with chronic diseases are more knowledgeable in OHIS. The phenomenon may be explained by the fact that health information needs trigger the OHIS process. Patients with chronic conditions have more explicit information needs than general consumers, including information on disease causes, lab testing results, and coping strategies [7-9]. Online sources are more convenient and accessible than formal health care services, so patients are assumed to perform OHIS frequently.

However, some contradictory evidence suggests that patients with chronic conditions may not necessarily seek health information. For example, McCloud and colleagues [10] conducted a mail-based survey in the United States and found that 1 in 3 cancer survivors intentionally avoided cancer-related information. Li et al [11] carried out a randomized field experiment in China and revealed that people avoid information on cancer and diabetes tests even when there is no monetary or transaction cost. A recent metareview concluded that health status is not a strong predictor of health information seeking [12]. Therefore, aside from health information needs, research questions of whether and why patients with chronic conditions seek health information online remain unresolved.

The existing research has applied many well-established theories to the portrayal of health behaviors among general consumers, such as the health belief model (HBM), social support theory, and health literacy. However, few attempts have been made to integrate these theories to understand health information behaviors comprehensively. Therefore, this paper aims to integrate 2 long-standing theories (ie, the health belief model and social support theory) and a critical health literacy perspective to understand online health information seeking among patients with chronic conditions.

Research Model and Hypotheses

OHIS Among Patients With Chronic Conditions

Patients with chronic conditions have long-term health management demands; thus, many health experts call for patient activation, an ideal state wherein patients know how to manage their conditions, keep functioning, and prevent health declines [13]. The extrinsic needs related to health management (eg, to get better informed and to manage chronic conditions) and intrinsic motivations (eg, to seek social support) motivate patients to perform OHIS [14].

Moreover, the internet provides patients with a supportive environment for OHIS. Conventional online health information sources include general search engines [15], medical databases [16], online forums [17], and so forth. Recently, social media has become one of the most popular online health information sources among users [18]. Song et al [19,20] suggest that although many social media platforms were not intentionally designed for OHIS, the rich sets of technological affordances embedded in these platforms allow users to search for health-related content and facilitate user engagement. For example, YouTube empowers patients in chronic condition management [21], and TikTok has also been a critical channel for delivering chronic disease information [22].

HBM As an Explanatory Framework in Health Behavior Research

Historically, the HBM has been widely used to understand why patients engage in proactive health behaviors. Social psychologists developed the HBM in the 1950s to explain preventive health behaviors [23]. The model assumes that the intentions of taking proactive health actions rely more on individual beliefs about a particular condition than the objective facts of the condition [24]. According to the HBM, people's proactive health behaviors are primarily determined by their *perceived susceptibility* to disease-related conditions, *perceived severity* of the consequences of disease-related conditions, *perceived benefits* of the behaviors in reducing the threats, and *perceived barriers* to the negative aspects of the health behaviors [25].

Numerous studies have investigated various health behaviors through the lens of HBM to contextualize health behaviors including a healthy diet [26], cancer screening [27], vaccination [28], medical help seeking [29], and preventive behaviors during epidemics [30]. For example, Hochbaum [31] applied the HBM when examining X-ray screening for tuberculosis and found that perceived susceptibility to tuberculosis and perceived benefits of screening varied across participants who had and had not received chest X-rays. More recently, Wong et al [28] employed the HBM to assess the acceptance of the COVID-19 vaccine and revealed that perceived severity of contracting COVID-19 and perceived benefits of receiving the vaccine

XSL•FO
RenderX

positively predicted vaccine acceptance. Overall, these studies produced internally consistent results that provided fairly strong support for HBM and informed the subsequent use of HBM to understand health behaviors. Despite the intensive use of HBM in health and medical contexts, the model is less adopted to investigate health information behaviors. Given the considerable explanatory power of HBM in health sciences, this study will employ the HBM to investigate OHIS intentions among patients with chronic conditions.

Although the HBM does not specify the variable ordering, it implicitly purports the idea that perceived susceptibility and severity jointly lead to a perception of the risk of disease, and perceived benefits influence an individual's assessment of the outcome of the proactive health behaviors [32]. As such, the risk-benefit consideration motivates the individual to take action. Noteworthily, the HBM does not provide rules of combinations of the constructs. For example, Harrison et al [33] did not include the cues to action and health motivation components in their analyses. Ahadzadeh et al [34] only included risk perceptions (ie, perceived susceptibility and perceived severity) when using the HBM. According to a recent systematic review [27], the risk-benefit aspect is the most frequently explored component in prior studies. Therefore, this study will also focus on the risk-benefit perspective.

The risk-benefit relationship posited by HBM has been partially examined in prior studies. For example, Ahadzadeh et al [34] found that risk perceptions had an indirect positive effect on Malaysian women's online health-related internet use. Mou et al [35] observed that perceived benefits of online health websites, perceived susceptibility, and perceived severity of one's health conditions were significant predictors of online health information seeking. Accordingly, our study proposes 2 hypotheses based on the parsimonious form of the HBM: (1) The OHIS of patients with chronic diseases is positively influenced by the perceived risk of chronic diseases (H1a) and the perceived benefits of performing OHIS (H1b); and (2) the perceived risk of chronic diseases of patients with chronic conditions is positively influenced by perceived susceptibility (H2a) and severity (H2b).

However, explicating the relationship between the HBM constructs cannot resolve all the theoretical limitations of the HMB. To overcome these constraints, researchers have often treated the HBM as an overarching framework [36] and combined its constructs with other theories [37]. For instance, Ahadzadeh et al [34] incorporated the HBM and the technology acceptance model to understand users' online health-related internet behaviors. Mou et al [35] integrated the HBM, the extended valence framework, and the perspective of self-efficacy to explain users' OHIS. Since prior work suggested that OHIS is associated with social support and health literacy [38], we will integrate the perspectives of social support and health literacy in this study.

Social Support in OHIS

Social support is often described as the comfort, help, or information that an individual obtains from others [39]. In offline settings, social support is often provided by friends and relatives [40]. In online environments, social media serves as an important source of social support for patients. For example, Zhang and He [41] found that people living with diabetes exchange medical and lifestyle information and provide and seek social support in Facebook groups. These Facebook diabetes groups share a broad variety of topics, such as nutrition, medications, blood glucose screening, and physical activity [42].

Social support has been extensively examined in health-related fields, with many studies finding positive associations between social support and people's physical and mental health [43,44]. The benefits of social support are especially evident in patients' self-management of chronic conditions [45]. However, despite its promising positive impacts, the mechanisms of how social support influences health behaviors remain underexplored. A couple of studies examined the direct associations between informational and emotional support and health behaviors or conditions. For example, Wang and Parameswaran [46] suggested that adequate online social support is correlated with better self-care behaviors of HIV patients. However, other studies revealed that the impacts of social support on health behaviors are mediated by different factors, such as health self-efficacy and health information seeking [47,48].

Although social support is a multifaceted concept with different subdimensions, informational and emotional supports are the most frequently studied aspects in the existing health literature [49]. Savolainen [50] found that dietary information seekers solicited emotional support in health blogs by describing their dieting problems, and readers responded by offering considerable informational and emotional support. Stellefson and Paige [42] surveyed the 34 largest diabetes support groups on Facebook and revealed that informational and emotional support exchanges were the 2 most common purposes for creating those groups. Therefore, this study will focus on these 2 main types of social support.

Regarding patients' motivations for seeking online sources for social support, some researchers suggest a compensation view and posit that online sources can fulfill patients' social support deficits from offline settings [51,52]. However, Guillory and Niederdeppe [53] found that patients who already had sufficient social support from families and friends were also likely to seek online health information. McKinley and Wright [47] assert that although their inconsistent findings cannot fully support the compensation view, they demonstrate that online social supports are helpful for the end users. Accordingly, we propose our third hypothesis (H3): The perceived benefits of online sources for patients with chronic conditions are positively influenced by online emotional support (H3a) and informational support (H3b).

Critical Health Literacy in OHIS

Health literacy refers to "the degree to which individuals can obtain, process, understand, and communicate about health-related information needed to make informed health decisions" [54]. According to Nutbeam [54], health literacy is a hierarchical concept consisting of multiple layers, depending on different levels of advancement of the literacy. While functional literacy refers to basic skills in reading and writing regarding health information, critical literacy refers to the

XSL•FO
RenderX

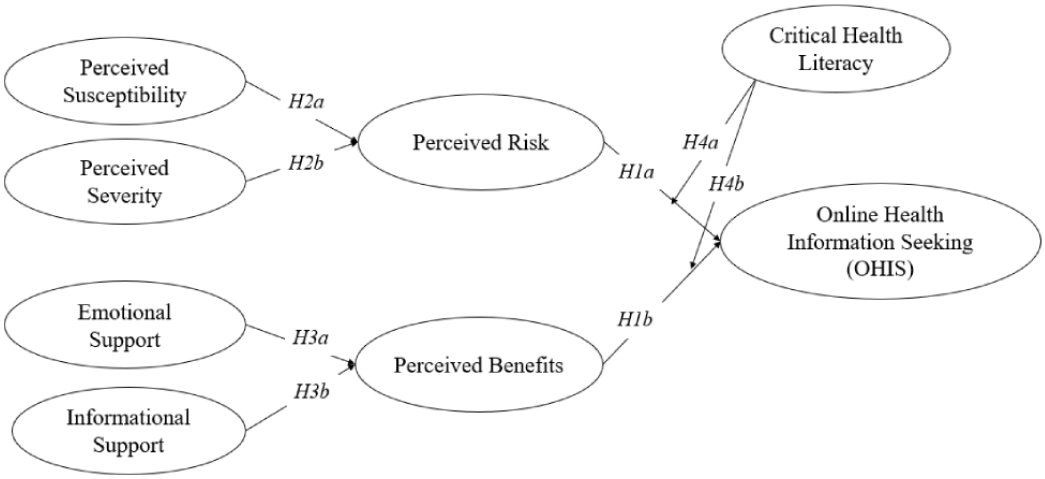
advanced cognitive skills in analyzing health information critically.

Early studies treated health literacy as a holistic concept and found varied associations between health literacy and patients' health behaviors [55]. However, many recent studies revealed that the different components of health literacy have different power in explaining health behaviors. For example, Heijmans and Waverijn [56] found that critical health literacy is related to self-management, but functional health literacy is not. Matsuoka and Tsuchihashi-Makaya [57] revealed similar findings that critical health literacy influences self-care and consulting behaviors but functional health literacy does not. Based on these findings, we argue that critical health literacy may influence patients' information behaviors.

Moreover, prior studies suggested that patients with chronic conditions were concerned about the information quality, although they mostly agreed that online health information was easy to find [58]. These findings indicated that some patients might be knowledgeable about their health conditions [9] and thus are more critical when it comes to health information assessment. Therefore, we posit that the effects of the perceived risk and benefits of OHIS are moderated by critical health literacy. When patients have higher critical health literacy, they are more cautious when choosing online health information sources and may turn to authoritative sources such as offline health care providers. Thus, we propose the following hypotheses (H4): Critical health literacy negatively moderates the associations between perceived risk (H4a) and perceived benefits (H4b) and patients' OHIS.

The research model and hypotheses are shown in Figure 1.

Figure 1. Proposed research model.



Methods

Measurement Instrument

Most of the construct items in this study were adapted from validated existing scales. Each item was measured following a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The 3 OHIS items were adapted from studies by Deng and Liu [48] and Li and Wang [59]. The 3 items measuring perceived risk were developed from Kahlor [60]. The perceived benefits scales were adjusted from McKinley and Wright [47]. The perceived susceptibility and severity were gauged based on studies by Ahadzadeh et al [34] and Shang and Zhou [61]. Measurements of emotional and informational support were derived from studies by Deng and Liu [48] and Li and Wang [59]. Three items for critical health literacy drew on the measurement developed by Ishikawa and Takeuchi [62] and converted into an index. The constructs and measures are shown in Table 1.

The questionnaire was formed in 2 stages. First, we used translation (from English to Chinese) and back-translation (from Chinese to English) techniques to design the questionnaire to ensure its reliability. Second, we invited 20 patients living with chronic disease to participate in a pilot survey. We gathered their feedback and suggestions during the completion of the initial questionnaire to further modify the questionnaire, which resulted in the final version of the questionnaire.

XSL•FO
RenderX

Table 1. Constructs and measures.

Constructs	Measures	References
Online health information seeking	• OHIS[a]1: I want to seek health information often on the internet. • OHIS2: I am willing to search the internet for relevant health information when I need it. • OHIS3: I will seek health information on the internet before making health decisions.	Deng and Liu [48]; Li and Wang [59]
Perceived risk	• PCR[b]1: I am constantly worried about my health condition. • PCR2: I fear that my chronic condition would probably attack or worsen. • PCR3: If my chronic disease attacks or worsens, it would have a serious impact on my work or life.	Kahlor [60]
Perceived benefits	• PBF[c]1: Health information on the internet could be useful for me. • PBF2: Health information on the internet could be helpful to me. • PBF3: Health information on the internet could help me become familiar with health knowledge.	McKinley and Wright [47]
Perceived susceptibility	• PSU[d]1: The health-related issues mentioned in the internet health information are likely to happen on me. • PSU2: There is a good possibility that I will experience the health-related issues mentioned in the internet health information. • PSU3: I am likely to contract the health-related issues mentioned in the internet health information.	Ahadzadeh et al [34]; Shang and Zhou [61]
Perceived severity	• PSE[e]1: The consequences of the health-related issues mentioned in the internet health information may be serious for me. • PSE2: Contracting the health-related issues mentioned in the internet health information would be likely to cause me major problems. • PSE3: Suffering from the health-related issues mentioned in the internet health information is a serious problem for me.	Shang and Zhou [61]
Emotional support	• ES[f]1: When faced with difficulties, some individuals on the internet comforted and encouraged me. • ES2: When faced with difficulties, some individuals on the internet expressed interest in and concern for my well-being. • ES3: When faced with difficulties, some individuals on the internet are on side with me.	Deng and Liu [48]; Li and Wang [59]
Informational support	• IS[g]1: When faced with difficulties, some individuals on the internet would offer suggestions when I needed help. • IS2: When faced with difficulties, some individuals on the internet would give me information to help me overcome the problem. • IS3: When faced with difficulties, some individuals on the internet would help me discover the cause and provide me with suggestions.	Li and Wang [59]
Critical health literacy	• CHL[h]1: Since being diagnosed with chronic diseases, I have considered whether the information was applicable to my situation. • CHL2: Since being diagnosed with chronic diseases, I have considered the credibility of the information. • CHL3: Since being diagnosed with chronic disease, I have checked whether the information was valid and reliable.	Ishikawa and Takeuchi [62]

[a]OHIS: online health information seeking.

[b]PCR: perceived risk.

[c]PBF: perceived benefit.

[d]PSU: perceived susceptibility.

[e]PSE: perceived severity.

[f]ES: emotional support.

[g]IS: informational support.

[h]CHL: critical health literacy.

XSL•FO
RenderX

Ethics Approval

This study was approved by the Institutional Review Boards of the School of Economics and Management of the Nanjing University of Science and Technology (20201101).

Data Collection

The questionnaire was distributed from 2 main channels. First, we recruited participants through online chronic disease health communities. Five typical online health forums (ie, diabetes, hypertension, chronic gastritis, hyperlipoidemia, and rhinitis) were chosen in each of the leading Chinese communities (ie, Baidu Tieba and Douban groups). We also distributed the questionnaire through chronic disease health groups on general social media platforms (eg, WeChat). Eligible participants were consumers with at least 1 chronic condition who sought health information online during the past 12 months. The questionnaire contained a consent form that included the details of the study. Participants who agreed to the consent continued to the questionnaire. Each participant received a cash incentive of 5 renminbi (RMB) (about US $0.8) after completing the questionnaire. We received 426 questionnaires from October 18 to 29, 2021. After eliminating incomplete and invalid questionnaires by applying the eligibility criteria, we finally obtained a sample consisting of 390 valid responses.

Statistical Analysis

The respondents' characteristics are illustrated in Table 2. Of the participants, 64.1% (n=250) were male, and 35.9% (n=140) were female. The age coverage was relatively broad, comprising young people under the age of 20 and older adults above the age of 60 years. Respondents' places of residence were relatively balanced, with 46.7% (n=182) of participants living in urban areas and 53.3% (n=208) living in rural areas. Approximately half (n=192, 49.2%) of the participants had college degrees. In terms of health status, 38.5% (n=150) of the participants reported feeling normal, 25.6% (n=100) felt bad, and 35.9% (n=140) felt good or very good. Participants reported various types of chronic conditions. Chronic gastritis (n=146, 37.4%) was the most frequently mentioned condition, followed by diabetes (n=114, 29.2%) and hyperlipidemia (n=98, 25%). About half (n=193, 49.5%) of the participants had 1 chronic condition, 31.79% (n=124) had 2, and 4% (n=17) had 4 or more conditions.

We also measured the types of health information that participants sought using a typology from Zhao and Zhao [38]. Participants most frequently sought health information about disease symptoms (n=209, 53.6%), medical resources (n=201, 51.5%), and health prevention (n=199, 51%). Additionally, we counted the online health information sources that the participants used. Medical and health apps (n=187, 48%) were the most frequently reported online health information source, followed by social question-and-answer platforms (n=179, 46%) and short video platforms (n=174, 44.6%). Regarding OHIS frequency, all the participants reported they had sought online health information at least once during the past 6 months, and 39.5% (n=154) participants reported that they had sought online health information relatively often or very frequently.

XSL•FO
RenderX

Table 2. Characteristics of respondents.

Measure and item	Value, n (%)
Sex	
Male	250 (64.1)
Female	140 (35.9)
Age	
＜20	13 (3.33)
20-29	131 (33.6)
30-39	137 (35.1)
40-49	58 (14.9)
50＜59	35 (9)
≥60	16 (4.1)
Place of residence	
Urban	182 (46.7)
Rural	208 (53.3)
Education level	
Junior high school or below	58 (14.9)
Senior high school	98 (25.1)
Technical secondary school	42 (10.8)
Associate degree	72 (18.5)
Bachelor’s degree	103 (26.4)
Master’s degree	17 (4.4)
Monthly income (RMB[a])	
＜1500	17 (4.4)
1500-2999	55 (14.1)
3000-3999	112 (28.7)
4000-4999	68 (17.4)
5000-5999	71 (18.2)
6000-6999	29 (7.4)
≥7000	38 (9.7)
Profession	
Currently in health care profession	46 (11.8)
Past worked in health care profession	208 (53.3)
Never worked in health care profession	136 (34.9)
Health status	
Very bad	17 (4.4)
Relatively bad	83 (21.3)
Normal	150 (38.5)
Relatively good	100 (25.6)
Very good	40 (10.3)
Type of chronic disease	
Chronic gastritis	146 (37.4)
Diabetes	114 (29.2)
Hyperlipoidemia	98 (25.1)

XSL•FO
RenderX

Measure and item	Value, n (%)
Hypertension	76 (19.5)
Rhinitis	72 (18.5)
Rheumatism	62 (15.9)
Lumbar disc bulging	37 (9.5)
Asthma	33 (8.5)
Chronic conjunctivitis	33 (8.5)
Other	10 (2.6)
Number of chronic diseases	
1	193 (49.5)
2	124 (31.8)
3	56 (14.4)
4	11 (2.8)
＞4	6 (1.5)
Type of health information	
Disease symptoms	209 (53.6)
Medical resource	201 (51.5)
Health prevention	199 (51)
Medication/treatment	111 (28.5)
Health promotion	94 (24.1)
Other	4 (1)
Source of health information	
Medical and health apps	187 (48)
Social question-and-answer platforms	179 (45.9)
Short video platforms	174 (44.6)
Social platforms	122 (31.3)
Search engines	111 (28.5)
News clients	56 (14.4)
Other	8 (2.1)
Frequency of searching	
Occasionally	83 (21.3)
Sometimes	153 (39.2)
Relatively often	127 (32.6)
Very frequently	27 (6.9)

[a]RMB: renminbi.

Results

Approach

We employed a partial least squares (PLS) approach to structural equation modeling (SEM) on testing the proposed model. Previous studies have shown that the PLS-SEM method is suitable for testing theoretically constructed models [63] and validating relatively complex models [64]. In addition, PLS-SEM can deal with nonnormally distributed samples, which is advantageous when processing relatively small sample sizes [65]. We used SmartPLS 3 software (SmartPLS GmbH) to analyze the data and test the structural model.

Measurement Model

Drawing on Shang and Zhou [61], we adopted reliability, convergent, and discriminant validity to evaluate the measurement model. Table 3 reports the reliability and convergence validity results. The reliability was judged based on the Cronbach alpha and composite reliability values. The results show that all Cronbach alpha and composite reliability values were greater than the proposed threshold of 0.7 [66],

XSL•FO
RenderX

indicating qualified reliability. The convergence validity was examined by the values of average variance extracted (AVE). The results show that AVEs were higher than the recommended value of 0.5 [67], and all indicator loadings exceeded the threshold of 0.7, suggesting satisfactory convergence validity.

The discriminant validity was checked by testing both the Fornell-Larcker criteria [68] and the heterotrait-monotrait ratio (HTMT) [69]. Table 4 suggested that the square root of AVE values for each construct exceeded all its correlation coefficients with other constructs, indicating promising discriminant validity [68]. Moreover, all HTMT values were below the recommended value of 0.85 (Table 5), suggesting good discriminant validity [69]. The foregoing results verify the discriminant validity of all the constructs in our study.

Table 3. Reliability and convergence validity.

Constructs and items	Indicator loading	Cronbach alpha	Composite reliability	AVE[a]
Perceived susceptibility		.814	.890	.729
PSU[b]1	.881			
PSU2	.795			
PSU3	.883			
Perceived severity		.852	.910	.772
PSE[c]1	.888			
PSE2	.861			
PSE3	.886			
Informational support		.831	.898	.747
IS[d]1	.883			
IS2	.832			
IS3	.878			
Emotional support		.856	.913	.777
ES[e]1	.896			
ES2	.861			
ES3	.888			
Perceived risk		.835	.901	.752
PCR[f]1	.882			
PCR2	.834			
PCR3	.885			
Perceived benefits		.821	.894	.737
PBF[g]1	.867			
PBF2	.834			
PBF3	.874			
Online health information seeking		.824	.895	.740
OHIS[h]1	.881			
OHIS2	.823			
OHIS3	.874			

[a]AVE: average variance extracted.

[b]PSU: perceived susceptibility.

[c]PSE: perceived severity.

[d]IS: informational support.

[e]ES: emotional support.

[f]PCR: perceived risk.

[g]PBF: perceived benefit.

[h]OHIS: online health information seeking.

XSL•FO
RenderX

Table 4. Discriminant validity (Fornell-Larcker criterion)[a].

Constructs	1	2	3	4	5	6	7
1. Emotional support	.881	—	—	—	—	—	—
2. Online health information seeking	.571	.860	—	—	—	—	—
3. Informational support	.621	.526	.864	—	—	—	—
4. Perceived benefits	.684	.660	.676	.858	—	—	—
5. Perceived risk	.526	.578	.461	.585	.867	—	—
6. Perceived severity	.522	.576	.460	.582	.717	.879	—
7. Perceived susceptibility	.513	.529	.442	.488	.698	.629	.854

[a]Values on the diagonal represent the square root of average variance extracted (AVE) for each construct.

Table 5. Discriminant validity (heterotrait-monotrait ratio).

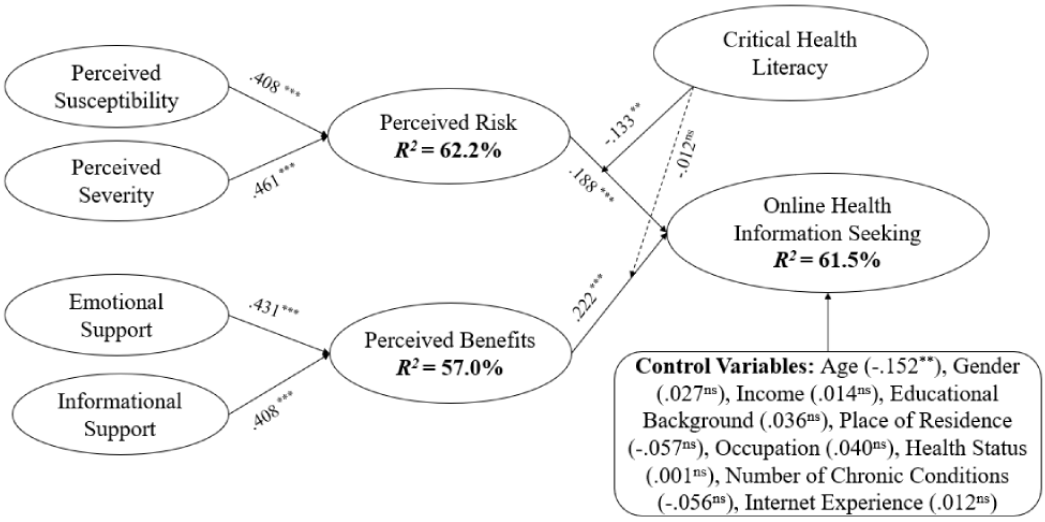
Items	1	2	3	4	5	6	7
1. Emotional support							
2.Online health information seeking	.677						
3. Informational support	.735	.633					
4. Perceived benefits	.816	.799	.815				
5. Perceived risk	.624	.692	.550	.707			
6. Perceived severity	.610	.687	.544	.694	.848		
7. Perceived susceptibility	.615	.643	.537	.598	.844	.754	

Structural Model

We adopted standard bootstrap in SmartPLS 3 on 5000 bootstrapping samples to examine the structural model's path coefficients and corresponding significance levels. Figure 2 shows the results of the PLS-SEM analysis, where perceived risk, perceived benefits, and online health seeking behavior are explained by the independent variables with variance values of 62.2%, 57%, and 61.5%, respectively, indicating a good explanation of the structural model.

The hypotheses testing results (Table 6) show that perceived risk (β=.188, *P*<.001) and perceived benefits (β=.222, *P*<.001) have significant positive effects on OHIS, supporting both H1a and H1b. As for health beliefs, perceived susceptibility (β=.408, *P*<.001) and perceived severity (β=.461, *P*<.001) significantly influence perceived risk, indicating that both H2a and H2b are supported. Concerning social support, both emotional support (β=.431, *P*<.001) and informational support (β=.408, *P*<.001) have positive effects on perceived risk, supporting H3a and H3b. Moreover, we tested the moderating effects of critical health literacy. The results show that critical health literacy (β=−.133, *P*=.002) has negative moderating effects on the relationship between perceived risk and OHIS, which supports H4a. However, critical health literacy cannot significantly moderate the relationship between perceived benefits and OHIS (β=−.012, *P*=.774). Therefore, H4b is not supported.

Figure 2. Structural model results. ns: nonsignificant. ***P<.001, **P<.01, and *P<.05.



XSL•FO
RenderX

Table 6. Hypotheses testing results.

Hypotheses	Paths	Path coefficients	*t*-statistic	*P* value	Hypothesis validation
H1a	PCR -> OHIS	.188	3.989	<.001	Supported
H1b	PBF -> OHIS	.222	3.632	<.001	Supported
H2a	PSU -> PCR	.408	7.743	<.001	Supported
H2b	PSE -> PCR	.461	8.852	<.001	Supported
H3a	ES -> PBF	.431	5.748	<.001	Supported
H3b	IS -> PBF	.408	5.761	<.001	Supported
H4a	PCR×CHL -> OHIS	−.133	3.097	.002	Supported
H4b	PBF×CHL -> OHIS	−.012	0.288	.774	Not supported

Discussion

Principle Findings

In this study, we investigated the effects of perceived risk and perceived benefits on OHIS among patients with chronic conditions. Based on HBM, we examined the influencing factors of perceived risk using 2 antecedents: perceived susceptibility and perceived severity. Additionally, drawing on social support theory, we explored the impact of informational and emotional support on perceived benefits of patients' OHIS. This study also focused on critical health literacy and how it moderates the effects of perceived risk and perceived benefits on OHIS. We proposed a research model by integrating the aforementioned theories and developed corresponding measurement instruments. Data were collected from online chronic disease communities and social media groups using the survey method and analyzed using the PLS-SEM method.

The results suggested that perceived risk (t=3.989, P<.001) and perceived benefits (t=3.632, P<.001) significantly affected patients' OHIS. Perceived susceptibility (t=7.743, P<.001) and perceived severity (t=8.852, P<.001) were found to significantly influence the perceived risk of chronic diseases. Informational support (t=5.761, P<.001) and emotional support (t=5.748, P<.001) also impacted the perceived benefits of online sources for patients. In addition, moderation analysis showed that critical health literacy significantly moderates the relationship between perceived risk and OHIS (t=3.097, P=.002) but not the relationship between perceived benefits and OHIS (t=0.288, P=.774).

Implications

This study makes contributions to both theory and practice. From a theoretical perspective, we extend the HBM into information behavior research by integrating it with the social support theory. The HBM suggests that belief in health risk predicts the likelihood of engaging in health-related behaviors [37]. Prior work shows that individuals with higher perceived risk have a stronger motivation to perform health-related behaviors and change their health conditions [34,70]. Among them, patient-initiated OHIS can undoubtedly meet patients' health information needs and promote positive health information behaviors to a certain extent. In addition to patients' spontaneous health beliefs, this paper argues that social determinants of health can largely contribute to patients' health information behaviors—social support as an intermediary social determinant predicts patients' OHIS. We believe this assertion can simultaneously enrich the HBM and literature on health information behaviors. Our empirical study confirms the validity of this extension. Wilson [71] suggested that the disciplines of health and medical sciences and information sciences share a prominent common interest in information behavior research, and the flows of ideas and theories from the community of interest would also benefit information behavior research.

Additionally, we contextualize health literacy in chronic diseases by proposing and testing how critical health literacy moderates the relationship between health beliefs and social support to patients' OHIS. Prior work has explored the measurement of critical health literacy for patients with chronic diseases and the impact on self-management of health [56,72]. However, few studies have analyzed the impact of critical health literacy on OHIS. Our analysis contributes to the literature by uncovering a negative moderating effect between perceived risk and OHIS. We speculated that patients with higher critical health literacy may also be more capable in health information seeking and source selections. When patients with higher critical health literacy perceive a greater health risk, they may not necessarily search for health information on the internet and social media, given the general information quality concerns with online sources; instead, they are likely to seek more professional medical advice and visit doctors directly. This finding allows us to reexamine the compound influences of OHIS and seek more theoretical support from a psychological perspective.

From a practical perspective, this study suggests that online health communities should provide sufficient social support to patients and create a reciprocal virtual community. This social support can come from high-quality content created by professionals or emotional support generated by the mutual help between patient-patient and doctor-patient interactions. Meanwhile, online health communities should encourage surrogate health information seeking among patients and enhance the sense of belonging to the virtual community through gamification incentives and participatory design methods.

Finally, online health platforms need to better segment their users by providing targeted professional services to differentiated patients according to their varied health literacy levels instead of the traditional demographic profiles. Patients can become well informed about their health conditions and

XSL•FO
RenderX

evolve into "expert patients." Expert patients with high health literacy usually have higher health information quality standards and prefer to go to offline professional medical institutions for consultation. Therefore, online health communities could consider inviting health care experts to carry out freemium consultations with more specialized, personalized, and accurate services to retain patients with higher critical health literacy and enhance their stickiness and loyalty to online health platforms.

Limitations and Future Work

This study has several limitations. First, the underlying influence mechanism between the 2 theories (ie, the health belief model and the social support theory) needs to be further empirically demonstrated. Future research could consider health beliefs as mediating constructs to unravel the effects of social determinants of health on individuals' perceived risks and benefits and further draw on social cognitive theory to empirically explore this mediating effect.

Second, we identified the moderating effect of critical health literacy in OHIS; however, the moderation analysis indicates that more contextualized measures are needed to validate the working mechanisms of critical health literacy. Future research needs to uncover how critical health literacy moderates the patients' OHIS intentions. Additionally, future research could further empirically analyze the constituent domains of critical health literacy [72] in terms of the dimensions of the constructs and how they are measured. Furthermore, researchers may also consider a randomized controlled trial to explore the effects of improved critical health literacy on OHIS.

Third, the generalizability may be limited as our sample is restricted to chronic disease patients in China. Our findings may not be applicable to other countries, regions, and contexts. Future work may conduct cross-cultural and cross-national comparisons to better generalize this study's results. Moreover, this is a cross-sectional study; due to the diversity of chronic diseases and the dynamic nature of chronic conditions, more longitudinal studies are needed in the future to reveal the dynamic effects of changes in health beliefs and social support on OHIS among patients with chronic diseases. Experience sampling methods and action research approaches are recommended to improve the validity of the research through multiwave data collection.

Conclusions

This paper contributes to the literature on OHIS by integrating the HBM and the social support theory. The integrated model suggested that health beliefs and social support positively impact OHIS among patients with chronic diseases. In particular, perceived susceptibility and severity can positively impact perceived risk, further influencing patients' OHIS. Informational support and emotional support can contribute to perceived benefits, further positively affecting patients' OHIS. This study also demonstrated critical health literacy's important negative moderating effects on the association between perceived risk and OHIS. Theoretical and practical implications for leveraging OHIS for patients with chronic diseases were also provided.

Acknowledgments

This work was jointly supported by grants from the National Natural Science Foundation of China (72204076; 72074112) and Research Funds of Humanities and Social Sciences for the Central Universities (B220201054).

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. About chronic diseases. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/chronicdisease/about/index.htm [accessed 2022-04-06]
2. SCIO holds press briefing on "Report on the Nutrition and Chronic Disease Status of Chinese Residents". China SCIO. URL: http://english.scio.gov.cn/pressroom/2020-12/23/content_77043604.htm [accessed 2022-04-06]
3. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. Lancet 2014 Jul;384(9937):45-52. [doi: 10.1016/s0140-6736(14)60648-6]
4. Hernandez P, Balter M, Bourbeau J, Hodder R. Living with chronic obstructive pulmonary disease: A survey of patients' knowledge and attitudes. Respir Med 2009 Jul;103(7):1004-1012. [doi: 10.1016/j.rmed.2009.01.018]
5. Kueh YC, Morris T, Ismail A. The effect of diabetes knowledge and attitudes on self-management and quality of life among people with type 2 diabetes. Psychol Health Med 2017 Feb 05;22(2):138-144. [doi: 10.1080/13548506.2016.1147055] [Medline: 26851120]
6. Madrigal L, Escoffery C. Electronic health behaviors among US adults with chronic disease: cross-sectional survey. J Med Internet Res 2019 Mar 05;21(3):e11240 [FREE Full text] [doi: 10.2196/11240] [Medline: 30835242]

XSL•FO
RenderX

7. Zhang Y, Yi S, Trace CB, Williams-Brown MY. Understanding the information needs of patients with ovarian cancer regarding genetic testing to inform intervention design: interview study. JMIR Cancer 2022 Feb 08;8(1):e31263 [FREE Full text] [doi: 10.2196/31263] [Medline: 35133282]
8. Pian W, Song S, Zhang Y. Consumer health information needs: A systematic review of measures. Inf Process Manag 2020 Mar;57(2):102077. [doi: 10.1016/j.ipm.2019.102077]
9. Cordier J. The expert patient: towards a novel definition. Eur Respir J 2014;44(4):853-857. [doi: 10.1183/09031936.00027414]
10. McCloud RF, Jung M, Gray SW, Viswanath K. Class, race and ethnicity and information avoidance among cancer survivors. Br J Cancer 2013 May 28;108(10):1949-1956 [FREE Full text] [doi: 10.1038/bjc.2013.182] [Medline: 23681189]
11. Li Y, Meng J, Song C, Zheng K. Information avoidance and medical screening: a field experiment in China. Management Science 2021 Jul;67(7):4252-4272. [doi: 10.1287/mnsc.2020.3723]
12. Chang C, Huang M. Antecedents predicting health information seeking: A systematic review and meta-analysis. IJIM 2020 Oct;54:102115. [doi: 10.1016/j.ijinfomgt.2020.102115]
13. Hibbard J, Stockard J, Mahoney E, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4p1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
14. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of "Dr Google": a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015 Dec 29;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
15. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014 Dec 02;16(12):e262. [doi: 10.2196/jmir.3706]
16. Zhang Y. Searching for specific health-related information in MedlinePlus: behavioral patterns and user experience. J Assn Inf Sci Tec 2014 Jan;65(1):53-68. [doi: 10.1002/asi.22957]
17. Hirvonen N, Tirroniemi A, Kortelainen T. The cognitive authority of user-generated health information in an online forum for girls and young women. JD 2019 Jan 7;75(1):78-98. [doi: 10.1108/jd-05-2018-0083]
18. De Choudhury M, Morris M, White R. Seeking and sharing health information online: comparing search engines and social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 26-May 1; Toronto, ON p. 1365-1376. [doi: 10.1145/2556288.2557214]
19. Song S, Zhao YC, Yao X, Ba Z, Zhu Q. Serious information in hedonic social applications: affordances, self-determination and health information adoption in TikTok. JD 2022 May 30;78(4):890-911. [doi: 10.1108/jd-08-2021-0158]
20. Song S, Zhao YC, Yao X, Ba Z, Zhu Q. Short video apps as a health information source: an investigation of affordances, user experience and users' intention to continue the use of TikTok. INTR 2021 Nov 12;31(6):2120-2142. [doi: 10.1108/intr-10-2020-0593]
21. Liu X, Zhang B, Susarlia A, Padman R. Go to YouTube and call me in the morning: use of social media for chronic conditions. MISQ 2020 Mar;44(1):257-283. [doi: 10.25300/misq/2020/15107]
22. Song S, Xue X, Zhao YC, Li J, Zhu Q, Zhao M. Short-video apps as a health information source for chronic obstructive pulmonary disease: information quality assessment of TikTok videos. J Med Internet Res 2021 Dec 20;23(12):e28318 [FREE Full text] [doi: 10.2196/28318] [Medline: 34931996]
23. Abraham C, Sheeran P. The health belief model. In: Cambridge Handbook of Psychology, Health and Medicine. Cambridge, UK: Cambridge University Press; 2007:97-102.
24. Chen M, Land KC. Testing the health belief model: LISREL analysis of alternative models of causal relationships between health beliefs and preventive dental behavior. Soc Psychol Q 1986 Mar;49(1):45-60. [doi: 10.2307/2786856]
25. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
26. Yazdanpanah M, Forouzani M, Hojjati M. Willingness of Iranian young adults to eat organic foods: application of the health belief model. Food Qual Prefer 2015 Apr;41:75-83. [doi: 10.1016/j.foodqual.2014.11.012]
27. Lau J, Lim TZ, Wong GJ, Tan KK. The health belief model and colorectal cancer screening in the general population: A systematic review. Prev Med Rep 2020 Oct 6;20:101223 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101223] [Medline: 33088680]
28. Wong MC, Wong EL, Huang J, Cheung AW, Law K, Chong MK, et al. Acceptance of the COVID-19 vaccine based on the health belief model: A population-based survey in Hong Kong. Vaccine 2021 Feb 12;39(7):1148-1156 [FREE Full text] [doi: 10.1016/j.vaccine.2020.12.083] [Medline: 33461834]
29. O'Connor PJ, Martin B, Weeks CS, Ong L. Factors that influence young people's mental health help-seeking behaviour: a study based on the Health Belief Model. J Adv Nurs 2014 Nov 09;70(11):2577-2587. [doi: 10.1111/jan.12423] [Medline: 24720449]
30. Al-Sabbagh MQ, Al-Ani A, Mafrachi B, Siyam A, Isleem U, Massad FI, et al. Predictors of adherence with home quarantine during COVID-19 crisis: the case of health belief model. Psychol Health Med 2022 Jan 11;27(1):215-227. [doi: 10.1080/13548506.2021.1871770] [Medline: 33427487]

XSL•FO
RenderX

31. Hochbaum GM. Public Participation in Medical Screening Programs: A Socio-psychological Study. Washington, DC: US Dept of Health, Education, and Welfare; 1958:1-23.
32. Armitage CJ, Conner M. Social cognition models and health behaviour: A structured review. Psychol Health 2000 Mar;15(2):173-189. [doi: 10.1080/08870440008400299]
33. Harrison JA, Mullen PD, Green LW. A meta-analysis of studies of the Health Belief Model with adults. Health Educ Res 1992 Mar;7(1):107-116. [doi: 10.1093/her/7.1.107] [Medline: 10148735]
34. Ahadzadeh AS, Sharif SP, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
35. Mou J, Shin D, Cohen J. Health beliefs and the valence framework in health information seeking behaviors. ITP 2016 Nov 07;29(4):876-900. [doi: 10.1108/itp-06-2015-0140]
36. Ng B, Kankanhalli A, Xu Y. Studying users' computer security behavior: A health belief perspective. Decis Support Syst 2009 Mar;46(4):815-825. [doi: 10.1016/j.dss.2008.11.010]
37. Champion V, Skinner C. The health belief model. In: Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2008:45-65.
38. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022 Feb 16;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]
39. Wallston BS, Alagna SW, DeVellis BM, DeVellis RF. Social support and physical health. Health Psychology 1983;2(4):367-391. [doi: 10.1037/0278-6133.2.4.367]
40. Wellman B, Wortley S. Different strokes from different folks: community ties and social support. Am J Sociol 1990 Nov;96(3):558-588. [doi: 10.1086/229572]
41. Zhang Y, He D, Sang Y. Facebook as a platform for health information and communication: a case study of a diabetes group. J Med Syst 2013 Jun;37(3):9942. [doi: 10.1007/s10916-013-9942-7] [Medline: 23588823]
42. Stellefson M, Paige S, Apperson A, Spratt S. Social media content analysis of public diabetes Facebook groups. J Diabetes Sci Technol 2019 May 31;13(3):428-438 [FREE Full text] [doi: 10.1177/1932296819839099] [Medline: 30931593]
43. Wang J, Mann F, Lloyd-Evans B, Ma R, Johnson S. Associations between loneliness and perceived social support and outcomes of mental health problems: a systematic review. BMC Psychiatry 2018 May 29;18(1):156 [FREE Full text] [doi: 10.1186/s12888-018-1736-5] [Medline: 29843662]
44. Hale CJ, Hannum JW, Espelage DL. Social support and physical health: the importance of belonging. J Am Coll Health 2010 Aug 07;53(6):276-284. [doi: 10.3200/jach.53.6.276-284]
45. Gallant MP. The influence of social support on chronic illness self-management: a review and directions for research. Health Educ Behav 2003 Apr 01;30(2):170-195. [doi: 10.1177/1090198102251030] [Medline: 12693522]
46. Wang X, Parameswaran S, Bagul D, Kishore R. Can online social support be detrimental in stigmatized chronic diseases? A quadratic model of the effects of informational and emotional support on self-care behavior of HIV patients. J Am Med Inform Assoc 2018 Aug 01;25(8):931-944 [FREE Full text] [doi: 10.1093/jamia/ocy012] [Medline: 29684190]
47. McKinley CJ, Wright PJ. Informational social support and online health information seeking: Examining the association between factors contributing to healthy eating behavior. Comput Hum Behav 2014 Aug;37:107-116. [doi: 10.1016/j.chb.2014.04.023]
48. Deng Z, Liu S. Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. Int J Med Inform 2017 Sep;105:98-109. [doi: 10.1016/j.ijmedinf.2017.05.014] [Medline: 28750916]
49. Deetjen U, Powell J. Informational and emotional elements in online support groups: a Bayesian approach to large-scale content analysis. J Am Med Inform Assoc 2016 May;23(3):508-513. [doi: 10.1093/jamia/ocv190] [Medline: 26911815]
50. Savolainen R. Dietary blogs as sites of informational and emotional support. Information Research 2010;15(4):438 [FREE Full text]
51. Kim SC, Shah DV, Namkoong K, McTavish FM, Gustafson DH. Predictors of online health information seeking among women with breast cancer: the role of social support perception and emotional well-being. J Comput Mediat Commun 2013 Jan 25;18(2):98-118 [FREE Full text] [doi: 10.1111/jcc4.12002] [Medline: 24634575]
52. Han JY, Wise M, Kim E, Pingree R, Hawkins RP, Pingree S, et al. Factors associated with use of interactive cancer communication system: an application of the comprehensive model of information seeking. J Comput Mediat Commun 2010 Apr;15(3):367-388 [FREE Full text] [doi: 10.1111/j.1083-6101.2010.01508.x] [Medline: 21760702]
53. Guillory J, Niederdeppe J, Kim H, Pollak JP, Graham M, Olson C, et al. Does social support predict pregnant mothers' information seeking behaviors on an educational website? Matern Child Health J 2014 Nov 27;18(9):2218-2225 [FREE Full text] [doi: 10.1007/s10995-014-1471-6] [Medline: 24671467]
54. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Prom Int 2000;15(3):a. [doi: 10.1093/heapro/15.3.259]
55. Taggart J, Williams A, Dennis S, Newall A, Shortus T, Zwar N, et al. A systematic review of interventions in primary care to improve health literacy for chronic disease behavioral risk factors. BMC Fam Pract 2012 Jun 01;13(1):49 [FREE Full text] [doi: 10.1186/1471-2296-13-49] [Medline: 22656188]

XSL•FO
RenderX

56. Heijmans M, Waverijn G, Rademakers J, van der Vaart R, Rijken M. Functional, communicative and critical health literacy of chronic disease patients and their importance for self-management. Patient Educ Couns 2015 Jan;98(1):41-48. [doi: 10.1016/j.pec.2014.10.006] [Medline: 25455794]
57. Matsuoka S, Tsuchihashi-Makaya M, Kayane T, Yamada M, Wakabayashi R, Kato NP, et al. Health literacy is independently associated with self-care behavior in patients with heart failure. Patient Education and Counseling 2016 Jun;99(6):1026-1032. [doi: 10.1016/j.pec.2016.01.003]
58. Mayoh J, Bond CS, Todres L. An innovative mixed methods approach to studying the online health information seeking experiences of adults with chronic health conditions. J Mix Methods Res 2012;6(1):21-33. [doi: 10.1177/1558689811416942]
59. Li Y, Wang X. Seeking health information on social media: a perspective of trust, self-determination, and social support. JOEUC 2018;30(1):1-22. [doi: 10.4018/joeuc.2018010101]
60. Kahlor L. PRISM: a planned risk information seeking model. Health Commun 2010 Jun;25(4):345-356. [doi: 10.1080/10410231003775172] [Medline: 20512716]
61. Shang L, Zhou J, Zuo M. Understanding older adults' intention to share health information on social media: the role of health belief and information processing. INTR 2020 Jul 27;31(1):100-122. [doi: 10.1108/intr-12-2019-0512]
62. Ishikawa H, Takeuchi T, Yano E. Measuring functional, communicative, and critical health literacy among diabetic patients. Diabetes Care 2008 May;31(5):874-879. [doi: 10.2337/dc07-1932] [Medline: 18299446]
63. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. JMTP 2011;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]
64. Chin WW, Marcolin BL, Newsted PR. A partial least squares latent variable modeling approach for measuring interaction effects: results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. Inf Syst Res 2003 Jun;14(2):189-217. [doi: 10.1287/isre.14.2.189.16018]
65. Khan GF, Sarstedt M, Shiau W, Hair JF, Ringle CM, Fritze MP. Methodological research on partial least squares structural equation modeling (PLS-SEM). INTR 2019 Jun 03;29(3):407-429. [doi: 10.1108/intr-12-2017-0509]
66. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: algebra and statistics. J Mark Res 1981 Aug;18(3):382-388. [doi: 10.2307/3150980]
67. Chin W. The partial least squares approach to structural equation modeling. In: Modern Methods for Business Research. New York, NY: Psychology Press; 1998:295-336.
68. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39-50. [doi: 10.1177/002224378101800104]
69. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J of the Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
70. Yun E, Park H. Consumers' disease information-seeking behaviour on the internet in Korea. J Clin Nurs 2010 Oct;19(19-20):2860-2868. [doi: 10.1111/j.1365-2702.2009.03187.x] [Medline: 20598000]
71. Wilson TD. The diffusion of information behaviour research across disciplines. Information Research 2018;23(4):isic1801 [FREE Full text]
72. Chinn D. Critical health literacy: a review and critical analysis. Soc Sci Med 2011 Jul;73(1):60-67. [doi: 10.1016/j.socscimed.2011.04.004] [Medline: 21640456]

Abbreviations

CDC: Centers for Disease Control and Prevention
HBM: health belief model
HTMT: heterotrait-monotrait ratio
OHIS: online health information seeking
PLS: partial least square
RMB: renminbi
SEM: structural equation modeling

Edited by T Leung; submitted 05.09.22; peer-reviewed by Y Zhao, M Morsa, X Han; comments to author 29.09.22; revised version received 06.10.22; accepted 20.10.22; published 02.11.22.

Please cite as:
Zhao YC, Zhao M, Song S
Online Health Information Seeking Among Patients With Chronic Conditions: Integrating the Health Belief Model and Social Support Theory
J Med Internet Res 2022;24(11):e42447
URL: https://www.jmir.org/2022/11/e42447
doi:10.2196/42447
PMID:36322124

XSL•FO
RenderX

©Yuxiang Chris Zhao, Mengyuan Zhao, Shijie Song. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Measuring the Digital Skills of Catalan Health Care Professionals as a Key Step Toward a Strategic Training Plan: Digital Competence Test Validation Study

Elisenda Reixach[1], PhD; Erik Andrés[1], BSc; Josuè Sallent Ribes[1], PhD; Montserrat Gea-Sánchez[2], PhD; Alícia Àvila López[2], BSc; Bea Cruañas[2], BSc; Anna González Abad[3], BSc; Ricard Faura[4], BSc; Montse Guitert[5], PhD; Teresa Romeu[5], PhD; Eulàlia Hernández-Encuentra[5], PhD; Sandra Bravo-Ramirez[5], BSc; Francesc Saigí-Rubió[6], PhD

[1]Fundació TIC Salut i Social, Generalitat de Catalunya, Barcelona, Spain

[2]Departament de Salut, Generalitat de Catalunya, Barcelona, Spain

[3]Servei Català de Salut, Barcelona, Spain

[4]Servei d'Inclusió i Capacitació Digital, Departament de la Vicepresidència i de Polítiques Digitals i Territori, Generalitat de Catalunya, Barcelona, Spain

[5]Faculty of Psychology and Education Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

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

Corresponding Author:
Francesc Saigí-Rubió, PhD
Faculty of Health Sciences
Universitat Oberta de Catalunya
Rambla del Poblenou, 156
Barcelona, 08018
Spain
Phone: 34 933 263 622
Email: fsaigi@uoc.edu

Abstract

Background: Despite Catalonia being an advanced region in terms of digital health adoption, the "Forum for Professional Dialogue" identified the need to improve information and communication technology (ICT) competences as one of the present and future challenges for health care professionals (HPs).

Objective: We aimed to validate the digital competence test developed ad hoc for this study and to measure the digital competence level of Catalan HPs to establish their current level as the baseline for designing a strategic training plan.

Methods: An exploratory observational study was conducted based on a voluntary survey where sociodemographic, professional and digital tool knowledge, digital tool use, and training needs data were collected and based on the score obtained from a digital competence test developed ad hoc. The digital competence test consisted of 2 "real-life scenarios" with 7 and 11 questions.

Results: In total, 803 HPs, of whom 612 (76.2%) were women, completed the survey between June 28 and July 16, 2021. Most participants self-rated their digital competence level as either intermediate (384/803, 47.8%) or basic (357/803, 44.5%). The mean score in the digital competence test was 22.6 (SD 4.3). Therefore, most participants displayed a basic level of digital competence. The internal consistency of the digital competence test was 0.66, and the discrimination index of all questions was ≥0.2 for all items except for 1 question.

Conclusions: This exploratory study highlights the need to improve the digital competence of HPs working in Catalonia, with special effort being made to provide training according to the specific needs of the different HP profiles. The results have informed the Health Plan for Catalonia 2021-2025 and lay the foundations for the development and deployment of a framework program for the digital competences of HPs. The developed digital competence test shows acceptable consistency for the objective pursued, although improvements are needed to fine-tune its accuracy.

(J Med Internet Res 2022;24(11):e38347) doi:10.2196/38347

KEYWORDS

digital health; eHealth; digital competences; digital literacy; information and communication technology; ICT; training

XSL•FO
RenderX

Introduction

Background

Digital health (eHealth) is changing the way prevention, diagnosis, treatment, and health monitoring are provided to patients [1,2], while allowing universal access to equal, qualified, and cost-effective health care [3-5]. However, unlocking the full potential of eHealth is only possible when all actors (health care professionals [HPs], patients, managers, and policy makers) are committed to accepting and adopting information and communication technologies (ICTs) as a different way of providing or receiving care. Embracing the digital culture and developing professionals' digital skills or competences to support digital transformation in the health care sector is fundamental to achieving this objective [6], because poor digital health competence is a common perceived barrier to the implementation of eHealth services [7-11].

Catalonia (Northeast Spain) is one of the most advanced regions in terms of digital health adoption across Europe [8,12,13]. Despite this, in 2018, the "Forum for Professional Dialogue" identified "the need to improve competences in ICTs to advance in their use and in the design of remote healthcare services" as one of the 17 main present and future challenges for HPs [14]. The "Digital Skills for HPs (COMPDIG-Salut)" project arose with the aim of meeting this challenge by addressing three objectives: (1) defining a specific digital competence framework for HPs; (2) creating a specific evaluation and accreditation model for HPs; and (3) drawing up actions to train and qualify HPs in digital competences. Therefore, knowing the current digital competence level of Catalan HPs is an essential first step upon which to build and address all COMPDIG-Salut project goals [15,16].

Despite the availability of a wide variety of free self-efficacy, knowledge-based, and performance-based digital competence assessments [16-22], these instruments were found to be too long, unvalidated, or too specific to measure HPs' digital literacy levels. In 2009, the Government of Catalonia launched the Accreditation of Competence in ICTs (ACTIC) certificate with the aim of assessing citizens' digital competence [23] on the basis of 3 levels: ACTIC 1-basic, ACTIC 2-intermediate, and ACTIC 3-advanced. The competences evaluated by this accreditation have been updated over the years and are currently aligned with the European Digital Competence Framework [24]. Currently, the attainment of the ACTIC 2-intermediate level certificate is a way of improving the professional development of HPs and the employability of graduates. However, the ACTIC intermediate certificate requires too much time and too many resources to complete; therefore, it was deemed unsuitable for the purposes of our research. Therefore, to assess the digital competences of HPs working in Catalonia, a digital competence test based on the ACTIC 2-intermediate certificate was developed ad hoc.

Objective

The objective of this work is two-fold: (1) to validate the digital competence test developed ad hoc, which combines both skills and self-assessment, and (2) to assess the current digital competence level of HPs working in Catalonia and to identify the areas needing improvement.

Methods

Study Design

An exploratory, observational study based on a web-based survey was conducted by Fundació TIC Salut i Social between June 28 and July 16, 2021, among HPs currently working in Catalonia. By law [25], the definition of HPs, who make up the study population, includes dentists, dental hygienists, dental technicians, dietitians-nutritionists, occupational therapists, nurses, opticians or optometrists, pharmacists, physicians, physiotherapists, podiatrists, speech therapists, and other health or clinical specialists such as biologists, physicists or chemists, and psychologists. According to the last available report (2017) [14], the HP population in Catalonia consisted of 121,039 professionals working for public and private health care providers. Participation was both voluntary and anonymous.

The survey was conducted, and the results were reported in accordance with "Good Practice in the conduct and reporting of survey research" where appropriate [26] and in compliance with the General Data Protection Regulation.

Survey Generation and Distribution

The survey included a section recording the participants' characteristics ("descriptive survey") and an objective digital competence test (Multimedia Appendix 1). The first section recorded demographic characteristics such as gender, age, professional profile, ownership of workplace (public or private [subsidized or nonsubsidized]), region, level of health care, professional experience, self-perception of digital competence, the use of digital tools for professional purposes, the need for training in digital tools for professional purposes, interest in receiving such training for personal purposes, and whether the participant had the ACTIC 2-intermediate certificate or an equivalent qualification.

The second section was the objective digital competence test. The ACTIC 2-intermediate level certificate served as the framework of reference to evaluate the digital competence level of HPs. The most appropriate and relevant digital competences for HPs were selected from those defined in ACTIC [27] to adapt the test to their reality and context (Table 1). Then, the indicators for each of these competences were selected to determine, in items, what aspects to evaluate. The indicators are observable characteristics and consist of specific tests, be they predefined measures or other types of qualitative information. After selecting the indicators, questions referring to them were formulated. These questions are related to the definition of observable behaviors that may be put into practice in different professional areas in the Catalan health care context. Observable behaviors are understood to be those practices or actions carried out by HPs during their professional activities (eg, searching for clinical information in databases, remote communication and collaboration with teams or patients, and using information management and content creation tools). As it was the achievement of competences that was being evaluated, the suitability of defining evaluation scenarios allowing

XSL•FO
RenderX

respondents to be faced with challenges they needed to resolve was assessed. Attempting to resolve situations that are similar to real ones and giving the best digital response to the proposed challenges allows the degree of achievement of the indicators to be evaluated more effectively. It also enables other competences to be put into practice, such as problem-solving, critical thinking, and the analysis and responsible use of ICTs.

After formulating the questions for each indicator, it was necessary to close the loop by reviewing the entire process using a methodological process whereby the specification of new elements improves upon previous ones [28,29]. Test development followed an iterative process of expert consultation, prior pilot tests, and item review. For cross-validation, 8 members of the COMPDIG-Salut project were asked to evaluate the overall proposal and specifically whether the questions contemplated the defined indicators for the corresponding competence. The experts had to answer the survey questions and suggest any changes they deemed relevant by answering an open-ended question.

The digital competence test that we developed consisted of 2 “real-life” scenarios adapted to the health care sector with 7 and 11 self-developed questions. Each question had 4 possible answers, with a score ranging from 0 to 1 or 2, resulting in a total maximum score of 35. The final scores were classified into 3 levels: initial (2-9.9), basic (10-24.9), and intermediate (25-35). The survey took approximately 20 minutes to complete.

XSL•FO
RenderX

Table 1. Development of survey questions by competence.

Accreditation of Competence in Information and Communication Technologies competence and intermediate level indicator	Survey
1.1 Searching for, selecting, and comparing information with digital tools	
The respondent uses advanced search parameters (language, update, publication date, region, Boolean operators, etc) using an assistant or different menus and options to optimize searches and readjust search criteria.	Case 1: Question 1
The respondent is critical of the information and recognizes the limits of the internet as a single source.	Case 1: Question 2
1.2 Organizing information and data with digital tools	
The respondent structures and classifies data coherently and accessibly using generic tools (spreadsheet or database) and specific tools (bookmark manager, contact manager, expense tracker, etc) to make searching easier.	Case 1: Question 6
1.3 Analyzing, exploiting, and visualizing data with digital tools	
The respondent gathers data with digital tools (forms, surveys, etc) for specific objectives.	Case 1: Question 5
The respondent uses and combines formulas and functions to perform simple operations.	Case 2: Question 7
2.1 Interacting and sharing information and digital content	
The respondent acts as an example of communication for the rest of the digital community according to the context and the tool used.	Case 1: Question 4
2.2 Collaborating with others via digital technologies	
The respondent identifies and uses digital tools, resources, and strategies to improve efficiency in the performance of tasks in collaboration with others.	Case 2: Question 1
The respondent interacts by using the most appropriate (synchronous or asynchronous) communication tools and their advanced functions (user groups, broadcast groups, mailing groups, distribution lists, web-based meetings, etc) effectively.	Case 2: Questions 3 and 4
3.1 Creating and publishing digital content	
The respondent selects and evaluates the most appropriate resources and applications for the objective pursued and the type of digital content to optimize creations.	Case 1: Question 7
3.2 Designing, integrating, and reworking digital content in various formats	
The respondent creates and publishes complex content suited to the audiences, objectives, or purposes thereof, seeking the most appropriate content in each case (using a template).	Case 1: Question 3
The respondent uses and evaluates repositories of audiovisual content (images, audio recordings, GIFs[a], videos, templates, etc) to design or rework digital content.	Case 2: Question 5
4.1 Protecting digital systems, devices, and content	
The respondent protects digital files and content to prevent unauthorized access by third parties.	Case 2: Question 9
4.2 Protecting personal data and privacy	
The respondent identifies critical points and suggests improvements for protecting personal data and privacy. Applying best practices, the respondent customizes the privacy settings of digital tools and environments and the permissions of apps to protect their identity and the privacy of the content they generate.	Case 2: Question 10
4.3 Acting in a civic manner in the digital environment	
The respondent uses licenses and attribution systems suited to their objectives when publishing content in digital environments.	Case 2: Question 2
The respondent promotes coexistence in the digital environment.	Case 2: Question 8
5.1 Understanding the basics and using digital technology	
The respondent uses digital technology and its environment autonomously.	Case 2: Question 11
5.2 Identifying personal and professional needs and applying digital solutions	
The respondent is up to date with the latest technological trends and compares and evaluates digital devices and tools to select those that best meet their personal needs (leisure, health, protection, sports, emotional, etc) and professional needs (training, job search, productivity, and time management).	Case 2: Question 6

[a]GIF: Graphics Interchange Format.

Data Collection

The survey was created and distributed to HPs using Microsoft Forms, together with an invitation email presenting the objective and characteristics of the study. Several meetings were held with HP associations and health service providers to explain our intention to conduct this study and that we would need their help to reach HPs. The invitation was distributed through the

XSL•FO
RenderX

Catalan Department of Health and the Catalan Health Service, which sent it to the corresponding professional associations and to the human resources departments of public and private health care providers (hospitals, consortia, etc). Each institution decided how to disseminate the study among the professionals.

Before participating in the study, the participants had to provide consent for the study sponsor to process the information collected. The time taken to complete the survey (in minutes) was recorded.

Sample Size

The sample size was calculated so that it would be representative of the population of Catalan HPs (both in size and distribution). Of a total of 121,039 HPs [30], 36,520 (30.17%) were physicians, 45,995 (38%) were nurses, and 38,524 (31.83%) were other HPs. In 2019, 96,105/121,039 (79.40%) and 11,014/121,039 (9.10%) of these HPs worked in the Barcelona and Girona health care regions, respectively. The remaining (13,920/121,039, 11.50%) worked in other health care regions [14,30]. The minimum and maximum sample sizes were calculated using the Cochran formula, considering a 95% power and a 10% and 5% margin of error, respectively. This resulted in a sample size of 304 to 906 and 34 to 101 HPs in the Barcelona and Girona health care regions, respectively, and 45 to 134 HPs in other health care regions. The minimum and maximum overall sample sizes were in the range of 383 to 1141 HPs.

Statistical Analysis

Percentages and mean (SD) were used to summarize categorical and continuous variables, respectively. Categorical variables were compared using the chi-square test, whereas continuous variables were compared using the *t* test or ANOVA (for 2 or >2 comparators, respectively). These comparisons were 2-sided. The analysis of subgroups with a score of <25 in the digital competence test was performed using a 1-sided *t* test. Given that no maximum time to complete the survey was established and that the filling in of fields could be interrupted for personal or professional reasons, atypical observations where this was likely to have happened were removed to estimate a more realistic mean completion time. Outliers were filtered using the Hampel identifier. The internal consistency of the survey was analyzed using the greatest lower bound, given the lack of homogeneity of the scoring scale. The discriminatory index of each question of the ACTIC-derived digital competence test was calculated according to the study by Taib et al [18], where a score of ≥0.2 indicated good discrimination between HPs with intermediate and basic levels. Statistical analyses were performed using the R (version 4.11; R Foundation for Statistical Computing) software. A *P* value of <.05 was considered significant.

Ethics Approval

No ethics approval was required due to the type and nature of the study as the Catalan Department of Health is responsible for formulating the general criteria for health planning, setting the objectives, and the levels to be achieved in the topics that are included in the Health Plan for Catalonia [31]. All participants were informed about the study's purposes and that their participation was voluntary. Data protection treatment was informed to the participant and before accessing the survey, participants had to provide acceptance.

Results

Between June 28 and July 16, 2021, a total of 1009 potential participants accessed the survey, of whom 922 (91.40%) gave their consent to participating in it. Of these, 803 (79.6%) participants were classified as HPs according to the legal definition and constituted the study population.

Descriptive Analysis

Table 2 shows the demographic and professional characteristics of the participants, most of whom were women (612/803, 76.2%), aged between 36 and 55 years (438/803, 54.5%). Nursing was the most common professional profile (227/803, 28.3%), followed by physicians (176/803, 21.9%). Nearly half of the participants worked in a subsidized private center. A total of (478/803, 59.5%) participants worked in specialized health care settings. Barcelona city was the most common work setting (209/803, 26%), followed by Camp de Tarragona (148/803, 18.4%). The mean length of professional experience was 19.6 (SD 11) years.

XSL•FO
RenderX

Table 2. Demographic and professional characteristics of participants (N=803).

Variables	Values
Gender, n (%)	
Women	612 (76.2)
Men	188 (23.4)
Nonbinary	3 (0.4)
Age (years), n (%)	
18-25	24 (3)
26-35	169 (21)
36-45	258 (32.1)
46-55	180 (22.4)
56-65	155 (19.3)
>65	17 (2.1)
Health care professional profiles, n (%)	
Nurse	227 (28.3)
Physician	176 (21.9)
Physiotherapist	80 (10)
Occupational therapist	74 (9.2)
Podiatrist	58 (7.2)
Dietitian-nutritionist	49 (6.1)
Speech therapist	38 (4.7)
Psychologist	35 (4.4)
Pharmacist	32 (4)
Biologist	15 (1.9)
Dental hygienist	12 (1.5)
Physicist or chemist	3 (0.4)
Optician-optometrist	2 (0.2)
Dentist	1 (0.1)
Dental technician	1 (0.1)
Workplace ownership, n (%)	
Public	257 (32)
Private (subsidized)	365 (45.5)
Private (nonsubsidized)	170 (21.2)
Do not know or no answer	11 (1.4)
Level of health care,[a] n (%)	
Mental health and addictions	67 (8.3)
Hospital or specialized care	478 (59.5)
Primary care	199 (24.8)
Social health	125 (15.6)
Work setting,[b] n (%)	
Alt Pirineu i Aran	26 (3.2)
Barcelona city	209 (26)
Camp de Tarragona	148 (18.4)

XSL•FO
RenderX

Variables	Values
Catalunya Central	59 (7.3)
Girona	67 (8.3)
Lleida	42 (5.2)
Metropolitan (north)	119 (14.8)
Metropolitan (south)	73 (9)
Terres de l'Ebre	53 (6.6)
Do not know or no answer	7 (0.9)
Professional experience (years), mean (SD)	19.6 (11)

[a]Multiple responses were allowed.

[b]Catalan health regions.

Table 3 shows the information collected on digital competence and on the use of, training needs for, and interest in digital tools. Most participants self-rated their digital competence as intermediate (384/803, 47.8%) or basic (357/803, 44.5%). Nearly half of the participants (394/803, 49.1%) did not know about ACTIC certification. Office tools (Microsoft Office, email, etc) were the most frequently used professional digital tools (750/803, 93.4%), followed by social media (700/803, 87.1%) and electronic health records (574/803, 71.5%). The most in-demand training topics were tools for disease prevention and health promotion (383/803, 47.7%), office tools (358/803, 44.6%), electronic health records (346/803, 43.1%), remote follow-up tools (300/803, 37.4%), and decision-making support tools (247/803, 30.8%). In relation to the use of digital tools for professional purposes and training needs, expressed for the largest groups of HPs in this study, we found that the tools that nurses (227/803, 28.3%) used the most (>30%) were office tools (206/227, 90.7%), social networks (198/227, 87.2%), electronic health records (192/227, 84.6%), healing support tools (83/227, 36.6%) and health promotion tools (80/227, 35.2%). Of these, the ones with more training needs were health promotion tools (106/227, 46.7%), office tools (94/227, 41.4%), electronic health records (93/227, 41%), healing support tools (91/227, 40.1%), and social networks (41/227, 18.1%). For physicians (176/803, 21.9%), we found that office tools (169/176, 96%), electronic health records (161/176, 91.5%), social networks (154/176, 87.5%), prescription tools (152/176, 86.4%), remote follow-up tools (68/176, 38.6%), epidemiological register tools (63/176, 35.8%), decision-making support tools (58/176, 33%), and health promotion tools (55/176, 31.3%) were the most used tools. Of these, the ones with more training needs were decision-marking support tools (81/176, 46%), office tools (78/176, 44.3%), electronic health records (77/176, 43.8%), remote follow-up tools (74/176, 42%), health promotion tools (69/176, 39%), prescription tools (58/176, 33%), epidemiological register tools (48/176, 27.3%), and social networks (40/176, 22.7%). As for physiotherapists (80/803, 10%), the most used tools were office tools (74/80, 93%), social networks (69/80, 86%), electronic health records (49/80, 61%), and health promotion tools (31/80, 39%), whereas health promotion tools (45/80, 39%), office tools (37/80, 46%), electronic health records (37/80, 46%), and social networks (36/80, 45%) were the most in-demand training topics. For occupational therapists (74/803, 9.2%), we found that (72/74, 97%), social networks (59/74, 80%), and electronic health records (49/74, 66%) were the most used tools, whereas office tools (30/74, 41%), electronic health records (30/74, 41%), and social networks (18/74, 24%) were the most in-demand training topics. As for podiatrists (58/803, 7.2%), the most used tools were social networks (53/58, 91%), office tools (51/58, 88%), electronic health records (29/58, 50%), and prescription tools (19/58, 33%), whereas electronic health records (25/58, 43%), prescription tools (20/58, 35%), office tools (16/58, 28%), and social networks (15/58, 26%) were the most in-demand training topics. Moreover, for dietitian-nutritionists (49/803, 6.1%), we found that social networks (46/49, 94%), office tools (44/49, 90%), electronic health records (26/49, 53%), health promotion tools (21/49, 43%), and remote follow-up tools (17/49, 35%) were the most used tools, whereas health promotion tools (34/49, 69%), electronic health records (24/49, 49%), office tools (23/49, 47%), and social networks (20/49, 41%) were the most in-demand training topics.

Disaggregated information relating to "Others" participants (biologists and dietitians or nutritionists, pharmacists, physiotherapists, dental hygienists, speech therapists, podiatrists, psychologists, and occupational therapists) can be found in Multimedia Appendix 2.

The greatest interest in receiving training in digital tools for personal purposes was for presentation of digital content (426/803, 53.1%); data management (408/803, 50.1%); digital technology, computer, and operating system use (344/803, 42.8%); and processing of graphic, audio, and video information (341/803, 42.5%; Table 3).

XSL•FO
RenderX

Table 3. Participants' digital competences and use of, training needs for, and interest in digital tools (N=803).

Variables	Overall, n (%)	Physicians, n (%)	Nurses, n (%)	Others, n (%)
Self-perceived digital competence				
Advanced	54 (6.7)	14 (8)	14 (6.2)	26 (6.5)
Intermediate	384 (47.8)	80 (45.5)	104 (45.8)	200 (50)
Basic	357 (44.5)	81 (46)	107 (47.1)	169 (42.3)
No digital competence	8 (0.1)	1 (0.6)	2 (0.9)	5 (1.3)
Accreditation of Competence in Information and Communication Technologies-2 certificate or similar				
Yes	57 (7.1)	3 (1.7)	32 (14.1)	22 (5.5)
No	352 (43.8)	70 (39.8)	109 (48)	173 (42.3)
I do not know about the ACTIC certificate	394 (49.1)	103 (58.5)	86 (37.9)	205 (51.3)
Use of digital tools for professional purposes[a,b]				
Office tools (Microsoft Office, email, etc)	750 (93.4)	169 (96)	206 (90.7)	372 (93)
Social media	700 (87.1)	154 (87.5)	198 (87.2)	348 (87)
Electronic health records	574 (71.5)	161 (91.5)	192 (84.6)	218 (54.5)
Prescription tools	265 (33)	152 (86.4)	53 (23.3)	60 (15)
Health promotion tools	246 (30.6)	55 (31.3)	80 (35.2)	111 (27.8)
Remote follow-up of patients	219 (27.3)	68 (38.6)	50 (22)	101 (25.3)
Decision-making support tools	159 (19.8)	58 (33)	60 (26.4)	41 (10.3)
Training needs for professional purposes[a,b]				
Health promotion tools	383 (47.7)	69 (39.2)	106 (46.7)	205 (51.3)
Office tools (Microsoft Office, email, etc)	358 (44.6)	78 (44.3)	94 (41.4)	181 (45.3)
Electronic health records	346 (43.1)	77 (43.8)	93 (41)	172 (43)
Remote follow-up of patients	300 (37.4)	74 (42)	67 (29.5)	156 (39)
Decision-making support tools	247 (30.8)	81 (46)	61 (26.9)	102 (25.5)
Prescription tools	215 (26.8)	58 (33)	82 (36.1)	72 (18)
Social networks	212 (26.4)	40 (22.7)	41 (18.1)	132 (33)
Diagnostic support tools	196 (24.4)	67 (38.1)	37 (16.3)	89 (22.3)
Bioinformatic (Omics) tools	172 (21.4)	55 (31.3)	49 (21.6)	65 (16.3)
Epidemiological register tools	141 (17.6)	48 (27.3)	53 (23.3)	54 (13.5)
Healing support tools	158 (19.7)	18 (10.2)	91 (40.1)	29 (7.3)
Personal interest in digital training[a,b]				
Digital content presentation	426 (53.1)	81 (46)	126 (55.5)	219 (54.8)
Data management	408 (50.1)	97 (55.1)	113 (49.8)	198 (49.5)
Digital technology, computer, and operating system use	344 (42.8)	75 (42.6)	113 (49.8)	156 (39)
Web browsing and digital communication	344 (42.8)	78 (44.3)	105 (46.3)	161 (40.3)
Graphic, audio, and video information management	341 (42.5)	82 (46.6)	99 (43.6)	160 (40)
Written information management	301 (37.5)	59 (33.5)	88 (38.8)	154 (38.5)
Numeric information management	235 (29.3)	57 (32.4)	72 (31.7)	106 (26.5)
Culture, participation, and citizenship	244 (27.9)	40 (22.7)	70 (30.8)	114 (28.5)

[a]Multiple responses were allowed.

[b]Only those options to which >15% of participants responded are shown.

XSL•FO
RenderX

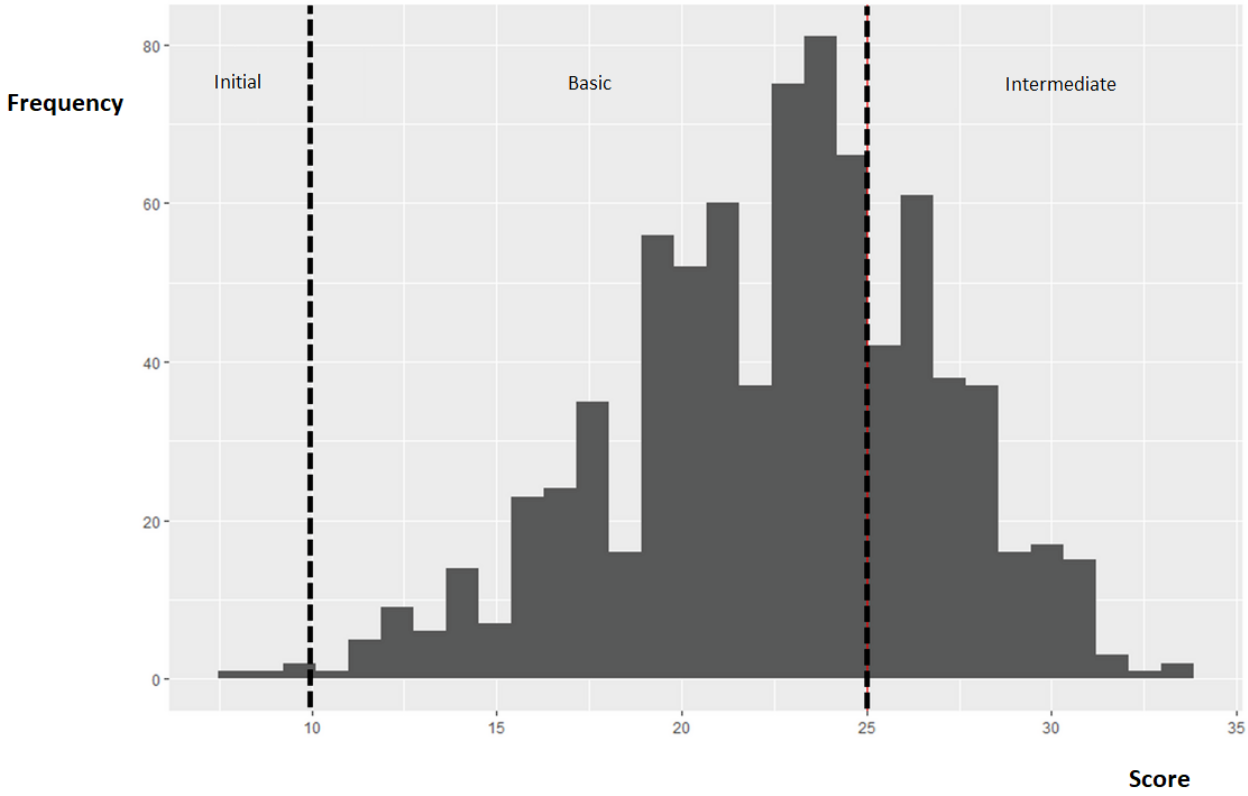
Digital Competence Test

The mean score in the digital competence test was 22.6 (SD 4.3), so most participants displayed a basic level (Table 4 and Figure 1). After removing outliers (122/803, 15%; see the *Methods* section), the mean time taken to complete the survey was 15.7 (SD 5.8) minutes.

Table 4. Digital competence test results (N=803).

Variable	Overall population	Accreditation of Competence in Information and Communication Technologies-2 certificate holders (n=57)
Score, mean (SD)	22.6 (4.3)	23.4 (3.9)
Score range, n (%)		
<10 (initial level)	2 (0.2)	0 (0)
10-24.9 (basic level)	535 (66.7)	34 (59.6)
≥25 (intermediate level)	266 (33.1)	23 (40.4)
Time to complete the survey, (minutes), mean (SD)	15.7 (5.8)	11.8 (3.6)

Figure 1. Distribution of scores achieved in the digital competence test.



The test score was higher in men (mean 23.2, SD 4.2, vs mean 22.4, SD 4.3 in women; *P*=.03), in younger participants (*P*<.001), and those with a high self-perceived level of digital competence (*P*<.001). We also observed differences among the various HP profiles, with other HP profiles scoring higher than nurses and physicians. Most of the analyzed subgroups had scores that were significantly <25 (Table 5).

A multivariate linear regression analysis confirmed the significant differences found in age, HP profile, and self-perception, but this was not confirmed for the gender variable.

In total, 7.1% (57/803) of participants had the ACTIC-2 certificate. Their mean score in the digital competence test was 23.4, SD 3.9), with 59.6% (34/57) having a basic level. The mean time taken to complete the survey was 11.8 (SD 3.6) minutes (Table 4). Of those 57 participants, 13 (22.8%) said their self-perceived knowledge was advanced, and 33 (57.9%) said their self-perceived knowledge was intermediate.

The internal consistency of the digital competence test as measured by greatest lower bound was 0.66 (acceptable consistency). The discrimination index of all questions except for one for all items was ≥0.2.

XSL•FO
RenderX

Table 5. Scores achieved in the digital competence test according to participants' characteristics (N=803)[a].

Characteristics	Value, n (%)	Score, mean (SD)	*P* value	Intermediate level achievement, *P* value
Gender			.03	
Women	612 (76.2)	22.4 (4.3)		<.001
Men	188 (23.4)	23.2 (4.2)		<.001
Age (years)			<.001	
18-25	24 (3)	24.6 (3.2)		.28
26-35	169 (21.1)	23.5 (4)		<.001
36-45	258 (32.1)	22.6 (4.7)		<.001
46-55	180 (22.4)	22.5 (4.1)		<.001
56-65	155 (19.3)	21.6 (4.1)		<.001
>65	17 (2.1)	21.5 (4.9)		.004
Health care professional profiles			<.001	
Nurse	227 (28.3)	21.7 (4.2)		<.001
Physician	176 (21.9)	22.5 (4.2)		<.001
Other	400 (49.8)	23.1 (4.4)		<.001
Biologist	15 (1.9)	23.3 (5.1)		.10
Dietitian-nutritionist	49 (6.1)	24 (4.3)		.05
Pharmacist	32 (4)	23.6 (4.9)		.06
Physiotherapist	80 (10)	22.6 (4)		<.001
Dental hygienist	12 (1.5)	20 (5)		.003
Speech therapist	38 (4.7)	24.6 (3.9)		.24
Podiatrist	58 (7.2)	22.3 (4.6)		<.001
Psychologist	35 (4.4)	23.8 (5.1)		.09
Occupational therapist	74 (9.2)	23 (4)		<.001
Workplace ownership			.40	
Public	257 (32)	22.5 (4.4)		<.001
Private (subsidized)	365 (45.5)	22.5 (4.3)		<.001
Private (nonsubsidized)	170 (21.2)	23.1 (4.3)		<.001
Level of health care			.24	
Mental health and addictions	67 (8.3)	23.5 (3.9)		.001
Hospital or specialized care	478 (59.5)	22.5 (4.4)		<.001
Primary care	199 (24.8)	22.5 (4.3)		<.001
Social health	125 (15.6)	23 (4.4)		<.001
Work setting			.08	
Alt Pirineu i Aran	26 (3.2)	21.8 (2.9)		<.001
Barcelona city	209 (26)	23.5 (4.3)		<.001
Camp de Tarragona	148 (18.4)	22 (4.2)		<.001
Catalunya Central	59 (7.4)	22.8 (4.5)		<.001
Girona	67 (8.3)	22.3 (4.5)		<.001
Lleida	42 (5.2)	21.7 (4.4)		<.001
Metropolitan (north)	119 (14.8)	22.6 (4.2)		<.001
Metropolitan (south)	73 (9.1)	23 (5.1)		<.001
Terres de l'Ebre	53 (6.6)	22 (3.9)		<.001

XSL•FO
RenderX

Characteristics	Value, n (%)	Score, mean (SD)	*P* value	Intermediate level achievement, *P* value
Self-perceived digital competence			<.001	
Advanced	54 (6.7)	24.6 (3.8)		.21
Intermediate	384 (47.8)	23.6 (3.9)		<.001
Basic	357 (44.5)	21.4 (4.3)		<.001
No digital competence	8 (1)	16.1 (6.3)		.002
Accreditation of Competence in Information and Communication Technologies-2 certificate or similar			.11	
Yes	57 (7.1)	23.4 (3.9)		.002
No, I do not know about the Accreditation of Competence in Information and Communication Technologies certificate	746 (92.9)	22.5 (4.4)		<.001

[a]Small samples were not analyzed.

Discussion

Principal Findings

Globally, there is a need to improve the digital skills of HPs through dedicated training to fully exploit the potential of digital technologies and to be able to provide the best possible care using such technologies [3,11,32-34]. Multiple initiatives have been implemented to address this need [35]. For example, the EU*US eHealth Work Project devoted its efforts to identifying gaps and defining competences and developed a free introductory web-based course in eHealth [36]. Another major effort is that of Health Education England, which had defined a digital capability framework for improving the digital literacy of the health and care workforce and is currently testing a tool to self-assess digital literacy [37].

Our exploratory study in a large cohort representative of the Catalan HP population has provided valuable information regarding their digital competence level and training needs and revealed the consistency of the ad hoc digital competence test for the pursued objective. Nearly all participants (801/803, 99.8%) had either the basic or the intermediate level of competence, indicating that this HP group is already fulfilling the objectives set by the European Skills Agenda [38], whose aim is to ensure that at least 70% of adults have basic digital skills by 2025. However, the results should be interpreted with caution given the voluntary nature of participation, which might have given rise to a bias in assessing the participants' digital competence.

Some of the findings deserve further discussion. We only found statistically significant differences in the level of digital competence by age (higher in younger ages), self-rated digital competence (higher in those rated as advanced), and professional profile (higher in HP profiles other than nurses and physicians). Conversely, we found no differences based on workplace ownership (public vs private), level of health care, or work setting, which points to a homogeneous population in this sense. Most of the analyzed subgroups obtained an overall score that was statistically <25 (basic level). Exceptions to this included the youngest participants (aged 18-25 years); those with an advanced level of self-reported digital competence; and certain HP profiles such as biologists, dietitians or nutritionists, pharmacists, speech therapists, and clinical psychologists. Although it cannot be ruled out that these professional profiles may be more highly trained in digital competences as a result of their profession, the small size of these populations may have contributed to a biased result.

Our study also provides valuable information on the most frequently used professional digital tools (Office tools, social media, and electronic health records) and the most requested type of training (tools for disease prevention and health promotion, Office tools, electronic health records, and remote follow-up tools). The training needs were found to be broader than the nature of the main tools used, and these needs were inversely proportional to the frequency of use of specific digital tools. Both findings are indicative of HPs' interest in digital competence, which goes beyond the limitations imposed by their current professional digital skills or even by their current clinical practice. The results of an in-depth analysis of the most frequent HP profiles revealed interesting differences in the use of digital tools and training needs, with physicians displaying a broader range in terms of digital tools used and training needs [39].

Interestingly, only 7.1% (57/803) of the participants held an ACTIC-2 certificate, accrediting an intermediate level. This finding has several possible interpretations. On the one hand, the ACTIC certificate is voluntary and was conceived as a standard tool to prove citizens' digital competences when applying for a job or job promotion. This may explain its reduced representation in our sample and the fact that nearly 50% (394/803) of the HPs participating in our study were not aware of its existence. Not holding an ACTIC-2 certificate only denotes that the participants had not had access to this voluntary certificate. This may explain the similar scores observed between the participants with or without ACTIC-2. Of the 6.7% (54/803) of participants who rated their digital competence as advanced, only 24% (13/55) had the ACTIC-2 certificate. Objectively speaking, holding this certificate translated into a higher rate of participants scoring ≥25 (intermediate level: 23/57, 40%, vs 266/803, 33.1%) and lesser time taken to complete the survey (mean 11.8, SD 3.6 minutes vs mean 15.7, SD 5.8 minutes).

XSL•FO
RenderX

However, the statistical significance of these differences was not analyzed.

The findings of our study are highly valuable to the COMPDIG-Salut project as it establishes the basis for planning and deciding on specific strategic actions and policies to improve the digital competences of Catalan HPs. These are likely to be similar to other HPs in other EU countries. Furthermore, the results revealed the HPs' training needs from both professional and personal perspectives, which may serve as a starting point for designing tailored training actions.

Comparison with similar studies is hindered by differences in the populations included, the methodology, and the questionnaires used. Further analyses through an in-depth examination of the answers obtained for each of the formulated questions are currently underway to identify specific competence areas and on which training should be primarily focused.

Strengths and Limitations

Our study has several strengths, including a large sample size and sociodemographic representativeness of the main HP profiles in Catalonia. The distribution of the HPs' professional profiles was skewed with respect to the real Catalan scenario, with profiles of HPs other than nurses and physicians being overrepresented.

One of the novelties of our study is the way in which digital competence is assessed, that is, subjectively by self-efficacy (the most common method) and objectively by an ad hoc survey. Although the digital competency tool is yet to be fully validated, its acceptable consistency reinforces our findings and supports its validity for use in evaluating the digital competence level in the research setting. Owing to its approach, no barriers were anticipated for its local adaptation. Moreover, all the questions answered by the participants, except one, showed sufficient discriminatory power. However, given that the tool is a generic one aimed at all HPs, the adaptation of assessment scenarios and activities to different HP profiles will enable further refinement of the results. Resolving these weaknesses in the tool will improve its accuracy.

Our study also has several limitations, some of which have already been addressed. Assuming that the gold standard of testing should be a practical test and not a survey-based case test, some of these limitations are related to the study design: these mainly include aspects related to data collection, as no information was available for the percentage of people willing to participate, the differences between respondents and nonrespondents, including how they were approached, and the response rate. Other limitations are inherent to voluntary surveys, including concerns not only about the truthfulness of the answers provided and the attention placed on certain answers but also about selection bias, as we expect HPs with higher digital skills to have been more likely to participate in the study. Differences in the interpretation of questions posed should also be considered. Finally, the exploratory nature of this study leads to hypothesis-generating conclusions rather than definitive conclusions.

Conclusions

Knowing the digital competence level of HPs is fundamental for promoting relevant strategic policies and actions to ensure that the right resources and conditions are in place for good professional performance. Such strategies would include the design and provision of specific training to qualify and accredit the digital competences of HPs.

This exploratory study highlights the need to improve the digital competences of HPs working in Catalonia. The results have informed the Health Plan for Catalonia 2021-2025 [31] and lay the foundations for the development and deployment of a framework program for HPs' digital competences that should include assessment indicators and standards to meet the COMPDIG-Salut project's goals. On the basis of the definition of this digital competence framework, training methodologies and content will be developed for implementation in bachelor's degree programs (the basic educational level for students of the various health care professions) and continuing education programs for working HPs, which in both cases must include the assessment and accreditation of digital competences.

The digital competence test showed acceptable consistency for the objective pursued, although improvements are needed to fine-tune its accuracy. The findings of this study lay the foundations for designing a strategic plan for training Catalan HPs.

Acknowledgments

The authors thank all health care professionals who participated in the study and the different organizations who helped in the dissemination thereof.

Data Availability

We cannot disclose the entire digital competence test, as this would potentially invalidate its use in further research. An extract of the test can be found in Multimedia Appendix 1. However, the test may be made available upon request. Please contact the principal investigator for further information.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Questionnaire on the level of digital competences of health professionals.
[DOCX File , 51 KB - jmir_v24i11e38347_app1.docx]

Multimedia Appendix 2
"Others" participants' digital competences and use of, training needs for, and interest in digital tools.
[DOCX File , 22 KB - jmir_v24i11e38347_app2.docx]

References

1. eHealth: Digital health and care. European Commission. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/overview_en [accessed 2022-07-26]
2. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly B, et al. Trends in the use of telehealth during the emergence of the COVID-19 pandemic - United States, January-March 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 30;69(43):1595-1599 [FREE Full text] [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]
3. Machleid F, Kaczmarczyk R, Johann D, Balčiūnas J, Atienza-Carbonell B, von Maltzahn F, et al. Perceptions of digital health education among European medical students: mixed methods survey. J Med Internet Res 2020 Aug 14;22(8):e19827 [FREE Full text] [doi: 10.2196/19827] [Medline: 32667899]
4. Telemedicine: Opportunities and developments in Member State. World Health Organization. URL: https://www.afro.who.int/publications/telemedicine-opportunities-and-developments-member-state [accessed 2021-10-15]
5. Mouratidis K, Papagiannakis A. COVID-19, Internet, and mobility: the rise of telework, telehealth, e-learning, and e-shopping. Sustain Cities Soc 2021 Nov;74:103182 [FREE Full text] [doi: 10.1016/j.scs.2021.103182] [Medline: 34540566]
6. Sharma V, Moulton G, Ainsworth J, Augustine T. Training digitally competent clinicians. BMJ 2021 Mar 25;372:n757. [doi: 10.1136/bmj.n757] [Medline: 33766891]
7. 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]
8. Saigí-Rubió F, Vidal-Alaball J, Torrent-Sellens J, Jiménez-Zarco A, López Segui F, Carrasco Hernandez M, et al. Determinants of Catalan public primary care professionals' intention to use digital clinical consultations (eConsulta) in the post-COVID-19 context: optical illusion or permanent transformation? J Med Internet Res 2021 May 31;23(6):e28944 [FREE Full text] [doi: 10.2196/28944] [Medline: 34097638]
9. Lolich L, Riccò I, Deusdad B, Timonen V. Embracing technology? Health and social care professionals' attitudes to the deployment of e-Health initiatives in elder care services in Catalonia and Ireland. Technol Forecast Soc Change 2019 Oct;147:63-71. [doi: 10.1016/j.techfore.2019.06.012]
10. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
11. Socha-Dietrich K. Empowering the health workforce to make the most of the digital revolution. Organisation for Economic Co-operation and Development. 2021 Nov 25. URL: https://www.oecd-ilibrary.org/content/paper/37ff0eaa-en [accessed 2021-10-12]
12. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
13. Temprana-Salvador J, López-García P, Castellví Vives J, de Haro L, Ballesta E, Rojas Abusleme M, et al. DigiPatICS: digital pathology transformation of the Catalan health institute network of 8 hospitals-planification, implementation, and preliminary results. Diagnostics (Basel) 2022 Mar 30;12(4):852 [FREE Full text] [doi: 10.3390/diagnostics12040852] [Medline: 35453900]
14. Reunió plenària I Fòrum de diàleg professional. Departament de Salut. URL: http://salutweb.gencat.cat/ca/el_departament/eixos-xiv-legislatura/forum-dialeg-professional/etapes/treballs-fase-anterior/i-forum-diagnostic/ [accessed 2022-07-27]
15. Committee on Digital Skills for Health Professionals. Digital skills for health professionals. European Health Parliament. 2016 Jun 29. URL: https://www.healthparliament.eu/digital-skills-health-professionals/ [accessed 2022-07-22]
16. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the eHealth literacy assessment toolkit. J Med Internet Res 2018 May 10;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
17. Digital Skills Assessment Guidebook. DocPlayer. URL: https://docplayer.net/186932834-Digital-skills-assessment-guidebook.html [accessed 2022-07-26]
18. MyDigiSkills. Asociación de Universidades Populares de Extremadura. URL: https://mydigiskills.aupex.org [accessed 2022-07-26]
19. Skills and knowledge assessment and development framework. EU*US eHealth Work. URL: http://www.ehealthwork.eu/Skills%20and%20Knowledge%20Assessment%20and%20Development%20Framework/index.html [accessed 2022-07-26]

XSL•FO
RenderX

20. Test Ikanos de competencias digitales - Gaitasun digitalen ikanos testa. Eusko Jaurlaritza. URL: https://test.ikanos.eus/ [accessed 2022-07-26]
21. Skov A. The Digital Competence Wheel. Center for Digital Dannelse. URL: https://digital-competence.eu/ [accessed 2022-07-26]
22. Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health Informatics J 2020 Mar;26(1):592-612 [FREE Full text] [doi: 10.1177/1460458219839613] [Medline: 30983476]
23. What is ACTIC. gencat. URL: http://actic.gencat.cat/en/actic_informacio/actic_que_es_l_actic_/index.html [accessed 2022-07-26]
24. The Digital Competence Framework. European Commission. URL: https://joint-research-centre.ec.europa.eu/digcomp/digital-competence-framework_en [accessed 2022-07-26]
25. Ley 44/2003, de 21 de noviembre, de ordenación de las profesiones sanitarias. Agencia Estatal Boletín Oficial del Estado. URL: https://www.boe.es/buscar/act.php?id=BOE-A-2003-21340 [accessed 2022-07-26]
26. Kelley K, Clark B, Brown V, Sitzia J. Good practice in the conduct and reporting of survey research. Int J Qual Health Care 2003 Jun;15(3):261-266. [doi: 10.1093/intqhc/mzg031] [Medline: 12803354]
27. ACTIC contents. Generalitat de Catalunya. 2021 Mar 15. URL: https://actic.gencat.cat/web/.content/01_informacio/documents/arxius/proposta_continguts21.pdf [accessed 2021-03-12]
28. Guitert M, Romeu T, Colas JF. Basic digital competences for unemployed citizens: conceptual framework and training model. Cogent Educ 2020 Apr 16;7(1):1748469. [doi: 10.1080/2331186x.2020.1748469]
29. Fernández M, Romeu T, Guitert M, Simón J, Ojando E. Challenges of the New Accreditation in Digital Competences for Citizens (ACTIC): Update, Validation and Results. Zenodo. 2022 May 23. URL: https://zenodo.org/record/6574366#.Y1r6eexBxpQ [accessed 2021-03-10]
30. Padró municipal d'habitants - Població a 1 de gener. Per sexe, edat quinquennal i regions sanitàries. gencat. Barcelona, Spain: Institut d'Estadística de Catalunya URL: https://www.idescat.cat/pub/?id=pmh&n=13312 [accessed 2021-10-26]
31. Pla de salut de Catalunya 2021-2025. Generalitat de Catalunya. Departament de Salut. Barcelona, Spain: Generalitat de Catalunya; 2021. URL: https://salutweb.gencat.cat/web/.content/_departament/pla-de-salut/pla-de-salut-2021-2025/pla-salut-catalunya-2021-2025.pdf [accessed 2021-04-13]
32. Brunner M, McGregor D, Keep M, Janssen A, Spallek H, Quinn D, et al. An eHealth capabilities framework for graduates and health professionals: mixed-methods study. J Med Internet Res 2018 May 15;20(5):e10229 [FREE Full text] [doi: 10.2196/10229] [Medline: 29764794]
33. Keep M, Janssen A, McGregor D, Brunner M, Baysari MT, Quinn D, et al. Mapping eHealth education: review of eHealth content in health and medical degrees at a metropolitan tertiary institute in Australia. JMIR Med Educ 2021 Aug 19;7(3):e16440 [FREE Full text] [doi: 10.2196/16440] [Medline: 34420920]
34. Echelard JF, Méthot F, Nguyen HA, Pomey MP. Medical student training in eHealth: scoping review. JMIR Med Educ 2020 Sep 11;6(2):e20027 [FREE Full text] [doi: 10.2196/20027] [Medline: 32915154]
35. Nazeha N, Pavagadhi D, Kyaw BM, Car J, Jimenez G, Tudor Car L. A digitally competent health workforce: scoping review of educational frameworks. J Med Internet Res 2020 Nov 05;22(11):e22706 [FREE Full text] [doi: 10.2196/22706] [Medline: 33151152]
36. EU*US eHealth Work. URL: http://www.ehealthwork.eu/ [accessed 2022-07-26]
37. Digital literacy of the wider workforce. Health Education England. 2018 Feb 12. URL: https://www.hee.nhs.uk/our-work/digital-literacy [accessed 2022-07-26]
38. European Skills Agenda - Employment, Social Affairs and Inclusion. European Commission. URL: https://ec.europa.eu/social/main.jsp?catId=1223 [accessed 2022-07-26]
39. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 07;160(1):48-54 [FREE Full text] [doi: 10.7326/M13-1531] [Medline: 24573664]

Abbreviations

ACTIC: Accreditation of Competence in Information and Communication Technology
COMPDIG-Salut: Digital Skills for HPs
HP: health care professional
ICT: information and communication technology

XSL•FO
RenderX

Edited by G Eysenbach; submitted 30.03.22; peer-reviewed by G Evangelinos, K Vossen, L Kayser, F Lopez Segui; comments to author 12.06.22; revised version received 03.08.22; accepted 13.10.22; published 30.11.22.

Please cite as:
Reixach E, Andrés E, Sallent Ribes J, Gea-Sánchez M, Àvila López A, Cruañas B, González Abad A, Faura R, Guitert M, Romeu T, Hernández-Encuentra E, Bravo-Ramirez S, Saigí-Rubió F
Measuring the Digital Skills of Catalan Health Care Professionals as a Key Step Toward a Strategic Training Plan: Digital Competence Test Validation Study
J Med Internet Res 2022;24(11):e38347
URL: https://www.jmir.org/2022/11/e38347
doi:10.2196/38347
PMID:36449330

©Elisenda Reixach, Erik Andrés, Josuè Sallent Ribes, Montserrat Gea-Sánchez, Alícia Àvila López, Bea Cruañas, Anna González Abad, Ricard Faura, Montse Guitert, Teresa Romeu, Eulàlia Hernández-Encuentra, Sandra Bravo-Ramirez, Francesc Saigí-Rubió. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Generation of a Lung Cancer Health Factor Distribution Using Patient Graphs Constructed From Electronic Medical Records: Retrospective Study

Anjun Chen[1], PhD; Ran Huang[1], MSc; Erman Wu[1], MD; Ruobing Han[2], MA; Jian Wen[3], PhD; Qinghua Li[4], MD; Zhiyong Zhang[4], PhD; Bairong Shen[1], PhD

[1]Institutes for System Genetics, West China Hospital, Chengdu, China

[2]iHealthd Shanghai Inc, Shanghai, China

[3]Guilin Medical University Affiliateted Hospital, Guilin, China

[4]Guilin Medical University, Guilin, China

Corresponding Author:
Bairong Shen, PhD
Institutes for System Genetics
West China Hospital
2222 Xingchuan Road
Chengdu, 610212
China
Phone: 86 15995854635
Email: bairong.shen@scu.edu.cn

Abstract

Background: Electronic medical records (EMRs) of patients with lung cancer (LC) capture a variety of health factors. Understanding the distribution of these factors will help identify key factors for risk prediction in preventive screening for LC.

Objective: We aimed to generate an integrated biomedical graph from EMR data and Unified Medical Language System (UMLS) ontology for LC, and to generate an LC health factor distribution from a hospital EMR of approximately 1 million patients.

Methods: The data were collected from 2 sets of 1397 patients with and those without LC. A patient-centered health factor graph was plotted with 108,000 standardized data, and a graph database was generated to integrate the graphs of patient health factors and the UMLS ontology. With the patient graph, we calculated the connection delta ratio (CDR) for each of the health factors to measure the relative strength of the factor's relationship to LC.

Results: The patient graph had 93,000 relations between the 2794 patient nodes and 650 factor nodes. An LC graph with 187 related biomedical concepts and 188 horizontal biomedical relations was plotted and linked to the patient graph. Searching the integrated biomedical graph with any number or category of health factors resulted in graphical representations of relationships between patients and factors, while searches using any patient presented the patient's health factors from the EMR and the LC knowledge graph (KG) from the UMLS in the same graph. Sorting the health factors by CDR in descending order generated a distribution of health factors for LC. The top 70 CDR-ranked factors of disease, symptom, medical history, observation, and laboratory test categories were verified to be concordant with those found in the literature.

Conclusions: By collecting standardized data of thousands of patients with and those without LC from the EMR, it was possible to generate a hospital-wide patient-centered health factor graph for graph search and presentation. The patient graph could be integrated with the UMLS KG for LC and thus enable hospitals to bring continuously updated international standard biomedical KGs from the UMLS for clinical use in hospitals. CDR analysis of the graph of patients with LC generated a CDR-sorted distribution of health factors, in which the top CDR-ranked health factors were concordant with the literature. The resulting distribution of LC health factors can be used to help personalize risk evaluation and preventive screening recommendations.

(J Med Internet Res 2022;24(11):e40361) doi:10.2196/40361

XSL•FO
RenderX

KEYWORDS

lung cancer; risk factor; patient graph; UMLS knowledge graph; Unified Medical Language System; connection delta ratio; EMR; electronic health record; EHR; electronic health record; cancer

Introduction

Early lung cancer (LC) detection is a key strategy to combat this deadly disease worldwide [1]. The National Lung Screening Trial in the United States and similar clinical trials around the world have shown an approximately 20% reduction in mortality from LC as a result of screening with low-dose computed tomography [2]. Based on these studies, LC screening medical guidelines as well as statistical risk prediction models including $PLCO_{M2012}$ have been implemented to recommend screening for smokers [3]. However, screening is not commonly recommended for nonsmokers even though they represent a significant percentage of patients with LC worldwide, 15%-20% among male patients and over 50% among female patients [4]. In addition, adoption of LC screening is still very low. For example, only approximately 5% of the at-risk population received their annual screening in the United States [5].

Risk-based or personalized screening approaches are being studied to overcome these challenges [6]. We believe that a deeper understanding of the spectrum of risk factors for LC and applying technologies such as machine learning and knowledge graphs (KGs) will generate more cost-effective screening solutions.

KGs have been widely applied in biomedical research. For interpreting proteomics data, a large-scale clinical KG has been plotted from biomedical data using the Neo4j tool [7]. Open-source graph databases and tools including Neo4j have made it easier to build and analyze KGs [8]. Studies have also demonstrated that construction of high-quality patient KGs from electronic medical records (EMRs) using rudimentary concept extraction is feasible and that the KGs can be used to predict diagnosis on the basis of symptoms [9]. Even though graphical representation of patient data holds the promise to illuminate insights in health care and to transform such insights gleaned from EMR data into actionable knowledge, the application of EMR-wide graphs for studying individual disease diagnosis journeys or treatment processes is still limited [10]. A graphical data model has been constructed, integrating clinical and molecular data of patients with non–small cell LC in the Cancer Genome Atlas LC data sets [11]. Another recent study of synthetic patients proposed a new graphical method to identify any particular disease's potential risk factor distribution from EMR (personal communication by A Chen, March 1, 2022).

The Unified Medical Language System (UMLS) ontology, freely available from the National Library of Medicine, is a KG consisting of millions of nodes and relationships [12]. It forms the foundation of interoperable biomedical information systems and services, including electronic health records. Connecting the UMLS KG to patient graphs may enable semantic search of patient data and support clinical decision-making [13].

This study aimed to construct a patient health factor graph for LC from a hospital EMR and integrate it with the UMLS KG for graph search and risk factor analysis. Through graph search, the study also aimed to generate a distribution of LC health factors, which was expected to help implement personalized LC risk evaluation for preventive screening.

Methods

EMR Health Factor Data Collection

We deidentified the patient records from January 2018 to June 2021 and saved them on a secured data server controlled by the hospital's informatics department. The data set had approximately 1 million patients and 7 million encounters including both outpatients and inpatients, in which patient names, dates of birth, contacts, and addresses were removed. The original identifiers of patients and encounters were replaced by irrelevant random numbers. Before using the data, our research team members were trained in the hospital's patient data security and privacy policy.

Because the EMR data had no usable codes associated with the diagnoses, synonyms of LC in Chinese were used to search for patients with LC. A total of 1397 patients with LC aged ≥30 years were included in the target data set. The same number (n=1397) of patients without LC and aged ≥30 years were randomly selected as control (or background) patients for comparison purposes.

Deidentified records of outpatient and inpatient visits, diagnoses, laboratory tests, and procedures were imported into a custom data collection tool on the secured data server. The data tool automatically extracted laboratory test data and saved them in the database. Researchers manually selected data from text records and entered them into the database. Because the records were not coded, practical rules were developed to improve consistency in the data collection process. Synonyms were automatically converted to "local standard terms" and the resulting data were called "local standard data." For each patient, only data from before the final diagnosis of LC were collected for studying disease risk factors, and a patient diagnosis journey (PDJ) object was created in the data tool to contain 1 or multiple encounters leading to the final diagnosis. When exporting PDJ data to a CSV file for analysis, only the latest data for each health factor in PDJ were selected. The final raw data set contained near 50,000 data from patients with LC and over 58,000 data from background patients. There were over 3000 different health factors identified in these data.

Patient Graph Construction

To simplify the patient graph, continuous numerical data were converted to categorical data. For example, values of age were converted to categories (ranges), including 30-50, 50-70, and >70 years; the value of drinking was "true" if the patient consumed >1 drink per day; the value of smoking was "true" if the patient smoked >1 cigarette per day. Laboratory findings from the EMR were already recorded as categorical variables: normal or abnormal; true or false; positive or negative; high,

XSL•FO
RenderX

medium, or low; and up, down, or normal. After value conversion, approximately 93,000 standard data for about 550 factors (ie, codes) that appeared in at least 10 patients with LC were selected and saved into a factor import CSV file. The format of the factor import file was as follows: virtual-id, category, code, term, value, unit, converted-value, and date. Patients with LC and background patients (N=2794) were both saved in a patient import file, one patient per line, with the following format: virtual-id, LC-label (1 for LC, 0 for background), and factor-count.

We used the Neo4j Desktop tool (version 4.4) available freely from Neo4j Inc, which is a graph database with a graphical user interface (Neo4j Browser) to query with Cypher language and view graphs. It provides an application programming interface through a Python driver. It can load data from CSV files to construct graphs. In our patient-centered graph model, each patient was represented by a "Patient" node (total of 2794 patient nodes), while health factor and value pairs were represented by 650 factor nodes. Because all values were categorical and some health factors had more than 1 piece of categorical data, the number of factor-value pair nodes increased from 550 to 650. The health factors were further subdivided into the following categories: Condition, Symptom, Observation, History, RiskFactor, Labtest, Procedure, Medication, and Treatment. The graph drew over 93,000 connections from patients to factors. Constraints were created on each label to ensure uniqueness. Patient nodes required virtual-id while all factor nodes required category, code, and converted-value as node key.

UMLS Disease Subgraph Construction

The UMLS 2020AB release was downloaded from the National Library of Medicine's UMLS website and installed locally by following the provided instructions. The local UMLS ontology had 2.8 million concepts, 8.3 million terms, and 39.1 million relationships. For generating an LC UMLS subgraph, we directly used the concept file MRCONSO.RRF and relation file MRREL.RRF in rich release format to generate Neo4j graph import files. The LC codes were first expanded to a more complete set of LC codes using the UMLS hierarchy (Table 1). We then used the expanded concept unique identifiers to find all horizontal relations (approximately 1100) between these LC target concepts and other biomedical concepts from over 39 million relations in UMLS ontology. The relations discovered were filtered by a selected set of UMLS relationship attributes for biological or medical concepts (Textbox 1); these were categorized into either biological concept relationships (called "biorel") or medical concept relationships (called "medrel"). To visualize this simple categorization of biomedical knowledge, we added RelCat nodes between TargetConcept nodes and related Concept nodes in the UMLS subgraph as shown in Figure 1. We then introduced a single AbstractPatient node to connect with all LC TargetConcept nodes. Connecting the patient nodes in EMR graph to the single AbstractPatient node resulted in an integrated biomedical graph that can present any patient's health factors together with biomedical knowledge from UMLS ontology for LC.

Table 1. Expanded lung cancer concepts in the Unified Medical Language System (UMLS) hierarchy.

UMLS concept unique identifiers	Term	SNOMEDCT code
C0581834	Suspected lung cancer	162573006
C0242379	Malignant neoplasm of lung	363358000
C0149925	Small cell carcinoma of lung	254632001
C0007131	Non-Small Cell Lung Carcinoma	254637007
C0152013	Adenocarcinoma of lung (disorder)	254626006
C0149782	Squamous cell carcinoma of lung	254634000
C1306460	Primary malignant neoplasm of lung	93880001
C0153676	Secondary malignant neoplasm of lung	94391008

XSL•FO
RenderX

Textbox 1. List of Unified Medical Language System (UMLS) relationship attributes and categories.

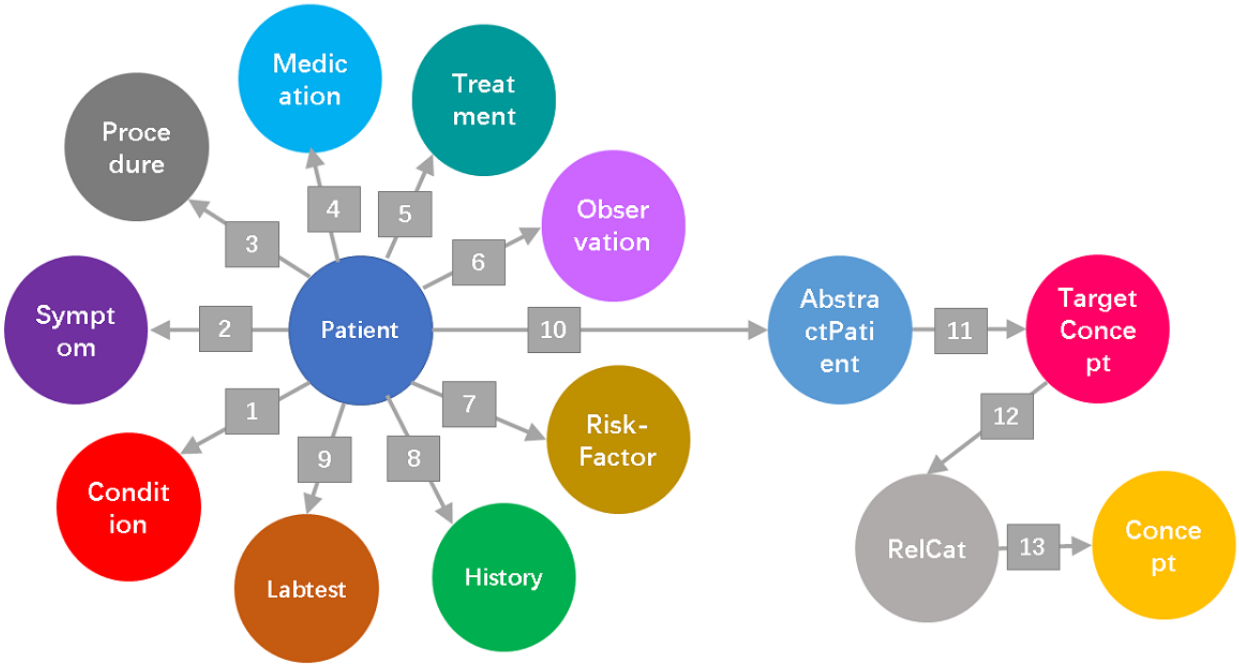
Biological concept relationships:

- gene_associated_with_disease
- gene_involved_in_pathogenesis_of_disease
- gene_mapped_to_disease
- gene_product_malfunction_associated_with_disease
- gene_product_is_biomarker_of
- may_be_cytogenetic_abnormality_of_disease
- may_be_molecular_abnormality_of_disease

Medical concept relationships:

- may_treat
- regimen_has_accepted_use_for_disease
- has_associated_finding
- associated_finding_of
- associated_disease
- is_finding_of_disease
- related_to
- clinically_associated_with
- co-occurs_with
- may_be_associated_disease_of_disease
- may_be_finding_of_disease

Figure 1. Biomedical graph model for the integration of the electronic medical record patient graph with the Unified Medical Language System knowledge graph of lung cancer. Numbered relationship labels are listed in Table 2.



XSL•FO
RenderX

Table 2. Node and relationship labels in the integrated biomedical graph model (shown in Figure 1).

Number	From node label	Relationship labels	To node label
1	Patient	HAS_CONDITION	Condition
2	Patient	HAS_SYMPTOM	Symptom
3	Patient	HAS_PROCEDURE	Procedure
4	Patient	HAS_MEDICATION	Medication
5	Patient	HAS_TREATMENT	Treatment
6	Patient	HAS_OBSERVATION	Observation
7	Patient	HAS_RISKFACTOR	RiskFactor
8	Patient	HAS_HISTORY	History
9	Patient	HAS_LABTEST	Labtest
10	Patient	INSTANCE_OF	AbstractPatient
11	AbstractPatient	MAY_HAVE_TARGET	TargetConcept
12	TargetConcept	HAS_RELCAT	RelCat
13	RelCat	HAS_RELA	Concept

Patient Health Factor Distribution

We developed a Python script to automatically query the patient graph with each of the health factors. The number of connections from each factor to LC target patients (depicted as "TPC" in equation 1) and background patients (depicted as "BPC" in equation 1) in the search results were counted separately. For each factor, the delta of patient connection counts was calculated by subtracting the number of background patient connections from that of the target patient connections. Division of the delta by the total number of patient connections yielded the "connection delta ratio" (CDR), a relative measure of the strength of connections from a factor to the target patient. Sorting factors by CDR and plotting a graph of CDR versus the sorted factors yielded a distribution of LC health factors from high to low strength.

$$CDR = (TPC - BPC) / (TPC + BPC) \quad \textbf{(1)}$$

A CDR between 1 and 0 implied that the factor was more related to the target patient, 1 being most related. A CDR below 0 implied that the factor was more related to the background patient.

In this study, factors with a CDR of >0.5 and having connections with at least 10 patients with LC were selected for literature verification. The local standard terms were first translated to English and the corresponding UMLS concepts as well as standard codes from SNOMEDCT_US, LOINC, or RxNORM if possible. We then searched the research literature on Google, Google Scholar, PubMed for each health factor and reviewed the published studies to verify whether the health factors were confirmed risk factors, correlated with LC, were unrelated to LC, or had an unsure relationship with LC. If a factor's relationship with LC was inconclusive in existing research reports, the factor was tagged as "unsure." For example, to look up the factor "Hypocalcemia," search terms included "Lung cancer risk factor Hypocalcemia" and other variations if necessary.

Ethical Considerations

This retrospective study of EMR patient data has been approved by the institutional review board of Guilin Medical University Associated Hospital in China (QTLL202139).

Results

Integrated Graph Model of the EMR Patient Graph and the UMLS KG

To study the spectrum of health factors related to LC in the hospital EMR, we applied a new graph method that we recently developed using synthetic patient data. Figure 1 shows the graph model integrating the EMR patient graph and UMLS knowledge subgraph for LC. The patient graph is patient-centered with patient nodes connecting to different categories of health factor nodes. Table 2 lists the relationships between nodes, as generated in the graph database. The UMLS subgraph in this model is focused on the horizontal biomedical relationships between LC nodes and related concept nodes. Such an integration model enables the presentation of a patient's actual health factors together with the UMLS KG's related biomedical factors in the same graph.

Patient Health Factor Graph Based on EMR Data

From the hospital EMR, 1397 patients with LC were selected along with the same number of background patients without LC. After deidentified data of laboratory tests and procedures were integrated into the corresponding encounters, a total of 108,000 standard data for various categories of health factors were extracted from patient encounters. Although over 3000 different factors were collected, only approximately 550 factors shared by at least 10 patients with LC were used for building the patient health factor graph.

The patient health factor graph was constructed by importing patient properties for the patient nodes and factor properties for the corresponding health factor nodes. The resulting patient graph had 93,000 relations between the 2794 patient nodes and

XSL•FO
RenderX

650 factor-value pair nodes. Table 3 lists several examples of Cypher queries for searching patients with various factors. For example, clinicians can easily search for patients with LC with 1 or more co-occurring diseases (Figure 2), with 1 or more nonlaboratory factors (symptoms, medical histories, and observations; Figure 3), or laboratory tests (Figure 4). One can also easily search for any number of health factors shared by patients among patients with LC.

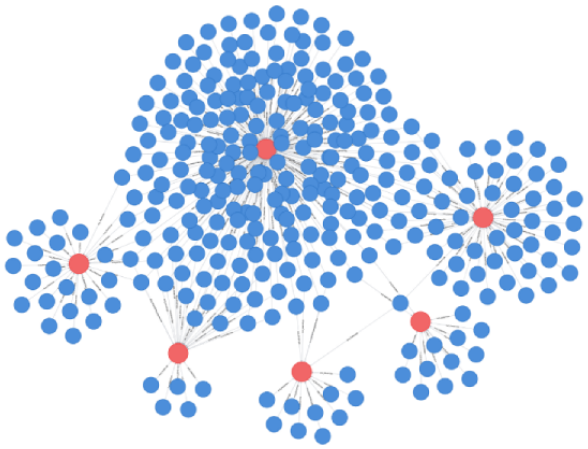
Table 3. Examples of graph search tasks and queries using Cypher language.

Number	Graph search task	Cypher query[a,b]
1	• Search for patients with LC with 1-6 co-occurring diseases and present the topology. • C-389764: Hypocalcemia • C-172569: Bacterial Infection • C-765209: Obstructive pneumonia • C-305976: Pneumothorax • C-352894: Leukopenia • C-654730: Pneumonia	match (p:Patient {label:'1'})-->(f {cat: 'dac'}) where f.code = 'C-389764' or f.code = 'C-172569' or f.code = 'C-765209' or f.code = 'C-305976' or f.code = 'C-352894' or f.code = 'C-654730' return p, f;
2	• Search for patients with LC with 1-5 nonlaboratory factors and present the topology • C-549780: Pain • C-289547: Bloodstained sputum • C-127089: Hoarseness • C-029761: Productive Cough • C-294680: Swollen Lymph Node in head and neck	match (p:Patient {label:'1'})-->(f) where (f.code = 'C-549780' and f.valcvt = 'true') or (f.code = 'C-289547' and f.valcvt='true') or (f.code = 'C-127089' and f.valcvt='true') or (f.code = 'C-029761' and f.valcvt='true') or (f.code = 'C-294680' and f.valcvt='true') return p, f;
3	• Search for patients with LC with 1-5 laboratory test values and present the topology. • C-659218: Hepatitis B virus • C-493765: Squamous cell carcinoma antigen • C-573086: Neuron-specific enolase measurement • C-120948: Gastrin-releasing peptide precursor increased • C-814793: Mycoplasma pneumoniae antibody	match (p:Patient {label:'1'})-->(f {cat: 'lab'}) where (f.code = 'C-659218' and f.valcvt = 'true') or (f.code = 'C-493765' and f.valcvt = 'up') or (f.code = 'C-573086' and f.valcvt = 'up') or (f.code = 'C-120948' and f.valcvt = 'abnormal') or (f.code = 'C-814793' and f.valcvt = 'abnormal') return p, f;
4	• Search for 1 patient, show the electronic medical record health factor graph and the Unified Medical Language System knowledge graph together	match (p:Patient {label:'1', vpid:'_8908085766'})-->(f) match (p)-->(ap:AbstractPatient)-->(tc:TargetConcept)-->(cr:RelCat)-->(c:Concept) return p, f, ap, tc, cr, c;

[a]Using Neo4j Cypher query language.

[b]Patient with LC: label=1; background patient: label=0. Factor property f.code: unique local code. Factor property f.valcvt: converted value.

Figure 2. Topology of an example patient graph searched with 6 disease factors. Search query 1 in Table 3 was used. Patient nodes are shown in blue and factor nodes are shown in red. Lines represent relationships between a patient and factors.



XSL•FO
RenderX

Figure 3. Topology of an example patient graph searched with 5 nonlaboratory factors. Search query 2 in Table 3 was used. Patient nodes are shown in blue and factor nodes are shown in pink. Lines represent relationships between a patient and factors.

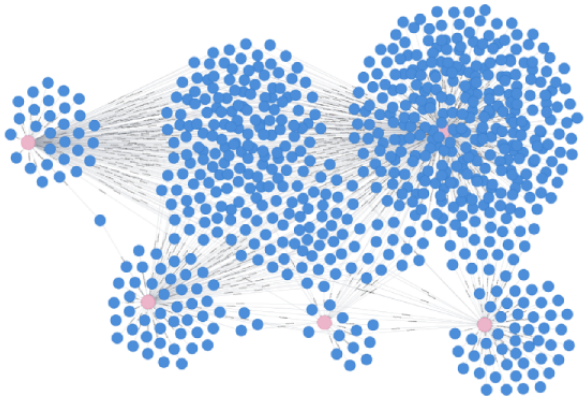
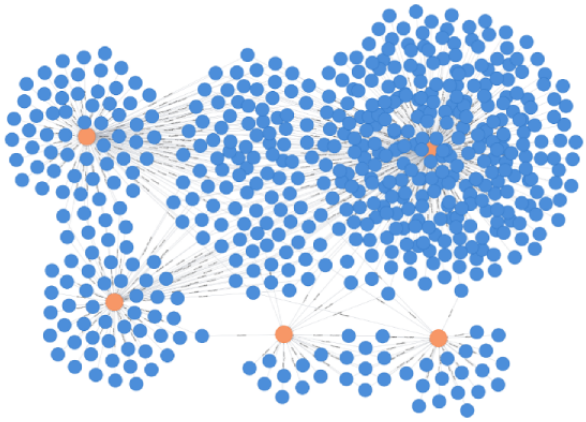


Figure 4. Topology of an example patient graph searched with 5 laboratory factors. Search query 3 in Table 3 was used. Patient nodes are shown in blue and factor nodes are shown in orange. Lines represent relationships between a patient and factors.



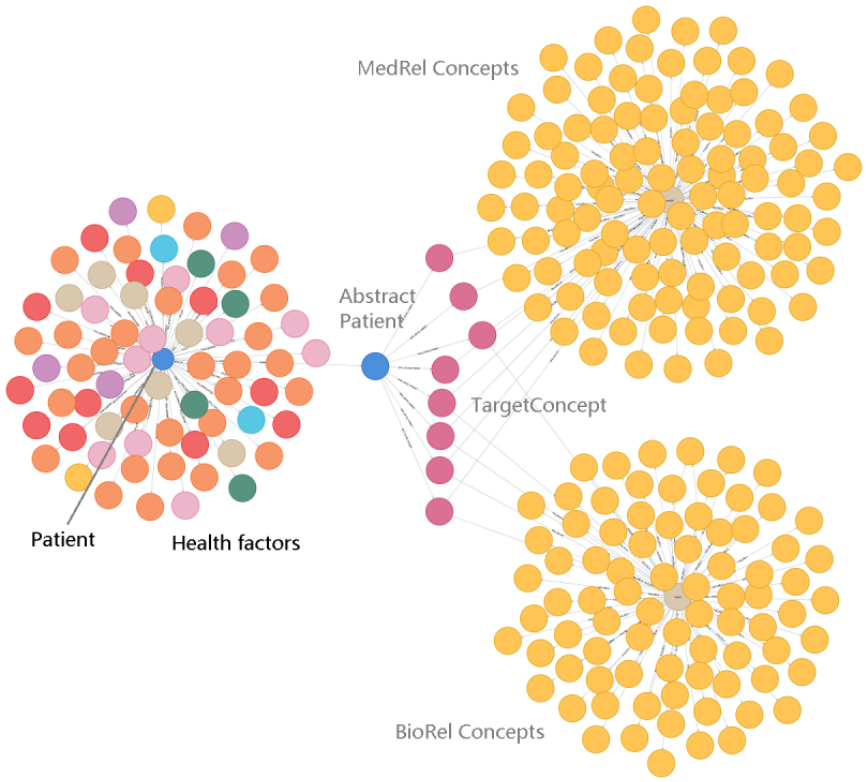
Integration of the EMR Patient Graph With the UMLS Subgraph

As the largest integrated biomedical ontology, the UMLS graph contains hierarchies of diseases and horizontal relationships with other entities. Within a disease family such as LC, the various types of LCs are horizontally connected to a myriad of related biomedical concepts including genes, proteins, symptoms, observations, medication, and treatments. This study is focused on the UMLS knowledge subgraph containing horizontal relationships for LC. Using the UMLS LC hierarchy, the target LC codes found in EMRs were expanded to 8 main LC concepts (Table 1). From these concepts, approximately 1100 relations were identified in the UMLS ontology. Most of the relations were hierarchical—for example, a parent-child relationship—and thus the relations were further filtered by the biomedical relationships that we were interested in (Textbox 1). The resulting UMLS LC biomedical subgraph had 8 LC concept nodes, 187 related biomedical concepts, and 188 horizontal biomedical relations (Figure 5).

Through a single AbstractPatient node, the EMR patient graph was connected to the UMLS subgraph for LC. Search query 4 in Table 3 and its search result in Figure 5 show an example presentation of both actual patient's health factors in the EMR and relevant biomedical knowledge in the UMLS in the same graph.

XSL•FO
RenderX

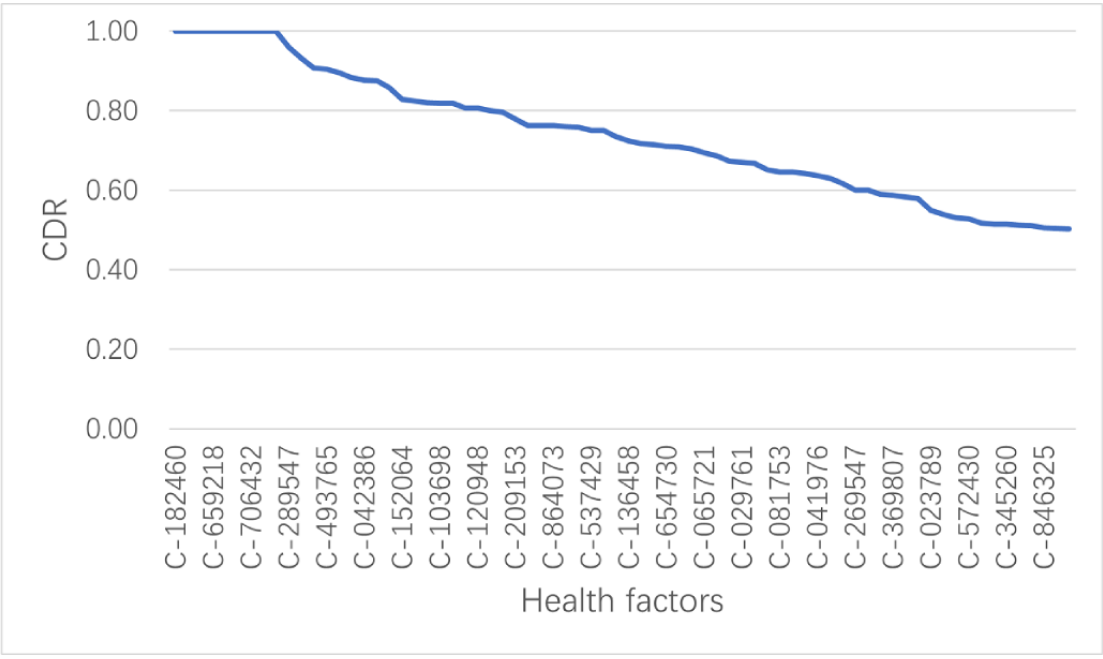
Figure 5. Example search result of the integrated biomedical graph. Search query 4 in Table 3 was used to search 1 specific ID of a patient with lung cancer. Left side: health factors from the electronic medical record of one patient with lung cancer. Right side: lung cancer biomedical knowledge from the Unified Medical Language System. Middle: single AbstractPatient as the connection. BioRel: biological concept relationship; MedRel: medical concept relationship.



Generation of the Distribution of LC Health Factors From the EMR

With the patient health factor graph, we searched for patients with LC and background patients with each of the health factors and its value. The connection delta ratios were calculated for each factor from the number of connections to patients with LC and the number of connections to background patients. Sorting factors by CDR in descending order generated a distribution of health factors for LC found in the EMR. The complete distribution of top-ranked factors over a CDR cutoff of 0.5 are shown in Table A1 in Multimedia Appendix 1 and plotted in Figure 6. As examples, up to 5 top health factors in each category are shown in Table 4. For understanding LC risk factors, this distribution excluded the various cancers, all procedures and medications related to cancers, and treatments.

Figure 6. Distribution curve of lung cancer health factors sorted by the connection delta ratio (CDR; cutoff=0.5). Only partial codes are visible on the x-axis. The full spectrum of lung cancer health factors can be found in Table A1 in Multimedia Appendix 1.



XSL•FO
RenderX

Table 4. Partial distribution of lung cancer health factors sorted by category and connection delta ratio (cutoff=0.5) as examples. The full distribution of lung cancer health factors is provided in Table A1 in Multimedia Appendix 1.

Category[a]	Local code	Term	Value	Connection delta ratio	Tag
dac	C-182460	Left lung pulmonary obstructive pneumonia	TRUE	1.00	confirmed
dac	C-248056	Right lung pulmonary obstructive pneumonia	TRUE	1.00	confirmed
dac	C-765209	Obstructive pneumonia	TRUE	1.00	confirmed
dac	C-305976	Pneumothorax	TRUE	0.93	correlated
dac	C-172569	Bacterial Infection	TRUE	0.88	correlated
lab	C-659218	Hepatitis B virus	TRUE	1.00	correlated
lab	C-493765	Squamous cell carcinoma antigen	up	0.90	confirmed
lab	C-573086	Neuron-specific enolase measurement	up	0.82	correlated
lab	C-952408	Non-small cell lung cancer associated-antigen	up	0.82	confirmed
lab	C-103698	Superoxide dismutase measurement	down	0.82	correlated
obs	C-039824	Mediastinal mass	TRUE	1.00	confirmed
obs	C-706432	Lung mass	TRUE	1.00	confirmed
obs	C-748932	Lung mass found in checkup	TRUE	1.00	confirmed
obs	C-134276	Lung shadow	TRUE	0.91	confirmed
obs	C-706281	Bronchial stenosis	TRUE	0.89	correlated
rf	C-902187	Smoking	TRUE	0.50	confirmed
smp	C-549780	Pain	TRUE	1.00	confirmed
smp	C-289547	Bloodstained sputum	TRUE	0.96	confirmed
smp	C-152064	Hemoptysis (cough up blood)	TRUE	0.83	correlated
smp	C-243071	Shoulder Pain	TRUE	0.82	confirmed
smp	C-127089	Hoarseness	TRUE	0.80	correlated

[a]Categories include condition (dac), laboratory test (lab), observation (obs), risk factor (rf), and symptom (smp).

We checked the medical literature for any associations between these top CDR-ranked health factors and LC [14-26]. This literature review confirmed that 70 out of the 71 factors (Table A1 in Multimedia Appendix 1) were LC risk factors or were correlated with LC. The relationship between 1 factor, laboratory test for immunoglobulin E levels, and LC was unsure according to the literature [27]. This high degree of concordance between the results of our CDR analysis and the literature suggests that the patient graph CDR method was effective in generating a reliable distribution of LC health factors from EMR patient data.

Discussion

Using hospital EMR patient data and applying the new patient graph CDR method recently developed from synthetic patient data, this study was able to construct an integrated biomedical graph for LC. From searching the graph, the study created a distribution of health factors for LC, which were verified through literature review. Our results show that the new strategy of first using synthetic patients for method development and then applying the methods with real patient data is valid and effective.

This study has implications for hospitals with regard to harnessing KG databases and technologies. First, generating an integrated biomedical graph with hospital EMR data may enable medical professionals to view individual patient's health factor graphs along with the related UMLS KGs for comprehensive comparisons. Current medical concept nodes horizontally related to the LC nodes are mostly genes and gene-related biological information, as well as drugs and treatment-related information from the UMLS ontology (see Figure 5). Since the UMLS is updated quarterly, the LC integrated biomedical graph will grow as the UMLS grows. Thus, this KG integration offers a new way for hospitals to bring continuously updated international standard biomedical knowledge to patient care. The current graph model is designed specifically for searching risk factors; however, it can be modified for other clinical information tasks. It may also be integrated with cancer-associated lifestyle KGs for disease management information [28].

The second implication of this study may be applying the CDR-ranked distribution of health factors to build more effective or practical machine learning models for LC risk prediction. Because the distribution ranks factors from higher to lower relative strength, they may be used to help select more health factors to build prediction models; that is, feature engineering. For example, we have an ongoing project experimenting with the factor distribution in building LC risk prediction machine learning models. Knowing the risk factors actually found in the EMR data, we could focus on these risk factors and reduce the

XSL•FO
RenderX

variables from over 100 to less than 30 in the machine learning models that were generated from EMR-wide data. To increase the LC screening rate in larger populations, machine learning models with a small number of variables for which data can be readily available in community and rural clinics are necessary.

In addition, the patient health factor graphs generated from EMR data may enable hospitals to study the effect of various types of factors in diagnosis, medication, treatment, and disease management. Such graph analysis complements existing statistical analysis. Traditionally, studies on individual risk factors are hypothesis driven and use a clinical trial or case-control study design [29]. The literature found in this study for verification of the health factor distribution collectively indicate the use of this approach [14-27]. Because this study's patient graph method is EMR data driven, it can reveal potential new risk factors or inconclusive risk factors that deserve additional research. For example, the factor "laboratory test for immunoglobulin E levels" was tagged as "unsure" in the distribution because prior studies were inclusive. Our CDR analysis suggests that this immunoglobulin E factor requires further clinical validation [30].

Because EMR data sometimes have biases and missing data, the EMR data–driven patient graph CDR method has limitations. CDR is a simple measurement of a factor's relative strength, but caution should be taken when considering factors with a high CDR but a small number of connections. The higher the number of connections, the more reliable the CDR. Hence, studies should set a cutoff for the CDR as well as the minimal number of connections to ensure that the study uses enough data. It is also important to recognize factors that might be affected by data biases and to exclude them from CDR analysis [31]. For EMRs lacking standardized and structured data, collecting standardized data is crucial but challenging. If a data collection pipeline is not fully automated, collecting enough unbiased standardized patient profile data will be a very time-consuming process.

In conclusion, by collecting standardized data of thousands of patients with and those without LC from EMRs, it was possible to generate a hospital-wide patient-centered health factor graph for graph search and presentation. It was also practical to integrate the patient graph with the UMLS KG for LC, enabling hospitals to bring continuously updated international standard biomedical KGs from the UMLS to clinical care. Applying CDR analysis to the graph of patients with LC yielded a CDR-sorted distribution of health factors, where top CDR-ranked health factors showed a high degree of concordance with the literature. The resulting distribution of LC health factors can be used to help personalize risk evaluation and preventive screening recommendations.

Acknowledgments

This work was supported by the Sichuan Science and Technology Program (grant2020YFQ0019) and the Guilin Municipal Science and Technology Bureau, China (grant 20190219-2).

Authors' Contributions

AC designed the study and drafted the manuscript. R Huang wrote programs and analyzed data. EW and R Han collected the data. JW supervised the study. ZZ, QL, and BS proposed the study, obtained funding, and directed the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete lung cancer health factor distribution sorted by category (Cat) and connection delta ratio (CDR).
[PDF File (Adobe PDF File), 159 KB - jmir_v24i11e40361_app1.pdf]

References

1. Pinsky PF. Lung cancer screening with low-dose CT: a world-wide view. Transl Lung Cancer Res 2018 Jun;7(3):234-242 [FREE Full text] [doi: 10.21037/tlcr.2018.05.12] [Medline: 30050762]
2. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 04;365(5):395-409 [FREE Full text] [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]
3. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013 Feb 21;368(8):728-736 [FREE Full text] [doi: 10.1056/NEJMoa1211776] [Medline: 23425165]
4. Dubin S, Griffin D. Lung Cancer in Non-Smokers. Mo Med 2020;117(4):375-379 [FREE Full text] [Medline: 32848276]
5. Lung Cancer Screening. National Cancer Institute. Cancer Trends Progress Report. URL: https://progressreport.cancer.gov/detection/lung_cancer [accessed 2022-11-14]
6. Wang Z, Wang Y, Huang Y, Xue F, Han W, Hu Y, et al. Challenges and research opportunities for lung cancer screening in China. Cancer Commun (Lond) 2018 Jun 07;38(1):34 [FREE Full text] [doi: 10.1186/s40880-018-0305-0] [Medline: 29880036]

XSL•FO
RenderX

7. Santos A, Colaço AR, Nielsen AB, Niu L, Strauss M, Geyer PE, et al. A knowledge graph to interpret clinical proteomics data. Nat Biotechnol 2022 May;40(5):692-702 [FREE Full text] [doi: 10.1038/s41587-021-01145-6] [Medline: 35102292]
8. Neo4j Desktop User Interface Guide. Neo4j, Inc. URL: https://neo4j.com/developer/neo4j-desktop/ [accessed 2022-11-14]
9. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a Health Knowledge Graph from Electronic Medical Records. Sci Rep 2017 Jul 20;7(1):5994 [FREE Full text] [doi: 10.1038/s41598-017-05778-z] [Medline: 28729710]
10. Schrodt J, Dudchenko A, Knaup-Gregori P, Ganzinger M. Graph-Representation of Patient Data: a Systematic Literature Review. J Med Syst 2020 Mar 12;44(4):86 [FREE Full text] [doi: 10.1007/s10916-020-1538-4] [Medline: 32166501]
11. Tuck D. A cancer graph: a lung cancer property graph database in Neo4j. BMC Res Notes 2022 Feb 14;15(1):45 [FREE Full text] [doi: 10.1186/s13104-022-05912-9] [Medline: 35164854]
12. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
13. Martinez D, Otegi A, Soroa A, Agirre E. Improving search over Electronic Health Records using UMLS-based query expansion through random walks. J Biomed Inform 2014 Oct;51:100-106 [FREE Full text] [doi: 10.1016/j.jbi.2014.04.013] [Medline: 24768598]
14. Signs and symptoms of lung cancer. American Cancer Society. URL: https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/signs-symptoms.html [accessed 2022-11-14]
15. What are the risk factors for lung cancer? Centers for Disease Control and Prevention. URL: https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm [accessed 2022-11-14]
16. Song C, Lv J, Liu Y, Chen JG, Ge Z, Zhu J, China Kadoorie Biobank Collaborative Group. Associations Between Hepatitis B Virus Infection and Risk of All Cancer Types. JAMA Netw Open 2019 Jun 05;2(6):e195718 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.5718] [Medline: 31199446]
17. Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, et al. Non-small-cell lung cancer. Nat Rev Dis Primers 2015 May 21;1:15009. [doi: 10.1038/nrdp.2015.9] [Medline: 27188576]
18. Ferraz Gonçalves JA, Costa T, Rema J, Pinto C, Magalhães M, Esperança A, et al. Hypocalcemia in cancer patients: An exploratory study. Porto Biomed J 2019;4(4):e45 [FREE Full text] [doi: 10.1097/j.pbj.0000000000000045] [Medline: 31930179]
19. You D, Zhang M, He W, Wang D, Yu Y, Yu Z, et al. Association between dietary sodium, potassium intake and lung cancer risk: evidence from the prostate, lung, colorectal and ovarian cancer screening trial and the Women's Health Initiative. Transl Lung Cancer Res 2021 Jan;10(1):45-56 [FREE Full text] [doi: 10.21037/tlcr-20-870] [Medline: 33569292]
20. Dreyer L, Prescott E, Gyntelberg F. Association between atherosclerosis and female lung cancer--a Danish cohort study. Lung Cancer 2003 Dec;42(3):247-254. [doi: 10.1016/s0169-5002(03)00295-2] [Medline: 14644511]
21. Kang SW. Superoxide dismutase 2 gene and cancer risk: evidence from an updated meta-analysis. Int J Clin Exp Med 2015;8(9):14647-14655 [FREE Full text] [Medline: 26628947]
22. Unsal E, Atalay F, Atikcan S, Yilmaz A. Prognostic significance of hemostatic parameters in patients with lung cancer. Respir Med 2004 Feb;98(2):93-98 [FREE Full text] [doi: 10.1016/j.rmed.2003.07.001] [Medline: 14971870]
23. Ercan H, Mauracher L, Grilz E, Hell L, Hellinger R, Schmid J, et al. Alterations of the Platelet Proteome in Lung Cancer: Accelerated F13A1 and ER Processing as New Actors in Hypercoagulability. Cancers (Basel) 2021 May 08;13(9):2260 [FREE Full text] [doi: 10.3390/cancers13092260] [Medline: 34066760]
24. Cui M, Li N, Liu X, Yun Z, Niu Y, Zhang Y, et al. Platelet distribution width correlates with prognosis of non-small cell lung cancer. Sci Rep 2017 Jun 14;7(1):3456 [FREE Full text] [doi: 10.1038/s41598-017-03772-z] [Medline: 28615714]
25. Yan L, Hu Z. Diagnostic accuracy of human epididymis secretory protein 4 for lung cancer: a systematic review and meta-analysis. J Thorac Dis 2019 Jul;11(7):2737-2744 [FREE Full text] [doi: 10.21037/jtd.2019.06.72] [Medline: 31463101]
26. Schmidt-Hansen M, Berendse S, Hamilton W, Baldwin DR. Lung cancer in symptomatic patients presenting in primary care: a systematic review of risk prediction tools. Br J Gen Pract 2017 Jun;67(659):e396-e404 [FREE Full text] [doi: 10.3399/bjgp17X690917] [Medline: 28483820]
27. Helby J, Bojesen S, Nielsen S, Nordestgaard B. IgE and risk of cancer in 37 747 individuals from the general population. Ann Oncol 2015 Aug;26(8):1784-1790 [FREE Full text] [doi: 10.1093/annonc/mdv231] [Medline: 25969367]
28. Chen Y, Sinha B, Ye F, Tang T, Wu R, He M, et al. Prostate cancer management with lifestyle intervention: From knowledge graph to Chatbot. Clinical and Translational Dis 2022 Feb 20;2(1). [doi: 10.1002/ctd2.29]
29. Stampfer MJ, Ridker PM, Dzau VJ. Risk factor criteria. Circulation 2004 Jun 29;109(25 Suppl 1):IV3-IV5. [doi: 10.1161/01.CIR.0000133446.69171.7d] [Medline: 15226245]
30. Kantor E, Hsu M, Du M, Signorello L. Allergies and Asthma in Relation to Cancer Risk. Cancer Epidemiol Biomarkers Prev 2019 Aug;28(8):1395-1403 [FREE Full text] [doi: 10.1158/1055-9965.EPI-18-1330] [Medline: 31167755]
31. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res 2018 May 29;20(5):e185 [FREE Full text] [doi: 10.2196/jmir.9134] [Medline: 29844010]

XSL•FO
RenderX

Abbreviations

CDR: connection delta ratio
EMR: electronic medical record
KG: knowledge graph
LC: lung cancer
PDJ: patient diagnosis journey
UMLS: Unified Medical Language System

Edited by R Kukafka; submitted 21.06.22; peer-reviewed by Z Wang, I Mircheva; comments to author 30.08.22; revised version received 09.09.22; accepted 25.10.22; published 25.11.22.

Please cite as:
Chen A, Huang R, Wu E, Han R, Wen J, Li Q, Zhang Z, Shen B
The Generation of a Lung Cancer Health Factor Distribution Using Patient Graphs Constructed From Electronic Medical Records: Retrospective Study
J Med Internet Res 2022;24(11):e40361
URL: https://www.jmir.org/2022/11/e40361
doi:10.2196/40361
PMID:36427233

©Anjun Chen, Ran Huang, Erman Wu, Ruobing Han, Jian Wen, Qinghua Li, Zhiyong Zhang, Bairong Shen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Impact of Social Influence on the Intention to Use Physician Rating Websites: Moderated Mediation Analysis Using a Mixed Methods Approach

Bernhard Guetz[1], BA, MSc; Sonja Bidmon[1], PhD

Department of Marketing and International Management, Alpen-Adria-Universitaet Klagenfurt, Klagenfurt am Woerthersee, Austria

Corresponding Author:
Bernhard Guetz, BA, MSc
Department of Marketing and International Management
Alpen-Adria-Universitaet Klagenfurt
Universitaetsstrasse 65-67
Klagenfurt am Woerthersee, 9020
Austria
Phone: 43 6508611182
Email: beguetz@edu.aau.at

Abstract

Background: Physician rating websites (PRWs) have become increasingly important in the cross-section between health and digitalization. Social influence plays a crucial role in human behavior in many domains of life, as can be demonstrated by the increase in high-profile influential individuals such as social media influencers (SMIs). Particularly in the health-specific environment, the opinion of family and friends has a significant influence on health-related decisions. However, so far, there has been little discussion about the role of social influence as an antecedent of behavioral intention to use PRWs.

Objective: On the basis of theories of social psychology and technology acceptance and theories from the economic perspective, this study aimed to evaluate the impact of social influence on the behavioral intention to use PRWs.

Methods: We conducted 2 studies by applying a mixed methods approach including a total of 712 participants from the Austrian population. The impact of social influence on the behavioral intention to use PRWs was investigated through linear regression and mediation and moderated mediation analysis using the PROCESS macro 4.0 in SPSS 27 (IBM Corp).

Results: The 2 studies show similar results. In study 1, an experiment, no direct effect of social influence on the behavioral intention to use PRWs could be detected. However, an indirect effect of social influence on the behavioral intention to use PRWs via credibility (b=0.572; P=.005) and performance expectancy (b=0.340; P<.001) could be confirmed. The results of study 2, a cross-sectional study, demonstrate that social influence seems to have a direct impact on the behavioral intention to use PRWs (b=0.410; P<.001). However, when calculating the proposed mediation model, it becomes clear that this impact may partly be explained through the 2 mediator variables—credibility (b=0.208; P<.001) and performance expectancy (b=0.312; P<.001). In contrast to the observed direct and indirect effect, neither demographic nor psychographic variables have a significant moderating impact on the influencing chain in study 2.

Conclusions: This study provides an indication that social influence has at least an indirect impact on the behavioral intention to use PRWs. It was observed that this impact is exerted through credibility and performance expectancy. According to the findings of both studies, social influence has the potential to boost the use of PRWs. As a result, these web-based networks might be a promising future interface between health care and digitalization, allowing health care practitioners to gain a beneficial external impact while also learning from feedback. Social influence nowadays is not just limited to friends and family but can also be exerted by SMIs in the domain of PRW use. Thus, from a marketing perspective, PRW providers could think of collaborating with SMIs, and our results could contribute to stimulating discussion in this vein.

(J Med Internet Res 2022;24(11):e37505) doi:10.2196/37505

KEYWORDS

social influence; eHealth literacy; patient satisfaction; physician rating websites

XSL•FO
RenderX

Introduction

Background

The number of websites where patients can publicly share their health care experiences has grown rapidly in recent years [1]. These web-based platforms are characterized by patients sharing their subjective perceived health experience with the entire web-based community by creating qualitative reviews and quantitative ratings [2]. In addition, there is an increasing number of patients who use these websites to make their health-related decisions as well as to search for and select a suitable health service provider [3]. Especially for health care providers, those web-based portals represent a cost-efficient possibility to achieve a positive external impact [4]. This shows that physician rating websites (PRWs) provide the opportunity to evaluate incidents in the health sector and make evidence-based decisions by referring to existing evaluations on rating sites [5].

Theoretical Background

We built our conceptual framework on insights from several theoretical domains. These can be categorized as theories from the social psychological perspective, theories of technology acceptance, and theories from the economic perspective. From a social psychological perspective, the theory of reasoned action (TRA) [6] and its extension, the theory of planned behavior (TPB), can be used as a framework for this study [7]. According to the TRA, it is assumed that individuals' *attitudes* and *subjective norms* shape their *behavioral intentions* as well as their *behavior* [8]. It is proposed that individuals are more likely to perform a specific *behavior* if they have a positive *attitude* toward this behavior and believe that others want them to perform it (*subjective norm*) [9]. The TRA has formed the theoretical underpinning of many empirical studies so far. As the meta-analysis by Sheppard et al [10] in 1988 could demonstrate, the empirical results of several studies contribute to support the TRA [10]. However, the proposed influencing chain was further extended through the more sophisticated version, the TPB [11]. According to this theory, there are 3 independent core components that shape an individual's *behavioral intentions* [12]. These include *attitudes* and *subjective norms* but also *perceived behavioral control* [13]. In this context, again, *attitude* describes the individual's view of a particular behavior, and *subjective norms* describe what others might think about the particular behavior [14]. However, *perceived behavioral control* is an individual's sense of control over their own behavior and represents an exogenous variable that, in contrast to the other 2, affects both *behavioral intention* and *behavior* itself [7].

The second stream of theories that provide both the foundation and framework for this study are the theories of technology acceptance and technology use. These theories are based on the TPB and integrate further theories, factors, and modifications depending on their individual characteristics [15-18]. The pioneering theory of this kind was the technology acceptance model (TAM), in which it was assumed that *perceived usefulness* and *perceived ease of use* influence the attitude toward using [19,20]. According to this model, *attitude toward using* represents the decisive predictor for *actual system use* [21]. Further developments of the original TAM were published as TAM 2 [22] and TAM 3 [23]. The currently prevailing theory of technology acceptance was developed on this basis. The so-called Unified Theory of Acceptance and Use of Technology (UTAUT) [24,25] and its further development, the Extended UTAUT [26], propose factor models that are characterized by a vast number of independent influencing variables [27]. These include, for example, *performance expectancy* or *social influence* [28]. According to the UTAUT, these factors affect the *use behavior* regarding new or adapted technologies whereby the impact can be obtained directly or indirectly via *behavioral intention* [29].

Finally, from an economical perspective, the theory of information economics (TIE) [30] and the concept of source similarity [31] also contribute to the theoretical underpinning of this study. According to the TIE, consumers are not able to sufficiently assess the quality of credence products [30]. In the medical context, this means that patients are dependent on additional sources beyond their mere subjective perception to be able to assess the quality of the medical encounter and the medical treatment involved [32,33]. The concept of similarity, in which it is assumed that the recipient of a piece of advice evaluates the quality of the source, could be valuable in this context [34]. According to this theoretical concept, the decision for or against advice received is dependent on the transmitter's attributed expertise and similarity to the perceiver's point of view [35]. Sources that are perceived as similar seem to have a significant impact from the perceiver's perspective because of the ascription of similar needs and expectations [31]. These assumptions show that advice from a person with a high similarity might lead to a change in behavior [36] and, thus, inter alia could also contribute to an increase in the *behavioral intention* to use PRWs.

The effect of *social influence* on the *behavioral intention* to use a new or adapted technology is well known. For example, previous studies have already shown that *social influence* has an impact on the *behavioral intention* to use mobile-based assessments [37], Instagram messaging [38], web-based banking [39], eHealth services [40], social media [41], e-government [42], e-learning [43], and accounting platforms [44]. Although this review is not exhaustive and includes only a fraction of the studies that have investigated the relationship between *social influence* and the behavioral use intention, to the best of our knowledge, the relationship between *social influence* and *behavioral intention* to use PRWs has not been investigated so far. For this reason, we conducted 2 studies by applying a mixed methods approach to investigate the direct and indirect effects of *social influence* on the *behavioral intention* to use PRWs, including potential moderation effects. The following section elaborates on this.

Determinants of Behavioral Intention to Use PRWs

The *behavioral intention* to use a new or adapted technology refers to the strength of the ambition to perform a particular action [6,22]. On the basis of various studies in this field (eg, Krueger and Carsrud [45], Tonglet et al [46], Hardeman et al [47], Anderson and Schwager [48], Hoogenbosch et al [49],

XSL•FO
RenderX

and Venkatesh and Zhang [24]), it can be assumed that the *behavioral intention* to use a new technology plays a crucial role in forecasting individuals' actual or future use behavior.

Social Influence

Consumer decisions in the product or service sector are often strongly influenced by individuals who have an impact on the customers' behavior [50]. The whole boom of influencer marketing nowadays is more or less based on this fact. However, in the medical field, the degree of perceived uncertainty from the patients' perspective is frequently very high, which leads to the fact that *social influence* has a very strong effect on patients' decision-making behavior [51]. Moreover, studies have shown that individuals whom patients feel closely related to have a strong impact on a variety of health-related decisions such as the choice of physician, the therapeutic method, or the frequency of medical consultation [52,53]. Beyond that, *social influence* was represented as subjective norm and was already an integral part of the TPB [11] and of subsequent theories from the field of technology acceptance [25]. For this reason, we focused on *social influence* as the independent variable and propose the following hypothesis: a positive *social influence* leads to a higher *behavioral intention* to use PRWs (hypothesis 1).

Figure 1 shows the proposed direct effect of *social influence* on the *behavioral intention* to use PRWs.

In hypotheses 2 to 6, we describe indirect effects, which increase the *behavioral intention* to use PRWs through *credibility* and *performance expectancy* based on *social influence*. Owing to this effect in which *credibility* and *performance expectancy* act as mediator variables, a potential direct effect of *social influence* on the *behavioral intention* to use PRWs should be weakened [54]. This assumption is based on the fact that mediator variables generally affect the direct effect of the independent variable on the dependent variable [55]. As we assume that the impact of *social influence* on the *behavioral intention* to use PRWs can partly be explained through *credibility* and *performance expectancy*, we argue that, in the mediation model, the direct effect between *social influence* and the *behavioral intention* to use PRWs gets weaker compared with the single linear regression (hypothesis 1a).

Figure 1. Direct effect of social influence on the behavioral intention to use physician rating websites. H: hypothesis.

Credibility

The concept of *credibility* describes the level of believability of a transmitter judged by the information perceiver [56]. Perceived *credibility* has an important impact on the whole consumer decision-making process, especially in the case of decisions under conditions of uncertainty [57]. In this context, it was shown that *social influence* not only can change or reinforce the attitude toward new or unknown subjects but also affects subject attributes such as the *credibility* of an information source [58]. Thus, *credibility* is a construct that is strongly controlled by *social influence* [59]. This leads to our second hypothesis: a more positive *social influence* leads to a higher *credibility* of PRWs (hypothesis 2).

In addition to *social influence*, the impact of credibility on behavioral use intention has also been investigated in the past [60]. In the context of technology acceptance, it was shown that credibility exerts a direct impact on the *behavioral intention* to use a new or adapted technology [61]. For this reason, we propose the following: a higher *credibility* leads to a higher *behavioral intention* to use PRWs (hypothesis 3).

Performance Expectancy

With regard to use of new technologies, *performance expectancy* is based on the fact that the use of new systems and the associated change of behavior can lead to an improvement of the current state [25]. This means that the desire and motivation to use and accept a new application or technology increases with the potential benefits derived from its use [62]. As users create individual content on rating portals, the question of *credibility* regarding this content in particular and the evaluation portal as a whole is essential to assess the *performance expectancy* [63]. Therefore, an increase in credibility can lead to web-based portals being perceived as more useful [64,65]. Thus, we expect the following: a higher *credibility* leads to a higher *performance expectancy* toward PRWs (hypothesis 4).

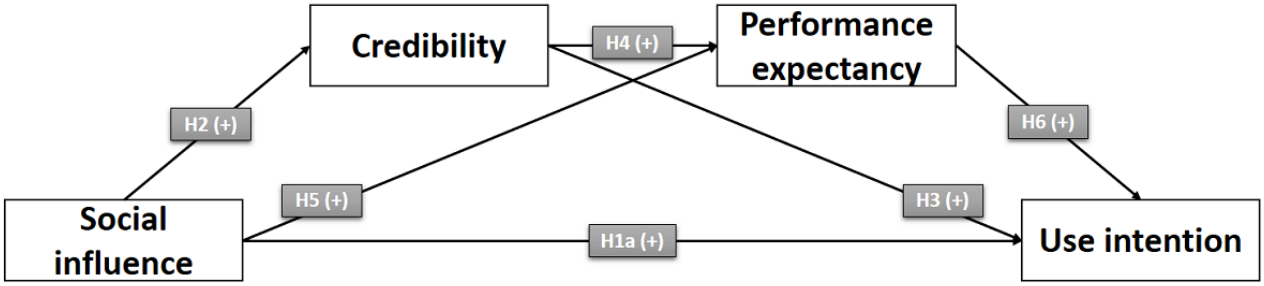
Nevertheless, alternative factors can also exert impact on *performance expectancy*. Fedorko et al [66] extended the UTAUT in the area of electronic banking and demonstrated that social influence had a positive effect on the expected performance. For this reason, the corresponding hypothesis is as follows: a more positive *social influence* leads to a higher *performance expectancy* regarding PRWs (hypothesis 5).

Studies in different fields have shown that *performance expectancy* has a strong effect on the *behavioral intention* to use new or adapted technologies [67]. In line with published results in the field of technology acceptance (eg, Anderson and Schwager [48], Carlsson et al [68], and Marchewka and Kostiwa [69]), we propose the following: a higher *performance expectancy* leads to a higher *behavioral intention* to use PRWs (hypothesis 6).

Figure 2 shows the proposed relationships between the constructs and, thus, the proposed influencing chain.

XSL•FO
RenderX

Figure 2. Conceptual serial mediation model. H: hypothesis.



Moderators of the Impact of Social Influence on the Behavioral Intention to Use PRWs

Age and *gender* are 2 demographic characteristics that have been discovered to affect the *behavioral intention* to use a new or adapted technology [70-73]. For this reason, we propose that the *age* of the participants moderates the effects between *social influence* and the *behavioral intention* to use PRWs as well as between *social influence* and *credibility* of PRWs (hypothesis 7a) and that the *gender* of the participants moderates the effects between *social influence* and the *behavioral intention* to use PRWs as well as between *social influence* and *credibility* of PRWs (hypothesis 7b).

In addition to these demographic characteristics, psychographic characteristics could also moderate the proposed effects. *eHealth literacy* describes the extent to which individuals are able to distinguish useful health-related information on the internet from less useful information [74]. Even though it was shown that *eHealth literacy* seems to exert little influence under certain conditions [75,76], it was also discovered that a high *eHealth literacy* can lead to an increase in the use behavior regarding health-related digital and mobile technologies [77,78]. In the context of PRWs, Schulz and Rothenfluh [79] observed that a higher *eHealth literacy* may mitigate the strength regarding the impact of individual reviews. This result shows that individuals with a high *eHealth literacy* seem to be less impressionable, especially in the medical web-based rating environment. On the basis of these results and assuming that individuals with a high *eHealth literacy* know where to find health-related information on the web, it is expected that a high *eHealth literacy* weakens the effects of our research model. For this reason, we propose that a high *eHealth literacy* weakens the effects between *social influence* and the *behavioral intention* to use PRWs as well as between *social influence* and *credibility* of PRWs (hypothesis 8a).

In addition to *eHealth literacy*, the level of skepticism regarding web-based reviews may also weaken the proposed effects in our research model. Consumer skepticism toward marketing and communication activities has a long research history [80]. Particularly in the digital environment, it has been shown that the level of skepticism toward web-based information can play a significant role in different types of decision-making [81,82]. A distinctive expression of consumer skepticism is *review skepticism*, which can be defined as mistrust toward electronic word of mouth in the context of web-based reviews [83]. In our study, review skepticism was conceptualized as a dispositional form of skepticism and not a situational one [81]. As it can be assumed that individuals who reveal a higher level of *review skepticism* are even more critical toward information on PRWs, we postulate that a high level of *review skepticism* weakens the effects between *social influence* and the *behavioral intention* to use PRWs as well as between *social influence* and *credibility* of PRWs (hypothesis 8b).

To check the hypotheses, a mixed methods approach was applied by conducting 2 studies with different target samples and an experimental as well as cross-sectional study approach. Both of the studies are explained in detail in the following sections.

Study 1

Methods

Study Design and Measures

To test the proposed hypothesized model as depicted in Figure 2, in a first step of our research endeavor, study 1 was conducted by performing a web-based questionnaire–based experiment with a between-subject design. Through randomized experimental manipulation, study participants were assigned to either the experimental group or the control group. After entering sociodemographic data, both groups received the following information: “Physician Rating Websites offer health care consumers the opportunity to evaluate their doctor anonymously. These evaluations could assist future or potential patients in decision-making regarding their future medical care.” In addition to that, the experimental group was asked to imagine that someone who influences their behavior or is important to them or whose opinion is appreciated has recommended the use of PRWs, whereas the control group did not receive this additional information. After that first part, respondents were asked to evaluate their perceived credibility and performance expectancy regarding PRW use as well as their behavioral intention to use PRWs.

The web-based questionnaire used in the experimental setting of study 1 was based on the adoption of established and validated scales [24,26,49,84-88]. The item wording of the questionnaire can be found in Multimedia Appendix 1 [24,26,49,84-88]. All items used were translated and back translated by both an English and a German native speaker who each had fluent language skills in their respective foreign language. To identify potential ambiguity in wording, a pretest with 20 participants was performed. After slight modifications based on the pretest results, the final version of the questionnaire was developed, which was then used for the main study.

XSL•FO
RenderX

Procedure

Data collection was performed using the web-based survey tool Google Forms. Respondents were invited to participate through various web-based channels such as email or social media (snowball sampling), and the survey was conducted over a period of 1 month, from April 15, 2019, to May 14, 2019.

Ethical Considerations

In Austria, there is no requirement to go through an institutional review board or an ethical committee when conducting research with human participants. The questionnaire and study methods adhere to Austrian and European Union privacy laws. The study, as well as the questionnaire, has received clearance from a number of academics and university professors. Participants were properly informed and instructed about their voluntary participation in a web-based survey and were also given the reassurance that their data would be managed with strict confidentiality using acceptable methods, processes, and protocols. Individuals who freely decided to participate in the survey were notified about it in written form before and after they completed the questionnaire. The data were handled in a strictly confidential and anonymous manner.

Data Check

At the beginning of the questionnaire, participants were informed that there were no right or wrong answers and that they would serve the objective of the survey best if they answered the questions honestly to minimize the potential risk of common method bias [89]. Participants were also informed that their information would be handled with complete secrecy using appropriate techniques, processes, and protocols [90].

Measurement Models

SPSS Statistics (version 27; IBM Corp) was used to test the hypotheses. The data were analyzed using linear regressions [91]. In addition to that, regression-based mediation analyses [92] were conducted using the PROCESS macro for SPSS [93]. Of the 92 models included in the PROCESS macro, which are also depicted in the appendix of the corresponding book by Hayes [94], we identified model number 6 ($Y=i_y+c'X+b_1M_1+b_2M_2+e_y$) as the appropriate model for the mediation analysis [92,94]. Furthermore, we included 5000 bootstraps and chose a 95% CI. *Social influence* was defined as independent variable (X), the *behavioral intention* to use PRWs was defined as dependent outcome variable (Y), and *credibility* as well as *performance expectancy* were both conceptualized as mediators (M_1 and M_2).

Results

Analysis of Used Concepts

The Cronbach α ranged from .88 (*performance expectancy*) to .98 (*behavioralintention* to use PRWs) for all multi-item measures. The evaluation of construct means shows a rather high *credibility* (4.04, SD 1.43), *performance expectancy* (4.39, SD 1.55), and *behavioralintention* to use PRWs (4.15, SD 1.96). Table 1 provides a summary of the model construct and measures, including means, SDs, and Cronbach α calculation results.

Table 2 shows how the individual constructs correlate with each other. It is evident that there is a high correlation between *performance expectancy* and *behavioralintention* to use PRWs. In contrast, there is a low correlation between *credibility* and *performance expectancy* as well as between *credibility* and *behavioral intention* to use PRWs [95].

XSL•FO
RenderX

Table 1. Model constructs and measures.

Variable and item	Value, mean (SD)	Cronbach α
Credibility		.96
PRWs[a] seem to be credible	4.54 (1.60)	
PRWs seem to be reliable	4.49 (1.56)	
PRWs seem to be honest	4.34 (1.65)	
PRWs seem to be sincere	4.06 (1.64)	
PRWs seem to be trustworthy	4.01 (1.65)	
PRWs seem to have expert knowledge	3.77 (1.73)	
PRWs seem to be experienced	3.72 (1.60)	
PRWs seem to contain knowledgeable content	3.93 (1.72)	
PRWs seem to be qualified	3.86 (1.65)	
PRWs seem to be knowledgeable	3.73 (1.70)	
Performance expectancy		.88
I think that PRWs are a useful tool	4.73 (1.70)	
By using PRWs, I feel like I have more control over my health	4.22 (1.70)	
Using PRWs will enhance my effectiveness in managing my health care	4.23 (1.80)	
Intention to use PRWs		.98
I intend to use PRWs in the future	4.14 (2.01)	
I will try to use PRWs	4.16 (1.98)	
I plan to use PRWs	4.13 (2.02)	

[a]PRW: physician rating website.

Table 2. Construct correlations[a].

Variable	Value, mean (SD)	Correlations (2-sided; 95% CI)		
		1	2	3
Credibility	4.04 (1.43)	1	0.29 (0.16-0.42)	0.23 (0.10-0.36)
Performance expectancy	4.39 (1.55)	0.29 (0.16-0.42)	1	0.53 (0.42-0.62)
Intention to use PRWs[b]	4.15 (1.96)	0.23 (0.10-0.36)	0.53 (0.42-0.62)	1

[a]All correlations have a *P* value of <.001.

[b]PRW: physician rating website.

Participant Characteristics

A total of 194 participants took part in the study. As the questionnaire was primarily sent out in the university environment, it can be assumed that most study participants were members (students and employees) of a midsized Austrian university. Table 3 provides the sample description.

XSL•FO
RenderX

Table 3. Sample description (N=194).

Sociodemographic characteristics	Participants, n (%)
Sex	
Female	106 (54.6)
Male	88 (45.4)
Age (years)	
20 to 24	31 (16)
25 to 29	18 (9.3)
30 to 34	26 (13.4)
35 to 39	56 (28.9)
40 to 44	15 (7.7)
45 to 49	2 (1)
50 to 54	8 (4.1)
55 to 59	13 (6.7)
≥60	20 (10.3)
Education	
Compulsory education	20 (10.3)
Vocational secondary education	9 (4.6)
Apprenticeship	29 (14.9)
High school	29 (14.9)
University degree	80 (41.2)
No answer	27 (13.9)
Marital status	
Single	43 (22.2)
Close-partnered	68 (35.1)
Married	42 (21.6)
Divorced	5 (2.6)
No answer	36 (18.6)
Occupation	
Salaried employee	102 (52.6)
Unemployed	3 (1.5)
Self-employed	8 (4.1)
In training (pupil or student)	47 (24.2)
Retired	3 (1.5)
No answer	31 (16)
Area of living	
Urban	116 (59.8)
Rural	74 (38.1)
No answer	4 (2.1)

Test of Hypotheses

In hypothesis 1, we propose that a positive *social influence* leads to a higher *behavioral intention* to use PRWs. This relationship could not be confirmed by the data as no direct effect of *social influence* on the *behavioral intention* to use PRWs could be found (b=−0.032; P=.91; SE −0.008; t_1=−0.114). For this reason, hypothesis 1 has to be rejected. Figure 3 shows the direct effect of *social influence* on the *behavioral intention* to use PRWs.

XSL•FO
RenderX

In hypothesis 1a, we assume that, in the mediation model, the direct effect between *social influence* and the *behavioral intention* to use PRWs gets weaker compared with the single linear regression. However, as the direct effect between *social influence* and the *behavioral intention* to use PRWs could not be demonstrated in hypothesis 1, it cannot be assumed that there is a direct effect between the independent and the dependent variable in the mediation model. This assumption is confirmed by the data as there is no significant direct effect between *social influence* and the *behavioral intention* to use PRWs in the mediation model (b=−0.037; P=.88). For this reason, hypothesis 1a has to be rejected.

Hypotheses 2 to 6 describe the indirect effect between *social influence* and the *behavioral intention* to use PRWs through 2 mediators (ie, *credibility* and *performance expectancy*). In this context, hypothesis 2 suggests that a positive *social influence* leads to a higher *credibility* of PRWs. The results referring to this assumption confirm that a positive *social influence* led to an increased *credibility* of PRWs (b=0.572; P=.005). This shows that, for the experimental group, PRWs seemed to be more credible than for the control group. Thus, hypothesis 2 is confirmed.

Referring to hypothesis 3, it is believed that a higher *credibility* also leads to a higher *behavioral intention* to use PRWs. However, this relationship could not be observed (b=−0.123; P=.18). For this reason, hypothesis 3 has to be rejected.

Hypothesis 4 examines the influence of *credibility* on the *behavioral intention* to use PRWs. The analysis in the mediation model confirms the expected relationship. Higher *credibility* led to an increase in *performance expectancy* (b=0.340; P<.001). For this reason, hypothesis 4 is confirmed.

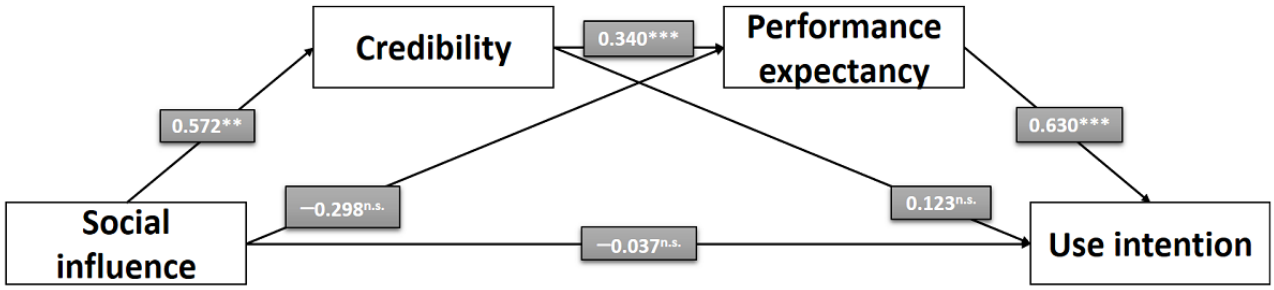
Hypothesis 5 assumes that a stronger *social influence* also leads to a direct increase in *performance expectancy*. This relationship could not be observed (b=−0.298; P=.17). For this reason, hypothesis 5 has to be rejected.

Hypothesis 6 examines the influence of *performance expectancy* on the *behavioral intention* to use PRWs. The analysis in the mediation model confirms the expected relationship. Higher *performance expectancy* led to a substantial increase in the *behavioral intention* to use PRWs (b=0.630; P<.001). Thus, hypothesis 6 is confirmed. Figure 4 shows the mediation model including estimates and P values for the linear regressions. In addition to that, Table 4 shows the detailed outcomes for our proposed mediation model, including model summary, SEs, and 2-tailed t test and P values.

Nevertheless, to check the external validity of the proposed effect chain and replicate the findings, another study was conducted. To assure external validity, another research setting (ie, a cross-sectional study instead of an experimental setting) was applied. In addition, the target sample of the second study should reflect a broader range of individuals, representing the general (web-based) population in a better way, and we intended to use a larger sample size instead of the small sample of 194 largely university members (students and employees) from a southern Austrian university.

Figure 3. Direct effect of social influence on the behavioral intention to use physician rating websites (study 1). *P<.05; **P<.01; ***P<.001; n.s.: not significant.

Figure 4. Serial mediation model (study 1). *P<.05; **P<.01; ***P<.001; n.s.: not significant.



XSL•FO
RenderX

Table 4. Serial mediation model outcomes.

Outcome variable and variable	Coefficient b[a] (SE; 95% CI)	*t* test (*df*)	*P* value
Credibility[b]			
Constant	3.740 (0.147; 3.450 to 4.029)	25.457 (1)	<.001
Social influence	0.572 (0.202; 0.175 to 0.970)	2.838 (1)	.005
Performance expectancy[c]			
Constant	3.176 (0.326; 2.533 to 3.819)	9.743 (2)	<.001
Social influence	−0.298 (0.218; −0.729 to 0.132)	−1.366 (2)	.17
Credibility	0.340 (0.077; 0.189 to 0.491)	4.438 (2)	<.001
Intention to use PRWs[d,e]			
Constant	0.904 (0.450; 0.018 to 1.790)	2.012 (3)	.046
Social influence	−0.037 (0.247; −0.525 to 0.451)	−0.149 (3)	.88
Credibility	0.123 (0.091; −0.056 to 0.301)	1.354 (3)	.18
Performance expectancy	0.630 (0.082; 0.469 to 0.791)	7.730 (3)	<.001

[a]Regression coefficient.

[b]R=0.201; R^2=0.040; P=.005.

[c]R=0.307; R^2=0.095; P<.001.

[d]R=0.532; R^2=0.283; P<.001.

[e]PRW: physician rating website.

Study 2

Methods

Study Design and Measures

In study 2, the constructs were once again measured using existing and validated measures [24,26,49,84-88,96-98]. The questionnaire used in study 2 can be found in Multimedia Appendix 2 [24,26,49,84-88,96-98]. Again, all items were translated and back translated by native English and German speakers who each had fluent language skills in their respective foreign language. The scale options in all variables of interest, apart from the demographic ones, ranged from 1 (“strongly disagree”) to 7 (“strongly agree”).

In contrast to the experimental research applied in study 1, the cross-sectional research approach applied in study 2 is less appropriate to indicate the direction of proposed effects. Thus, the direction of effects is not specified, and study models with alternative effect chains could be set up. However, the use of cross-sectional data is quite common in the analysis of mediation effects as cross-sectional studies are often the only feasible approach for certain topics [99-109] and for targeting larger sample sizes as we intended in study 2. Thus, the methodological approach of our cross-sectional study, which was used in study 2, is in line with existing literature using cross-sectional data to test hypothesized models [99-109]. Both of our studies apply different methodological approaches but lead to comparable results. This agreement in the results supports the validity of the proposed mediation model. However, it has been mentioned previously that cross-sectional data alone may not be suitable to confirm or reject hypotheses proposing a causal direction of effects. To act cautiously and anticipate points of criticism in this vein, we decided not to report in the Results section of study 2 whether hypotheses were confirmed or had to be rejected but to report whether the data were in line with the proposed hypotheses or not.

Procedure

Data collection was carried out in February 2021 and March 2021. The data were gathered using a web-based panel (Clickworker GmbH), which is a research crowdsourcing platform comparable with Amazon’s Mechanical Turk.

Ethical Considerations

The study methodology, questionnaire, and survey instrument adhered to European Union and Austrian privacy laws. The questionnaire did not address any sensitive subjects, and the evaluation process precluded drawing any conclusions about the survey respondents. The questions were kept generic, and there was never any risk of harm from answering them. None of the panel members were asked for any sensitive information, and they all agreed to the data gathering. As mentioned previously, the study was conducted using the crowdsourcing platform Clickworker. This platform has pledged to abide by the General Data Protection Regulation standards and has obtained ISO 27001 certification. Before being approved, all projects and orders must pass an auditing procedure. In this process, professional staff members review for any survey issues, requests for personal information, and instances of discriminatory or unethical content. Orders requesting personal information or containing offensive or unethical content are not accepted. Participants were given the proper information and instructions on their voluntary participation in a web-based survey as well as the knowledge that their data would be handled with utmost secrecy. The processing of the data was completely private and anonymous.

XSL•FO
RenderX

Data Check

The constructs included in study 2 were self-reported measures that are associated with the risk of common method bias [110]. To minimize the potential risk of common method bias, we informed participants that there were no right or wrong answers and that they would serve the purpose of the survey best if they answered the questions as honestly as possible [89]. In addition to that, participants were informed that their data would be treated with absolute confidentiality using suitable methods, procedures, and protocols [90].

Furthermore, we ran multiple pretests and eliminated unclear or imprecise data sets [111]. Using the survey tool LimeSurvey (LimeSurvey GmbH), the data were thoroughly checked for inconsistent answer patterns, flatliners, and very short answer times. In this context, we eliminated questionnaires from respondents who used descending or ascending numerical sequences for the items throughout the questionnaire (inconsistent answer patterns), consistently responded with the same answer (flatliners), or completed the questionnaire in <265 seconds (very short answer times). The minimum response time for answering the questionnaire was pretested by the authors.

By incorporating 3 security levels in the web-based questionnaire, we followed guidelines to reduce validity concerns when using crowdsourcing platforms (eg, see Aguinis et al [112]). First, logic tasks and attention tests were used to verify that survey participants' attention and integrity were maintained. Participants had to solve a mathematical equation to verify that they were human and that they were eligible before they could begin answering the questions. In addition to that, there was a check for attention. See Multimedia Appendix 3 for an example equation and the attention check. To verify that participants had read the introductory text of the third question group, they had to pick a specific answer choice in a specific question as part of this attention check (see Oppenheimer et al [113] and Kung et al [114]). Finally, each participant was assigned a cookie to prevent them from taking part again.

There were no missing data in the survey as the questionnaire instrument was not set up to allow for unanswered questions.

Measurement Models

SPSS Statistics (version 27; IBM Corp) was used to test the hypotheses. The data were analyzed using linear regressions [91]. In addition to that, regression-based mediation analyses [92] were conducted using the PROCESS macro for SPSS [93]. Of the 92 models included in the PROCESS macro, which are also depicted in the appendix of the corresponding book by Hayes [93], we again identified model number 6 ($Y=i_y+c'X+b_1M_1+b_2M_2+e_y$) for the mediation analysis and model number 8 ($Y=i_y+c'_1X+c'_2W+c'_3XW+b_1M_1+b_2M_2+e_y$) for the moderated mediation analysis [92,94] as the appropriate models in our case. Furthermore, we included 5000 bootstraps and chose a 95% CI. *Social influence* was defined as independent variable (X), the *behavioral intention* to use PRWs was defined as dependent outcome variable (Y), and *credibility* as well as *performance expectancy* both were conceptualized as mediators (M_1 and M_2). *Age*, *gender*, and *eHealth literacy* as well as *review skepticism* acted as moderator variables in our models. The analyses were performed in a hierarchical order starting with the basic mediation model. After that, the proposed moderators were included one after the other.

Results

Analysis of Used Concepts

The Cronbach α ranged from.79 (*performance expectancy*) to.97 (*behavioralintention* to use PRWs) for all multi-item measures. The evaluation of construct means shows a rather high *credibility* (4.44, SD 1.11) and *behavioralintention* to use PRWs (4.14, SD 1.90). The mean value of *performance expectancy* (3.62, SD 1.26) is slightly above the midpoint of the scale. However, on average, *social influence* (2.47, SD 1.56) in the domain of PRWs in real life (as opposed to the experimental manipulation of social influence in study 1) seems to be rather low. Table 5 provides a summary of the model construct and measures, including means, SDs, and Cronbach α calculation results.

Table 6 shows how the individual constructs correlate with each other. It is evident that there is a high correlation between *performance expectancy* and *behavioralintention* to use PRWs. By contrast, there is a low correlation between *social influence* and *credibility*. The remaining constructs are characterized by a medium correlation with each other [95].

XSL•FO
RenderX

Table 5. Model constructs and measures.

Variable and item	Value, mean (SD)	Cronbach α
Social influence	2.47 (1.56)	.96
People who influence my behavior think that I should use PRWs[a]	2.39 (1.57)	
People who are important to me think that I should use PRWs	2.49 (1.61)	
People whose opinion I value think that I should use PRWs	2.53 (1.66)	
Credibility	4.44 (1.11)	.92
PRWs seem to be credible	4.50 (1.76)	
PRWs seem to be reliable	4.37 (1.22)	
PRWs seem to be trustworthy	4.44 (1.21)	
Performance expectancy	3.62 (1.26)	.79
I think that PRWs are a useful tool	4.99 (1.50)	
By using PRWs, I feel like I have more control over my health	3.14 (1.71)	
Using PRWs will enhance my effectiveness in managing my health care	3.64 (1.67)	
Overall, PRWs will be useful in managing my health care	2.70 (1.55)	
Intention to use PRWs	4.14 (1.90)	.97
I intend to use PRWs in the future	4.24 (1.97)	
I will try to use PRWs	4.18 (1.92)	
I plan to use PRWs	4.00 (2.00)	

[a]PRW: physician rating website.

Table 6. Construct correlations[a].

Variable	Value, mean (SD)	Correlations (2-sided; 95% CI)			
		1	2	3	4
Social influence	2.47 (1.56)	1	0.29 (0.21-0.37)	0.49 (0.43-0.56)	0.41 (0.34-0.48)
Credibility	4.44 (1.11)	0.29 (0.21-0.37)	1	0.50 (0.42-0.56)	0.47 (0.40-0.54)
Performance expectancy	3.62 (1.26)	0.50 (0.43-0.56)	0.50 (0.43-0.56)	1	0.59 (0.53-0.64)
Intention to use PRWs[b]	4.14 (1.90)	0.41 (0.34-0.48)	0.47 (0.40-0.54)	0.59 (0.53-0.64)	1

[a]All correlations have a *P* value of <.001.

[b]PRW: physician rating websites.

Participant Characteristics

A total of 852 participants from Austria took part in the study, with 334 (39.2%) of them being eliminated as they were not able to pass the manipulation check (239/334, 71.6%) or were characterized by an implausible response behavior or insufficient answer time (95/334, 28.4%). See Multimedia Appendix 4 for a graphical data cleansing description. This data cleansing mechanism resulted in a total of 518 survey participants who form the calculation sample for this study. Table 7 provides the sample description.

XSL•FO
RenderX

Table 7. Sample description (N=518).

Sociodemographic characteristics	Participants, n (%)
Sex	
Female	289 (55.8)
Male	227 (43.8)
Intersex	2 (0.4)
Age (years)	
15 to 19	62 (12)
20 to 24	141 (27.2)
25 to 29	101 (19.5)
30 to 34	75 (14.5)
35 to 39	64 (12.4)
40 to 44	32 (6.2)
45 to 49	14 (2.7)
50 to 54	13 (2.5)
55 to 59	8 (1.5)
≥60	8 (1.5)
Education	
Compulsory education	27 (5.2)
Vocational secondary education	55 (10.6)
Apprenticeship	72 (13.9)
High school	214 (41.3)
University degree	150 (29)
Marital status	
Single	203 (39.2)
Close-partnered	218 (42.1)
Married	87 (16.8)
Divorced	10 (1.9)
Occupation	
Salaried employee	245 (47.3)
Unemployed	43 (8.3)
Self-employed	47 (9.1)
In training (pupil or student)	175 (33.8)
Retired	8 (1.5)
Area of living	
Urban	322 (62.2)
Rural	196 (37.8)

Test of Hypotheses

In hypothesis 1, we propose that a positive social influence leads to a higher behavioral intention to use PRWs. This relationship is in line with the data as it was shown that respondents whose social environment influenced them to a greater extent to use those websites seemed to have a higher behavioral intention to use PRWs than respondents who were less influenced to use PRWs by their social environment (b=0.503; P<.001; SE 0.049; t_1=10.197). Thus, the data were in line with hypothesis 1. Figure 5 shows the direct effect of social influence on the behavioral intention to use PRWs. However, in hypothesis 1a, we assume that the impact of *social influence* on *behavioral intention* to use PRWs is partly explained by the mediator variables *credibility* and *performance expectancy*. Therefore, this conclusion would lead to the direct impact of *social influence*

XSL•FO
RenderX

on the *behavioral intention* to use PRWs being weaker in the mediation model than in the simple linear regression. In this context, the data are in line with hypothesis 1a as the direct effect of *social influence* on the *behavioral intention* to use PRWs in the mediation model was weaker (*b*=0.177; *P*<.001) than the effect measured in the simple linear regression (*b*=0.410; *P*<.001). On the basis of these results, the data were in line with hypothesis 1a.

Hypotheses 2 to 6 describe the indirect effect between *social influence* and the *behavioral intention* to use PRWs through 2 mediators (ie, *credibility* and *performance expectancy*). In this context, hypothesis 2 suggests that a positive *social influence* leads to a higher *credibility* of PRWs. The results referring to this assumption suggest that, for respondents whose social environment influenced them to a greater extent to use those websites, PRWs seemed to be more credible than for respondents who were less influenced to use PRWs by their social environment (*b*=0.208; *P*<.001). Thus, the data are in line with hypothesis 2.

Hypothesis 3 assumes that higher *credibility* also increases the *behavioral intention* to use PRWs. Study participants who indicated that PRWs were more credible also seemed to have a higher *behavioral intention* to use PRWs (*b*=0.402; *P*<.001). For this reason, hypothesis 3 is in line with the data.

Referring to hypothesis 4, it is believed that a higher *credibility* also leads to a higher *performance expectancy* toward PRWs. Results show that respondents reporting a higher *credibility* regarding PRWs also reported a higher *performance expectancy* toward PRWs (*b*=0.431; *P*<.001). On the basis of these results, hypothesis 4 is in line with the data.

Hypothesis 5 assumes that a stronger *social influence* also leads to a direct increase in *performance expectancy*. In this context, it could be observed that respondents whose social environment influenced them to a greater extent to use those websites also reported a higher *performance expectancy* regarding PRWs (*b*=0.312; *P*<.001). For this reason, hypothesis 5 is in line with the data.

Hypothesis 6 examines the influence of *performance expectancy* on the *behavioral intention* to use PRWs. The analysis in the mediation model showed that respondents who reported a higher *performance expectancy* regarding PRWs also reported a higher *behavioral intention* to use PRWs (*b*=0.605; *P*<.001). Thus, hypothesis 6 is in line with the data. Figure 6 shows the mediation model, including estimates and *P* values for the linear regressions. In addition to that, Table 8 shows the detailed outcomes for our proposed mediation model, including model summary, SEs, and *t* test and *P* values.

Hypotheses 7 and 8 control for potential moderator effects in our model. In hypothesis 7a, we suggest that the *age* of the participants affects the effects between *social influence* and the *behavioral intention* to use PRWs as well as between *social influence* and *credibility* of PRWs. To test this hypothesis, we created a subsample consisting of 2 age groups. Respondents aged <39 years were assigned to the younger age group (443/518, 85.5%), and respondents aged ≥39 years were assigned to the older age group (75/518, 14.5%). No significant interaction effects could be demonstrated when examining *age* as a potential moderator variable. For this reason, hypothesis 7a had to be rejected. However, it could be shown that higher *age* has a negative influence on *performance expectancy* (*b*=−0.501; *P*=.04). Table 9 summarizes the results of the corresponding moderated mediation model.

In hypothesis 7b, we suggest that the *gender* of the participants affects the effects between *social influence* and the *behavioral intention* to use PRWs as well as between *social influence* and *credibility* of PRWs. To examine this assumption, we excluded participants who reported that they belonged to the *intersex* category. The exclusion was made as the response rate of this population group was very low (2/518, 0.4%). No significant interaction effects could be demonstrated when examining *gender* as a potential moderator variable. For this reason, hypothesis 7b had to be rejected. Table 10 summarizes the results of the corresponding moderated mediation model.

In hypothesis 8a, we suggest that a high level of *eHealth literacy* weakens the effects between *social influence* and the *behavioral intention* to use PRWs as well as between *social influence* and *credibility* of PRWs. Accordingly, a high *eHealth literacy* should weaken the direct as well as indirect effect in the moderated mediation model. Even though in the model it was shown that *eHealth literacy* had a positive effect on *credibility* (*b*=0.266; *P*=.002), this effect could not be observed between *eHealth literacy* and the *behavioral intention* to use PRWs (*b*=0.173; *P*=.17). In addition to that, the proposed interaction effects could also not be observed as *eHealth literacy* did not exert significant influence on the effect between *social influence* and *credibility* (*b*=0.018; *P*=.55) or the effect between *social influence* and the *behavioral intention* to use PRWs (*b*=0.005; *P*=.92). For this reason, hypothesis 8a had to be rejected. Table 11 summarizes the results of the corresponding moderated mediation model.

Finally, in hypothesis 8b, we suggest that a high level of *review skepticism* weakens the effects between *social influence* and the *behavioral intention* to use PRWs as well as between *social influence* and *credibility* of PRWs. Accordingly, a high level of *review skepticism* should weaken the direct as well as indirect effect in the moderated mediation model. Even though in the model it was shown that r*eview skepticism* had a negative effect on *credibility* (*b*=−0.254; *P*<.001), this effect could not be observed between *review skepticism* and the *behavioral intention* to use PRWs (*b*=−0.111; *P*=.24). In addition to that, the proposed interaction effects could also not be observed as *review skepticism* did not exert significant influence on the effect between *social influence* and *credibility* (*b*=0.025; *P*=.25) or the effect between *social influence* and the *behavioral intention* to use PRWs (*b*=0.017; *P*=.58). For this reason, hypothesis 8b had to be rejected. Table 12 summarizes the results of the corresponding moderated mediation model.

XSL•FO
RenderX

Figure 5. Direct effect of social influence on the behavioral intention to use physician rating websites (study 2). *P<.05; **P<.01; ***P<.001; n.s.: not significant.

Figure 6. Serial mediation model (study 2). *P<.05; **P<.01; ***P<.001; n.s.: not significant.

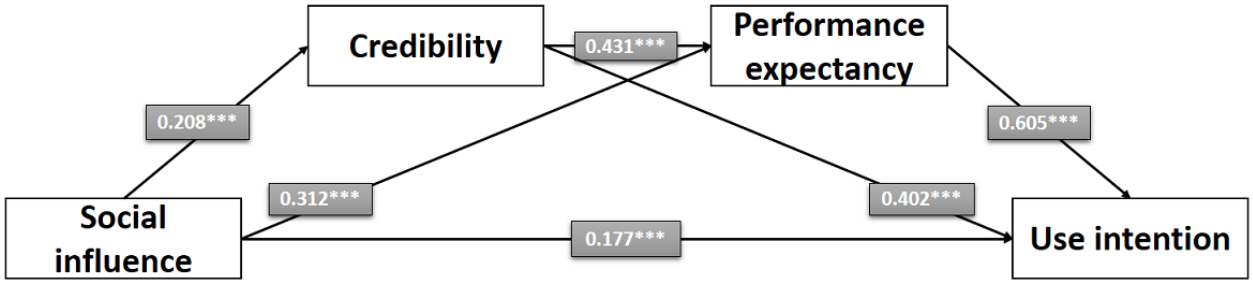


Table 8. Serial mediation model outcomes.

Outcome variable	Coefficient b[a] (SE; 95% CI)	*t* test (*df*)	*P* value
Credibility[b]			
Constant	3.925 (0.088; 3.751 to 4.098)	44.393 (1)	<.001
Social influence	0.208 (0.030; 0.148 to 0.267)	6.845 (1)	<.001
Performance expectancy[c]			
Constant	0.935 (0.181; 0.580 to 1.290)	5.172 (2)	<.001
Social influence	0.312 (0.030; 0.254 to 0.370)	10.590 (2)	<.001
Credibility	0.431 (0.041; 0.351 to 0.512)	10.516 (2)	<.001
Intention to use PRWs[d,e]			
Constant	−0.275 (0.274; −0.814 to 0.264)	−1.003 (3)	.32
Social influence	0.177 (0.048; 0.082 to 0.271)	3.665 (3)	<.001
Credibility	0.402 (0.067; 0.271 to 0.533)	6.012 (3)	<.001
Performance expectancy	0.605 (0.065; 0.477 to 0.734)	9.283 (3)	<.001

[a]Regression coefficient.
[b]R=0.289; R^2=0.083; P<.001.
[c]R=0.615; R^2=0.378; P<.001.
[d]R=0.637; R^2=0.406; P<.001.
[e]PRW: physician rating website.

XSL•FO
RenderX

Table 9. Results of the moderated mediation analyses with age as moderator.

Outcome variable	Coefficient b[a] (SE; 95% CI)	*t* test (*df*)	*P* value
Credibility[b]			
Constant	3.992 (0.097; 3.802 to 4.182)	41.299 (3)	<.001
Social influence	0.194 (0.033; 0.130 to 0.258)	5.952 (3)	<.001
Age	−0.383 (0.239; −0.853 to 0.087)	−1.600 (3)	.11
Interaction: (social influence × age)	0.071 (0.091; −0.108 to 0.249)	0.777 (3)	.44
Performance expectancy[c]			
Constant	2.714 (0.099; 2.520 to 2.908)	27.438 (3)	<.001
Social influence	0.383 (0.033; 0.318 to 0.449)	11.504 (3)	<.001
Age	−0.501 (0.245; −0.982 to −0.020)	–2.045 (3)	.04
Interaction: (social influence × age)	0.098 (0.093; −0.085 to 0.281)	1.054 (3)	.29
Intention to use PRWs[d,e]			
Constant	−0.124 (0.285; −0.682 to 0.435)	−0.435 (5)	.66
Social influence	0.152 (0.050; 0.053 to 0.251)	3.021 (5)	.003
Credibility	0.397 (0.067; 0.266 to 0.528)	5.942 (5)	<.001
Performance expectancy	0.596 (0.065; 0.468 to 0.725)	9.135 (5)	<.001
Age	−0.650 (0.331; −1.300 to 0.001)	−1.961 (5)	.05
Interaction: (social influence × age)	0.189 (0.125; −0.058 to 0.435)	1.503 (5)	.13

[a]Regression coefficient.

[b]R=0.299; R^2=0.090; P<.001.

[c]R=0.503; R^2=0.253; P<.001.

[d]R=0.640; R^2=0.410; P<.001.

[e]PRW: physician rating website.

XSL•FO
RenderX

Table 10. Results of the moderated mediation analyses with gender as moderator.

Outcome variable	Coefficient b[a] (SE; 95 CI)	*t* test (*df*)	*P* value
Credibility[b]			
Constant	3.916 (0.271; 3.384 to 4.448)	14.458 (3)	<.001
Social influence	0.222 (0.096; 0.034 to 0.410)	2.317 (3)	.02
Gender	0.002 (0.181; −0.353 to 0.357)	0.011 (3)	.99
Interaction: (social influence × gender)	−0.009 (0.061; −0.129 to 0.112)	−0.145 (3)	.88
Performance expectancy[c]			
Constant	2.539 (0.278; 1.992 to 3.085)	9.125 (3)	<.001
Social influence	0.454 (0.098; 0.260 to 0.647)	4.608 (3)	<.001
Gender	0.058 (0.186; −0.306 to 0.422)	0.313 (3)	.75
Interaction: (social influence × gender)	−0.034 (0.063; −0.158 to 0.090)	−0.542 (3)	.59
Intention to use PRWs[d,e]			
Constant	0.250 (0.445; −0.623 to 1.124)	0.563 (5)	.57
Social influence	0.169 (0.134; −0.096 to 0.420)	−1.233 (5)	.22
Credibility	0.407 (0.065; 0.280 to 0.535)	6.255 (5)	<.001
Performance expectancy	0.599 (0.062; 0.476 to 0.721)	9.609 (5)	<.001
Gender	−0.364 (0.240; −0.836 to 0.108)	−1.514 (5)	.13
Interaction: (social influence × gender)	0.078 (0.083; −0.144 to 0.180)	0.216 (5)	.83

[a]Regression coefficient.

[b]R=0.290; R^2=0.084; P<.001.

[c]R=0.495; R^2=0.245; P<.001.

[d]R=0.643; R^2=0.413; P<.001.

[e]PRW: physician rating website.

XSL•FO
RenderX

Table 11. Results of the moderated mediation analyses with eHealth literacy as moderator.

Outcome variable	Coefficient b[a] (SE; 95% CI)	*t* test (*df*)	*P* value
Credibility[b]			
Constant	2.463 (0.490; 1.501 to 3.425)	5.031 (3)	<.001
Social influence	0.111 (0.170; −0.224 to 0.445)	0.649 (3)	.52
eHealth literacy	0.266 (0.087; 0.095 to 0.438)	3.056 (3)	.002
Interaction: (social influence × eHealth literacy)	0.018 (0.030; −0.042 to 0.078)	0.592 (3)	.55
Performance expectancy[c]			
Constant	1.920 (0.519; 0.901 to 2.940)	3.700 (3)	<.001
Social influence	0.342 (0.181; −0.012 to 0.698)	1.899 (3)	.06
eHealth literacy	0.129 (0.092; −0.053 to 0.310)	1.394 (3)	.17
Interaction: (social influence × eHealth literacy)	0.011 (0.032; −0.052 to 0.074)	0.338 (3)	.74
Intention to use PRWs[d,e]			
Constant	−1.030 (0.719; −2.442 to 0.382)	−1.433 (5)	.15
Social influence	0.165 (0.244; −0.315 to 0.644)	0.674 (5)	.50
Credibility	0.355 (0.069; 0.220 to 0.490)	5.158 (5)	<.001
Performance expectancy	0.601 (0.065; 0.474 to 0.729)	9.267 (5)	<.001
eHealth literacy	0.173 (0.126; −0.074 to 0.420)	1.378 (5)	.17
Interaction: (social influence × eHealth literacy)	0.005 (0.043; −0.081 to 0.090)	0.103 (5)	.92

[a]Regression coefficient.

[b]R=0.399; R^2=0.159; P<.001.

[c]R=0.509; R^2=0.260; P<.001.

[d]R=0.643; R^2=0.414; P<.001.

[e]PRW: physician rating website.

XSL•FO
RenderX

Table 12. Results of the moderated mediation analyses with review skepticism as moderator.

Outcome variable	Coefficient b[a] (SE; 95% CI)	*t* test (*df*)	*P* value
Credibility[b]			
Constant	5.012 (0.288; 4.447 to 5.579)	17.403 (3)	<.001
Social influence	0.086 (0.092; −0.094 to 0.266)	0.936 (3)	.35
Review skepticism	−0.254 (0.066; −0.384 to 0.124)	−3.836 (3)	<.001
Interaction: (social influence × review skepticism)	0.025 (0.022; −0.018 to 0.068)	1.143 (3)	.25
Performance expectancy[c]			
Constant	3.158 (0.303; 2.564 to 3.752)	10.440 (3)	<.001
Social influence	0.307 (0.096; 0.118 to 0.496)	3.188 (3)	.002
Review skepticism	−0.126 (0.069; −0.262 to 0.011)	−1.809 (3)	.07
Interaction: (social influence × review skepticism)	0.022 (0.023; −0.023 to 0.067)	0.954 (3)	.34
Intention to use PRWs[d,e]			
Constant	0.277 (0.518; −0.740 to 1.294)	0.535 (5)	.59
Social influence	0.104 (0.131; −0.153 to 0.362)	0.797 (5)	.43
Credibility	0.382 (0.069; 0.247 to 0.516)	5.570 (5)	<.001
Performance expectancy	0.606 (0.065; 0.478 to 0.734)	9.287 (5)	<.001
Review skepticism	−0.111 (0.095; −0.298 to 0.075)	−1.174 (5)	.24
Interaction: (social influence × review skepticism)	0.017 (0.031; −0.044 to 0.078)	0.555 (5)	.58

[a]Regression coefficient.

[b]R=0.365; R^2=0.133; P<.001.

[c]R=0.501; R^2=0.251; P<.001.

[d]R=0.639; R^2=0.408; P<.001.

[e]PRW: physician rating website.

Discussion

Principal Findings

On the basis of the results of studies 1 and 2, *social influence* exerts a statistically significant impact on the *behavioral intention* to use PRWs [115,116]. However, in study 1, it was shown that this impact might only be exerted indirectly through the 2 mediator variables *credibility* and *performance expectancy*. Notwithstanding, when we tested the proposed chain of effects in study 2, we were able to reveal 2 further findings. On the one hand, we found a direct effect of *social influence* on the *behavioral intention* to use PRWs. However, this direct effect between *social influence* and the *behavioral intention* to use PRWs was significantly weakened in the mediation model. This result suggests that the direct effect between the independent and the dependent variable is at least partially explained by the 2 mediator variables [92,117]. Furthermore, the proposed indirect effect itself could again be observed in a significant expression.

In contrast to the successfully predicted direct and indirect effects between *social influence* and the *behavioral intention* to use PRWs, the proposed moderation effects could not be observed in our moderated mediation model. Even though *eHealth literacy* and *review skepticism* seem to affect the *credibility* of PRWs, we were not able to observe a significant effect of *age*, *gender*, *eHealth literacy*, or *review skepticism* on the proposed mediation model.

These findings strongly support theories from the social psychological perspective, theories of technology acceptance, and theories from the economic perspective. From a social psychological perspective, the results of the 2 studies support both the TRA and its extension, the TPB. Both theories serve as a basis for the factor model studied. However, the TRA can also generally be reconstructed by means of the factors investigated. As described in the Introduction, *social influence* is a continuation of *subjective norms* [25]. *Credibility* as well as *performance expectancy* may be considered as determinants of *attitude* toward behavior, and *behavioral intention* is a construct that is incorporated as a result of *attitudes* and *subjective norms* within the TRA [9]. The 2 studies also show a generally similar influencing chain. Thus, the results are in line with a large number of empirical studies in the field of social psychological theories (eg, Gotch and Hall [118], Fishbein [119], Ajzen et al [120], Lada et al [121], and Buttle and Bok [122]) and provide evidence for the validity of the TRA.

XSL•FO
RenderX

The results of the 2 studies also support theories of technology acceptance as it was shown that *social influence* affects the *behavioral intention* to use PRWs indirectly (studies 1 and 2) and directly (study 2). Furthermore, the impact of *performance expectancy* on the dependent variable could also be demonstrated in the factor model. Even though not all technology acceptance variables were considered in the factor model, these studies still provide evidence for the validity of the respective influencing paths included in the UTAUT.

Finally, the findings of both studies strongly support the TIE as they show that individuals in our study samples built their decision for or against the *behavioral intention* to use PRWs on *social influence* through *credibility* and *performance expectancy*. This outcome indicates that individuals may base their decision for or against the use of these web-based platforms at least partly on factors other than the performance quality of PRWs. In addition to that, our results support the similarity effect as we were able to show that advice from a person with a high similarity could lead to an increased *behavioral intention* to use PRWs.

Furthermore, a path model in the form examined in this study has never been tested in the context of PRWs. Therefore, the added value of the studies also lies in developing practical implications from the relationships that could specifically increase the degree of use of PRWs (eg, through the targeted use of social influence by social media influencers [SMIs]).

Limitations

Despite the significance of the findings, a number of limitations need to be considered. Particularly in web-based surveys, it is to be expected that research participants might be disinterested or respond in a one-sided manner. However, we tried our best to address and solve the issue of a possible common method bias. We attempted to minimize the risk of a potential common method bias by conducting information processes before and during the completion of the questionnaire by study participants. Furthermore, participants in a web-based survey may have a more thorough grasp of web-based issues. This might have resulted in a more prominent representation of the *behavioral intention* to use PRWs in the study population. In addition to that, by focusing on social influence as the only independent variable, we disregarded a number of alternative influencing variables. This may have led to a disproportionate impact of social influence on our study model. Finally, as was already explained in detail in the Methods section of study 2, it should be noted that a cross-sectional study, in contrast to the experimental study design of study 1, does not specify a direction of effects. In general, this means that the effect of the factors on each other could also be different from that proposed in the study model. An explanation could be that individuals tend to associate with people who share their attitudes and viewpoints (eg, see Bos et al [123]). Another explanation could be that people tend to exaggerate the degree to which their opinions and those of others are similar (eg, see Dunning et al [124]). We have made several attempts to address this criticism. First, the factor model of this study was built on established theories of social psychology and technology acceptance. In addition, we used a 2-step procedure to check the appropriateness of the proposed influencing chain. In a first step, we conducted a study applying an experimental setting to test the meaningfulness of the causal model. The cross-sectional study was the second step of our research endeavor. Although this approach lends considerable credence to the study model, there is still a certain residual risk that the influence paths of the integrated factors are not as interrelated as suggested.

Conclusions and Practical Implications

The aim of the bipartite research endeavor was to investigate if and how *social influence* affects the *behavioral intention* to use PRWs. In study 1, the proposed indirect effect between *social influence* and the *behavioral intention* to use PRWs could be demonstrated. Moreover, in study 2, almost all of our hypotheses were in line with the data. The proposed serial mediation model provides evidence for the validity of both the TRA and the UTAUT. Moreover, we were able to observe the proposed similarity effect as a positive *social influence* led to a higher *credibility* of PRWs in both studies. In this context, the TIE can serve as a profound theoretical framework in explaining the relationships between the constructs. By categorizing health care services as credence goods, this theoretical approach can make valuable contributions in explaining the impact of *social influence* on the *credibility* of PRWs as well as on the *performance expectancy* and *behavioral intention* to use PRWs. The most obvious finding to emerge from both studies is that *social influence* seems to exert an impact on the *behavioral intention* to use PRWs. However, in particular, we showed that, under certain conditions, this impact seems not to be exerted directly but indirectly through *credibility* and *performance expectancy*. To sum up, the evidence from both studies suggests that *social influence* could increase the use rates of PRWs enormously. Bearing in mind that SMIs develop a kind of parasocial relationship with their followers [125,126], it might be conceivable that *social influence* is not just limited to friends and family but could also be exerted by SMIs in the domain of PRW use. From a marketing perspective, PRW providers could think of collaborating with SMIs to boost use of PRWs in the future. With the onset of the ongoing pandemic, and especially in times of lockdowns and reduced personal contacts, SMIs have increasingly taken on the role of a kind of “homefluencers” [127], especially with regard to specific health-related issues such as vaccination in general [128]. Thus, the follower base of SMIs could also be used as a target group of electronic word of mouth to increase use of PRWs in the long run. However, special attention should be paid to choosing the most suitable SMI according to the fit between their personality, their follower base, and the specific PRW provider, as it has been widely investigated in commercial realms such as brand relationships [129-131]. Increased use of PRWs could be advantageous not only for PRW providers but also for patients and physicians. Higher PRW use could lead to a higher average number of ratings per physician, which could increase their representativeness [132]. From the physicians’ perspective, PRWs enable them to achieve a positive external impact and learn from feedback and offer them avenues to improve their service quality [133].

XSL•FO
RenderX

Directions for Future Research

This study investigated not only whether *social influence* exerts an impact on the *behavioral intention* to use PRWs but also how this influence emerges. However, when focusing on this crucial independent variable, several other possible influencing variables could be interesting as well. These include, in line with the UTAUT [25], for example, *effort expectancy*, *facilitating conditions*, *hedonic motivation*, or *habit* [24,26,49,134]. In future investigations, it might be possible to use additional moderator variables (eg, the area where people live [rural vs urban]). To sum up, further elaboration on the influencing chain to explain behavioral intention to use PRWs by including other variables of interest could be an important issue for future research.

Acknowledgments

The authors reported that there is no funding associated with the work featured in this paper.

Authors' Contributions

BG contributed to study design, data collection, data analysis, and writing of the paper. SB contributed to conceptualization, study design, and data collection and critically reviewed the paper. Both authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire for study 1.
[PDF File (Adobe PDF File), 249 KB - jmir_v24i11e37505_app1.pdf]

Multimedia Appendix 2
Questionnaire for study 2.
[PDF File (Adobe PDF File), 289 KB - jmir_v24i11e37505_app2.pdf]

Multimedia Appendix 3
Example equation and attention check.
[PDF File (Adobe PDF File), 60 KB - jmir_v24i11e37505_app3.pdf]

Multimedia Appendix 4
Graphical data cleansing description.
[PDF File (Adobe PDF File), 39 KB - jmir_v24i11e37505_app4.pdf]

References

1. Emmert M, McLennan S. One decade of online patient feedback: longitudinal analysis of data from a German physician rating website. J Med Internet Res 2021 Jul 26;23(7):e24229 [FREE Full text] [doi: 10.2196/24229] [Medline: 34309579]
2. Bäumer F, Kersting J, Kuršelis V, Geierhos M. Rate your physician: findings from a Lithuanian physician rating website. In: Information and Software Technologies. Cham: Springer; 2018.
3. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
4. McLennan S, Strech D, Kahrass H. Why are so few patients rating their physicians on German physician rating websites? A qualitative study. BMC Health Serv Res 2018 Aug 29;18(1):670 [FREE Full text] [doi: 10.1186/s12913-018-3492-0] [Medline: 30157842]
5. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. JAMA 2017 Feb 21;317(7):766-768 [FREE Full text] [doi: 10.1001/jama.2016.18553] [Medline: 28241346]
6. Hill RJ, Fishbein M, Ajzen I. Belief, attitude, intention and behavior: an introduction to theory and research. Contemporary Sociol 1977 Mar;6(2):244. [doi: 10.2307/2065853]
7. Ajzen I. From intentions to actions: a theory of planned behavior. In: Action Control. Berlin, Heidelberg: Springer; 1985.
8. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Hoboken, New Jersey, United States: Prentice-Hall; 1980.
9. Fishbein M. A theory of reasoned action: some applications and implications. Nebr Symp Motiv 1980;27:65-116. [Medline: 7242751]

XSL•FO
RenderX

10. Sheppard BH, Hartwick J, Warshaw PR. The theory of reasoned action: a meta-analysis of past research with recommendations for modifications and future research. J Consum Res 1988 Dec;15(3):325. [doi: 10.1086/209170]
11. 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]
12. Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. J Experimental Social Psychol 1986 Sep;22(5):453-474. [doi: 10.1016/0022-1031(86)90045-4]
13. Madden TJ, Ellen PS, Ajzen I. A comparison of the theory of planned behavior and the theory of reasoned action. Pers Soc Psychol Bull 2016 Jul 02;18(1):3-9. [doi: 10.1177/0146167292181001]
14. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]
15. King WR, He J. A meta-analysis of the technology acceptance model. Inform Manag 2006 Sep;43(6):740-755. [doi: 10.1016/j.im.2006.05.003]
16. Surendran P. Technology acceptance model: a survey of literature. Int J Bus Social Res 2012;2(4). [doi: 10.18533/ijbsr.v2i4.161]
17. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014 Feb 16;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
18. Mathieson K. Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. Inform Syst Res 1991 Sep;2(3):173-191. [doi: 10.1287/isre.2.3.173]
19. Davis F. A technology acceptance model for empirically testing new end-user information systems : theory and results. Massachusetts Institute of Technology. 1986. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2022-02-16]
20. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. Int J Man Mach Stud 1993 Mar;38(3):475-487. [doi: 10.1006/imms.1993.1022]
21. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
22. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
23. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decision Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
24. Venkatesh V, Zhang X. Unified theory of acceptance and use of technology: U.S. vs. China. J Global Inform Technol Manag 2014 Sep 09;13(1):5-27. [doi: 10.1080/1097198X.2010.10856507]
25. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
26. Venkatesh V, Thong JJY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
27. Chang A. UTAUT and UTAUT 2: a review and agenda for future research. The Winners 2012 Sep 28;13(2):10. [doi: 10.21512/tw.v13i2.656]
28. Venkatesh V, Thong J, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
29. Williams MD, Rana NP, Dwivedi YK. The unified theory of acceptance and use of technology (UTAUT): a literature review. J Ent Inform Manag 2015 Apr 13;28(3):443-488. [doi: 10.1108/JEIM-09-2014-0088]
30. Darby MR, Karni E. Free competition and the optimal amount of fraud. J Law Econ 1973 Apr;16(1):67-88. [doi: 10.1086/466756]
31. Simons H, Berkowitz N, Moyer R. Similarity, credibility, and attitude change: a review and a theory. Psychol Bulletin 1970 Jan;73(1):1-16. [doi: 10.1037/h0028429]
32. Lantzy S, Anderson D. Can consumers use online reviews to avoid unsuitable doctors? Evidence from ratemds.Com and the federation of state medical boards. Decision Sci 2019 Jun 26;51(4):962-984. [doi: 10.1111/deci.12398]
33. Dagger TS, Sweeney JC, Johnson LW. A hierarchical model of health service quality. J Service Res 2016 Jun 29;10(2):123-142. [doi: 10.1177/1094670507309594]
34. Gershoff AD, Broniarczyk SM, West PM. Recommendation or evaluation? Task sensitivity in information source selection. J Consum Res 2001 Dec 01;28(3):418-438. [doi: 10.1086/323730]
35. Ketelaar PE, Willemsen LM, Sleven L, Kerkhof P. The good, the bad, and the expert: how consumer expertise affects review valence effects on purchase intentions in online product reviews. J Comput Mediat Comm 2015 Nov 27;20(6):649-666. [doi: 10.1111/jcc4.12139]
36. Goldstein NJ, Cialdini RB, Griskevicius V. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. J Consum Res 2008 Oct 01;35(3):472-482. [doi: 10.1086/586910]
37. Nikou SA, Economides AA. Mobile-based assessment: investigating the factors that influence behavioral intention to use. Comput Educ 2017 Jun;109:56-73. [doi: 10.1016/j.compedu.2017.02.005]
38. Shen AX, Cheung CM, Lee MK, Chen H. How social influence affects we-intention to use instant messaging: the moderating effect of usage experience. Inf Syst Front 2009 Jun 24;13(2):157-169. [doi: 10.1007/s10796-009-9193-9]

XSL•FO
RenderX

39. Chaouali W, Ben Yahia I, Souiden N. The interplay of counter-conformity motivation, social influence, and trust in customers' intention to adopt internet banking services: the case of an emerging country. J Retailing Consum Serv 2016 Jan;28:209-218. [doi: 10.1016/j.jretconser.2015.10.007]
40. Boontarig W, Chutimaskul W, Chongsuphajaisiddhi V, Papasratorn B. Factors influencing the Thai elderly intention to use smartphone for e-Health services. In: Proceedings of the 2012 IEEE Symposium on Humanities, Science and Engineering Research. 2012 Presented at: 2012 IEEE Symposium on Humanities, Science and Engineering Research; Jun 24-27, 2012; Kuala Lumpur, Malaysia. [doi: 10.1109/shuser.2012.6268881]
41. Wang N, Sun Y. Social influence or personal preference? Examining the determinants of usage intention across social media with different sociability. Inform Develop 2016 Jul 09;32(5):1442-1456. [doi: 10.1177/0266666915603224]
42. Abdul Razak FZ, Abu Bakar A, Abdullah W. How perceived effort expectancy and social influence affects continuance of intention to use e-government. A study of a Malaysian government service. Electron Gov 2017;13(1):1. [doi: 10.1504/eg.2017.10004264]
43. Bakar AA, Razak F. The role of facilitating condition and social influence towards continuance intention to use elearning. Intl J Technical Res Appl 2014;2(1):12-14.
44. Cokins G, Oncioiu I, Türke MC, Topor DI, Căpuşneanu S, Pa tiu CA, et al. Intention to use accounting platforms in Romania: a quantitative study on sustainability and social influence. Sustainability 2020 Jul 30;12(15):6127. [doi: 10.3390/su12156127]
45. Krueger N, Carsrud A. Entrepreneurial intentions: applying the theory of planned behaviour. Entrepreneurship Regional Develop 1993 Jan;5(4):315-330. [doi: 10.1080/08985629300000020]
46. Tonglet M, Phillips P, Read A. Using the Theory of Planned Behaviour to investigate the determinants of recycling behaviour: a case study from Brixworth, UK. Resources Conservation Recycling 2004 Jun;41(3):191-214. [doi: 10.1016/j.resconrec.2003.11.001]
47. Hardeman W, Johnston M, Johnston D, Bonetti D, Wareham N, Kinmonth A. Application of the theory of planned behaviour in behaviour change interventions: a systematic review. Psychol Health 2002 Jan;17(2):123-158. [doi: 10.1080/08870440290013644a]
48. Anderson J, Schwager P. SME adoption of wireless LAN technology: applying the UTAUT model. In: Proceedings of the 7th Annual Conference of the Southern Association for Information Systems (SAIS). 2004 Presented at: 7th Annual Conference of the Southern Association for Information Systems (SAIS); Feb. 27-28, 2004; Savannah GA URL: https://aisel.aisnet.org/sais2004/6
49. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
50. Kim Y, Srivastava J. Impact of social influence in e-commerce decision making. In: Proceedings of the ninth international conference on Electronic commerce. 2007 Presented at: ICEC07: Ninth International Conference on Electronic Commerce; Aug 19 - 22, 2007; Minneapolis MN USA. [doi: 10.1145/1282100.1282157]
51. Wolters R, Wensing M, Van Weel GC, Van Der Wilt RP, Grol RP. Lower urinary tract symptoms: social influence is more important than symptoms in seeking medical care. BJU Int 2002;90(7):655-661. [doi: 10.1046/j.1464-410X.2002.02996.x]
52. McGlone TA, Butler ES, McGlone VL. Factors influencing consumers' selection of a primary care physician. Health Mark Q 2008 Sep 25;19(3):21-37. [doi: 10.1300/j026v19n03_03]
53. Shah J, Dickinson C. Establishing which factors patients value when selecting urology outpatient care. Br J Med Surg Urol 2010 Jan 01;3(1):25-29. [doi: 10.1016/j.bjmsu.2009.10.003]
54. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Personal Soc Psychol 1986;51(6):1173-1182. [doi: 10.1037/0022-3514.51.6.1173]
55. Zhao X, Lynch JG, Chen Q. Reconsidering Baron and Kenny: myths and truths about mediation analysis. J Consum Res 2010 Aug 01;37(2):197-206. [doi: 10.1086/651257]
56. O'keefe D. Persuasion. In: The International Encyclopedia of Communication. Hoboken, New Jersey: Wiley; 2012.
57. Erdem T, Swait J. Brand equity as a signaling phenomenon. J Consum Psychol 1998 Jan;7(2):131-157. [doi: 10.1207/s15327663jcp0702_02]
58. Metzger M, Flanagin A, Medders R. Social and heuristic approaches to credibility evaluation online. J Commun 2010;60(3):413-439.
59. Gass R, Seiter J. Persuasion, Social Influence, and Compliance Gaining. Boston, Massachusetts, United States: Allyn & Bacon; 2011.
60. Wang Y, Wang Y, Lin H, Tang T. Determinants of user acceptance of internet banking: an empirical study. Int J Serv Ind Manag 2003;14(5):501-519.
61. Luarn P, Lin H. Toward an understanding of the behavioral intention to use mobile banking. Comput Human Behav 2005 Nov;21(6):873-891. [doi: 10.1016/j.chb.2004.03.003]
62. Broman Toft M, Schuitema G, Thøgersen J. Responsible technology acceptance: model development and application to consumer acceptance of Smart Grid technology. Applied Energy 2014 Dec;134:392-400. [doi: 10.1016/j.apenergy.2014.08.048]

XSL•FO
RenderX

63. Haas C, Wearden S. E-credibility: building common ground in web environments. L1 Educ Stud Language Literature 2003:169-184.
64. Jensen ML, Averbeck JM, Zhang Z, Wright KB. Credibility of anonymous online product reviews: a language expectancy perspective. J Manag Inform Syst 2014 Dec 08;30(1):293-324. [doi: 10.2753/mis0742-1222300109]
65. Chen Y, Shang R, Li M. The effects of perceived relevance of travel blogs' content on the behavioral intention to visit a tourist destination. Comput Human Behav 2014 Jan;30:787-799. [doi: 10.1016/j.chb.2013.05.019]
66. Fedorko I, Bačik R, Gavurova B. Effort expectancy and social influence factors as main determinants of performance expectancy using electronic banking. Banks Bank Syst 2021 Apr 28;16(2):27-37. [doi: 10.21511/bbs.16(2).2021.03]
67. Im I, Hong S, Kang MS. An international comparison of technology adoption. Inform Manag 2011 Jan;48(1):1-8. [doi: 10.1016/j.im.2010.09.001]
68. Carlsson C, Carlsson J, Hyvonen K, Puhakainen J, Walden P. Adoption of mobile devices/services-searching for answers with the UTAUT. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06). 2006 Presented at: 39th Annual Hawaii International Conference on System Sciences (HICSS'06); Jan 04-07, 2006; Kauai, HI, USA. [doi: 10.1109/hicss.2006.38]
69. Marchewka J, Kostiwa K. An application of the UTAUT model for understanding student perceptions using course management software. Commun IIMA 2007;7:10 [FREE Full text]
70. Duarte P, Pinho JC. A mixed methods UTAUT2-based approach to assess mobile health adoption. J Bus Res 2019 Sep;102:140-150. [doi: 10.1016/j.jbusres.2019.05.022]
71. Macedo IM. Predicting the acceptance and use of information and communication technology by older adults: an empirical examination of the revised UTAUT2. Comput Human Behav 2017 Oct;75:935-948. [doi: 10.1016/j.chb.2017.06.013]
72. Owusu Kwateng K, Osei Atiemo KA, Appiah C. Acceptance and use of mobile banking: an application of UTAUT2. J Enterp Inform Manag 2019 Feb 11;32(1):118-151. [doi: 10.1108/jeim-03-2018-0055]
73. Yu C. Factors affecting individuals to adopt mobile bankingmpirical evidence from the UTAUT model. J Electron Commer Res J 2012;13(2):104 [FREE Full text]
74. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
75. Noblin AM, Wan TT, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13 [FREE Full text] [Medline: 23209454]
76. 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]
77. Cho YI, Lee SD, Arozullah AM, Crittenden KS. Effects of health literacy on health status and health service utilization amongst the elderly. Soc Sci Med 2008 Apr;66(8):1809-1816. [doi: 10.1016/j.socscimed.2008.01.003] [Medline: 18295949]
78. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar 31;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
79. Schulz PJ, Rothenfluh F. Influence of health literacy on effects of patient rating websites: survey study using a hypothetical situation and fictitious doctors. J Med Internet Res 2020 Apr 06;22(4):e14134 [FREE Full text] [doi: 10.2196/14134] [Medline: 32250275]
80. Chaudhary V, Sharma D, Kalro A. Consumer scepticism: a systematic review of literature. Mark Rev 2019 Dec 31;19(3):189-212. [doi: 10.1362/146934719X15774562877683]
81. Ahmad F, Guzmán F. Consumer skepticism about online reviews and their decision-making process: the role of review self-efficacy and regulatory focus. J Consum Mark 2021 Aug 11;38(5):587-600. [doi: 10.1108/jcm-09-2020-4119]
82. Sher PJ, Lee S. Consumer skepticism and online reviews: an elaboration likelihood model perspective. Soc Behav Pers 2009 Feb 01;37(1):137-143. [doi: 10.2224/sbp.2009.37.1.137]
83. Reimer T, Benkenstein M. When good WOM hurts and bad WOM gains: the effect of untrustworthy online reviews. J Bus Res 2016 Dec;69(12):5993-6001. [doi: 10.1016/j.jbusres.2016.05.014]
84. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
85. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
86. Liff JM, Chow W, Greenberg RS. Rural–urban differences in stage at diagnosis. Possible relationship to cancer screening. Cancer 1991 Mar 01;67(5):1454-1459. [doi: 10.1002/1097-0142(19910301)67:5<1454::aid-cncr2820670533>3.0.co;2-k]
87. Ohanian R. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. J Advert 1990 Oct;19(3):39-52. [doi: 10.1080/00913367.1990.10673191]
88. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]

XSL•FO
RenderX

89. Hulland J, Baumgartner H, Smith KM. Marketing survey research best practices: evidence and recommendations from a review of JAMS articles. J Acad Mark Sci 2017 Apr 10;46(1):92-108. [doi: 10.1007/s11747-017-0532-y]
90. MacKenzie S, Podsakoff P. Common method bias in marketing: causes, mechanisms, and procedural remedies. J Retail 2012 Dec;88(4):542-555. [doi: 10.1016/j.jretai.2012.08.001]
91. Seber G, Lee A. Linear Regression Analysis, Second Edition. Hoboken, New Jersey, United States: Wiley; 2003.
92. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis A Regression-Based Approach. New York, United States: Guilford Publications; 2022.
93. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach (3rd Edition). New York: The Guilford Press; 2022.
94. Hayes AF. An index and test of linear moderated mediation. Multivariate Behav Res 2015;50(1):1-22. [doi: 10.1080/00273171.2014.962683] [Medline: 26609740]
95. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Taylor & Francis; 2013.
96. Soellner R, Huber S, Reder M. The concept of eHealth literacy and its measurement. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
97. 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]
98. Skarmeas D, Leonidou CN. When consumers doubt, watch out! the role of CSR skepticism. J Bus Res 2013 Oct;66(10):1831-1838. [doi: 10.1016/j.jbusres.2013.02.004]
99. Storm V, Reinwand DA, Wienert J, Tan S, Lippke S. The mediating role of perceived social support between physical activity habit strength and depressive symptoms in people seeking to decrease their cardiovascular risk: cross-sectional study. JMIR Ment Health 2018 Nov 14;5(4):e11124 [FREE Full text] [doi: 10.2196/11124] [Medline: 30429112]
100. Roberts JA, David ME. Improving predictions of COVID-19 preventive behavior: development of a sequential mediation model. J Med Internet Res 2021 Mar 17;23(3):e23218 [FREE Full text] [doi: 10.2196/23218] [Medline: 33651707]
101. Kerr B, D'Angelo JD, Diaz-Caballero A, Moreno MA. College student problematic internet use and digital communication medium used with parents: cross-sectional study. JMIR Pediatr Parent 2020 Apr 23;3(1):e17165 [FREE Full text] [doi: 10.2196/17165] [Medline: 32324140]
102. Guan Y, Duan W. The mediating role of visual stimuli from media use at bedtime on psychological distress and fatigue in college students: cross-sectional study. JMIR Ment Health 2020 Mar 16;7(3):e11609 [FREE Full text] [doi: 10.2196/11609] [Medline: 32175912]
103. Audrain-Pontevia A, Menvielle L, Ertz M. Effects of three antecedents of patient compliance for users of peer-to-peer online health communities: cross-sectional study. J Med Internet Res 2019 Nov 11;21(11):e14006 [FREE Full text] [doi: 10.2196/14006] [Medline: 31710295]
104. Adebesin F, Mwalugha R. The mediating role of organizational reputation and trust in the intention to use wearable health devices: cross-country study. JMIR Mhealth Uhealth 2020 Jun 09;8(6):e16721 [FREE Full text] [doi: 10.2196/16721] [Medline: 32348260]
105. Liu S, Perdew M, Lithopoulos A, Rhodes RE. The feasibility of using Instagram data to predict exercise identity and physical activity levels: cross-sectional observational study. J Med Internet Res 2021 Apr 19;23(4):e20954 [FREE Full text] [doi: 10.2196/20954] [Medline: 33871380]
106. Duarte J, Pinto-Gouveia J, Cruz B. Relationships between nurses' empathy, self-compassion and dimensions of professional quality of life: a cross-sectional study. Int J Nurs Stud 2016 Aug;60:1-11. [doi: 10.1016/j.ijnurstu.2016.02.015] [Medline: 27297364]
107. Kristofferzon M, Engström M, Nilsson A. Coping mediates the relationship between sense of coherence and mental quality of life in patients with chronic illness: a cross-sectional study. Qual Life Res 2018 Jul;27(7):1855-1863 [FREE Full text] [doi: 10.1007/s11136-018-1845-0] [Medline: 29623597]
108. Dahlkvist E, Hartig T, Nilsson A, Högberg H, Skovdahl K, Engström M. Garden greenery and the health of older people in residential care facilities: a multi-level cross-sectional study. J Adv Nurs 2016 Sep;72(9):2065-2076. [doi: 10.1111/jan.12968] [Medline: 27028976]
109. Khan MN, Mahmood A. A distinctive approach to obtain higher page rank through search engine optimization. Sādhanā 2018 Mar 22;43(3). [doi: 10.1007/s12046-018-0812-3]
110. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
111. Chang S, van Witteloostuijn A, Eden L. From the editors: common method variance in international business research. J Int Bus Stud 2010 Feb 8;41(2):178-184. [doi: 10.1057/jibs.2009.88]
112. Aguinis H, Villamor I, Ramani R. MTurk research: review and recommendations. J Manag 2020 Nov 19;47(4):823-837. [doi: 10.1177/0149206320969787]
113. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: detecting satisficing to increase statistical power. J Exp Soc Psychol 2009 Jul;45(4):867-872. [doi: 10.1016/j.jesp.2009.03.009]

XSL•FO
RenderX

114. Kung FY, Kwok N, Brown DJ. Are attention check questions a threat to scale validity? Applied Psychol 2017 Aug 24;67(2):264-283. [doi: 10.1111/apps.12108]
115. VanderWeele TJ, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiol Methods 2014 Jan;2(1):95-115 [FREE Full text] [doi: 10.1515/em-2012-0010] [Medline: 25580377]
116. Steen J, Loeys T, Moerkerke B, Vansteelandt S. Flexible mediation analysis with multiple mediators. Am J Epidemiol 2017 Jul 15;186(2):184-193. [doi: 10.1093/aje/kwx051] [Medline: 28472328]
117. Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics 2015 Mar 28;71(1):1-14 [FREE Full text] [doi: 10.1111/biom.12248] [Medline: 25351114]
118. Gotch C, Hall * T. Understanding nature‐related behaviors among children through a theory of reasoned action approach. Environ Educ Res 2004 May;10(2):157-177. [doi: 10.1080/13504620242000198159]
119. Fishbein M. A reasoned action approach to health promotion. Med Decis Making 2008;28(6):834-844 [FREE Full text] [doi: 10.1177/0272989X08326092] [Medline: 19015289]
120. Ajzen I, Albarracin D, Hornik R. Prediction and Change of Health Behavior Applying the Reasoned Action Approach. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Taylor & Francis; 2007.
121. Lada S, Tanakinjal G, Amin H. Predicting intention to choose halal products using theory of reasoned action. Int J Islam Middle East Financ Manag 2009:66-76.
122. Buttle F, Bok B. Hotel marketing strategy and the theory of reasoned action. Int J Contemp Hosp Manag 1996:5-10.
123. van den Bos K, Lind EA, Bommelé J, VandeVondele SD. Reminders of behavioral disinhibition increase public conformity in the Asch paradigm and behavioral affiliation with ingroup members. Front Psychol 2015;6:837 [FREE Full text] [doi: 10.3389/fpsyg.2015.00837] [Medline: 26175695]
124. Dunning D, Van Boven L, Loewenstein G. Egocentric empathy gaps in social interaction and exchange. In: Advances in Group Processes. Bingley, United Kingdom: Emerald Group Publishing Limited; 2001.
125. Hwang K, Zhang Q. Influence of parasocial relationship between digital celebrities and their followers on followers' purchase and electronic word-of-mouth intentions, and persuasion knowledge. Comput Human Behav 2018 Oct;87:155-173. [doi: 10.1016/j.chb.2018.05.029]
126. Yuan S, Lou C. How social media influencers foster relationships with followers: the roles of source credibility and fairness in parasocial relationship and product interest. J Interact Advert 2020 Jul 13;20(2):133-147. [doi: 10.1080/15252019.2020.1769514]
127. R. Taylor C. Advertising and COVID-19. Int J Advert 2020 Jun 09;39(5):587-589. [doi: 10.1080/02650487.2020.1774131]
128. Lookadoo K, Hubbard C, Nisbett G, Wong N. We're all in this together: celebrity influencer disclosures about COVID-19. Atl J Commun 2021 Jul 30;30(4):397-418. [doi: 10.1080/15456870.2021.1936526]
129. Janssen L, Schouten A, Croes E. Influencer advertising on Instagram: product-influencer fit and number of followers affect advertising outcomes and influencer evaluations via credibility and identification. Intl J Advert 2021 Oct 28;41(1):101-127. [doi: 10.1080/02650487.2021.1994205]
130. Schouten A, Janssen L, Verspaget M. Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit. Int J Advert 2019 Jul 02;39(2):258-281. [doi: 10.1080/02650487.2019.1634898]
131. Shan Y, Chen K, Lin J. When social media influencers endorse brands: the effects of self-influencer congruence, parasocial identification, and perceived endorser motive. Intl J Advert 2019 Oct 18;39(5):590-610. [doi: 10.1080/02650487.2019.1678322]
132. Bidmon S, Terlutter R, Röttl J. What explains usage of mobile physician-rating apps? Results from a web-based questionnaire. J Med Internet Res 2014 Jun 11;16(6):e148 [FREE Full text] [doi: 10.2196/jmir.3122] [Medline: 24918859]
133. Bidmon S, Elshiewy O, Terlutter R, Boztug Y. What patients value in physicians: analyzing drivers of patient satisfaction using physician-rating website data. J Med Internet Res 2020 Feb 03;22(2):e13830 [FREE Full text] [doi: 10.2196/13830] [Medline: 32012063]
134. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 02;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]

Abbreviations

PRW: physician rating website
SMI: social media influencer
TAM: technology acceptance model
TIE: theory of information economics
TPB: theory of planned behavior
TRA: theory of reasoned action
UTAUT: Unified Theory of Acceptance and Use of Technology

XSL•FO
RenderX

Edited by A Mavragani, G Eysenbach; submitted 24.02.22; peer-reviewed by S McLennan, A Reviewer; comments to author 20.04.22; revised version received 23.05.22; accepted 11.10.22; published 14.11.22.

Please cite as:
Guetz B, Bidmon S
The Impact of Social Influence on the Intention to Use Physician Rating Websites: Moderated Mediation Analysis Using a Mixed Methods Approach
J Med Internet Res 2022;24(11):e37505
URL: https://www.jmir.org/2022/11/e37505
doi:10.2196/37505
PMID:36374547

©Bernhard Guetz, Sonja Bidmon. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Investigating Patients' Continuance Intention Toward Conversational Agents in Outpatient Departments: Cross-sectional Field Survey

Xingyi Li[1], MPH; Shirong Xie[2], BS; Zhengqiang Ye[2], MEng; Shishi Ma[3], MPH; Guangjun Yu[4], PhD

[1]School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Eye and Ear, Nose, and Throat Hospital, Fudan University, Shanghai, China

[3]Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[4]Shanghai Children's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Corresponding Author:
Guangjun Yu, PhD
Shanghai Children's Hospital
Shanghai Jiao Tong University School of Medicine
No 355 Luding Road
Shanghai, 200062
China
Phone: 86 18917762998
Email: gjyu@shchildren.com.cn

Abstract

Background: Conversational agents (CAs) have been developed in outpatient departments to improve physician-patient communication efficiency. As end users, patients' continuance intention is essential for the sustainable development of CAs.

Objective: The aim of this study was to facilitate the successful usage of CAs by identifying key factors influencing patients' continuance intention and proposing corresponding managerial implications.

Methods: This study proposed an extended expectation-confirmation model and empirically tested the model via a cross-sectional field survey. The questionnaire included demographic characteristics, multiple-item scales, and an optional open-ended question on patients' specific expectations for CAs. Partial least squares structural equation modeling was applied to assess the model and hypotheses. The qualitative data were analyzed via thematic analysis.

Results: A total of 172 completed questionaries were received, with a 100% (172/172) response rate. The proposed model explained 75.5% of the variance in continuance intention. Both satisfaction (β=.68; *P*<.001) and perceived usefulness (β=.221; *P*=.004) were significant predictors of continuance intention. Patients' extent of confirmation significantly and positively affected both perceived usefulness (β=.817; *P*<.001) and satisfaction (β=.61; *P*<.001). Contrary to expectations, perceived ease of use had no significant impact on perceived usefulness (β=.048; *P*=.37), satisfaction (β=−.004; *P*=.63), and continuance intention (β=.026; *P*=.91). The following three themes were extracted from the 74 answers to the open-ended question: personalized interaction, effective utilization, and clear illustrations.

Conclusions: This study identified key factors influencing patients' continuance intention toward CAs. Satisfaction and perceived usefulness were significant predictors of continuance intention (*P*<.001 and *P*<.004, respectively) and were significantly affected by patients' extent of confirmation (*P*<.001 and *P*<.001, respectively). Developing a better understanding of patients' continuance intention can help administrators figure out how to facilitate the effective implementation of CAs. Efforts should be made toward improving the aspects that patients reasonably expect CAs to have, which include personalized interactions, effective utilization, and clear illustrations.

(J Med Internet Res 2022;24(11):e40681) doi:10.2196/40681

KEYWORDS

conversational agent; continuance intention; expectation-confirmation model; partial least squares; structural equation modeling; chatbot; virtual assistant; cross-sectional; field study; optimization; outpatient; interview; qualitative; questionnaire; satisfaction; perceived usefulness; intention; adoption; attitude; perception

XSL•FO
RenderX

Introduction

Background

Tertiary hospitals in China are occupied with many outpatients every day, which results in long waiting times and limited physician-patient communication during consultations. This phenomenon is caused mainly by two aspects. First, the large population base has resulted in a growing demand for medical services. Second, physicians have to finish both consulting patients and filling out medical histories during a limited amount of time. A study found that in the consultation room, 66.5% of physicians' time was spent on communication and the examination of patients, and 20.7% of their time was spent on writing medical records [1]. Long waiting times, together with limited consultation times, further result in insufficient physician-patient communication and an incomplete understanding of conditions and diagnoses (ie, knowing all of the facts) [2,3]. Besides, during the ongoing COVID-19 pandemic, long waiting times have also put patients at risk for cross-infection [4].

Under national policies on the digital transformation of the health care industry, Shanghai, as a leading digital city, developed conversational agents (CAs) in outpatient departments, hoping to alleviate patient overload and improve communication efficiency. CAs are artificial intelligence programs that engage in dialogues with patients on mobile devices [5]. With these contextual question–answering agents, data on patients' symptoms and medical histories can be captured and delivered to physicians' workstations in structured forms before a consultation. During face-to-face consultations, physicians can rapidly gain an understanding of patients' general conditions and focus on other responsibilities [6], which has resulted in a man-machine integrated consultation model.

Prior studies have indicated that CAs can save time by reducing the time required for history taking, improve consultation efficiency, and enhance the completeness and accuracy of medical histories [7-10]. However, their potential has not been fully exploited, as the usage of CAs is often limited; 6 months after tertiary hospitals in Shanghai established CAs, the usage rates fell short of expectations (26% and 20%, respectively, for the second- and fourth-ranked hospitals). As end users, patients' continuance intention is essential for the sustainable development of CAs [11], yet limited studies are available.

Based on the abovementioned research background and motivations, this study has 3 aims. First, it attempts to identify key factors influencing users' continuance usage intention via a theoretical model. Second, it empirically examines the applicability of the model in the context of implementing CAs in outpatient departments. Third, it proposes corresponding managerial implications based on the results.

Theoretical Background and Model

The usage of information systems includes the following two stages: preacceptance (acceptance before a system's initial use) and postacceptance (acceptance after a system's initial use; ie, continuance).

Even though initial use is an important first step toward realizing information systems' success, it is mostly influenced by secondhand information from referent others or popular media rather than users' actual interactions with the information system. In contrast, continuance after a system's initial use is more realistic and unbiased, since it is grounded in users' firsthand experiences [12]. Therefore, the long-term viability and eventual success of information systems depend on users' continued use.

There has been a considerable body of theory-based research on information system use in recent years. Among the theoretical models, the Technology Acceptance Model (TAM) is commonly applied to understand the initial acceptance of information systems, including intelligent health service systems (eg, registration systems and patient portals) [13-16]. The TAM predicts users' initial use of information systems based on the following two constructs: perceived usefulness and perceived ease of use [17].

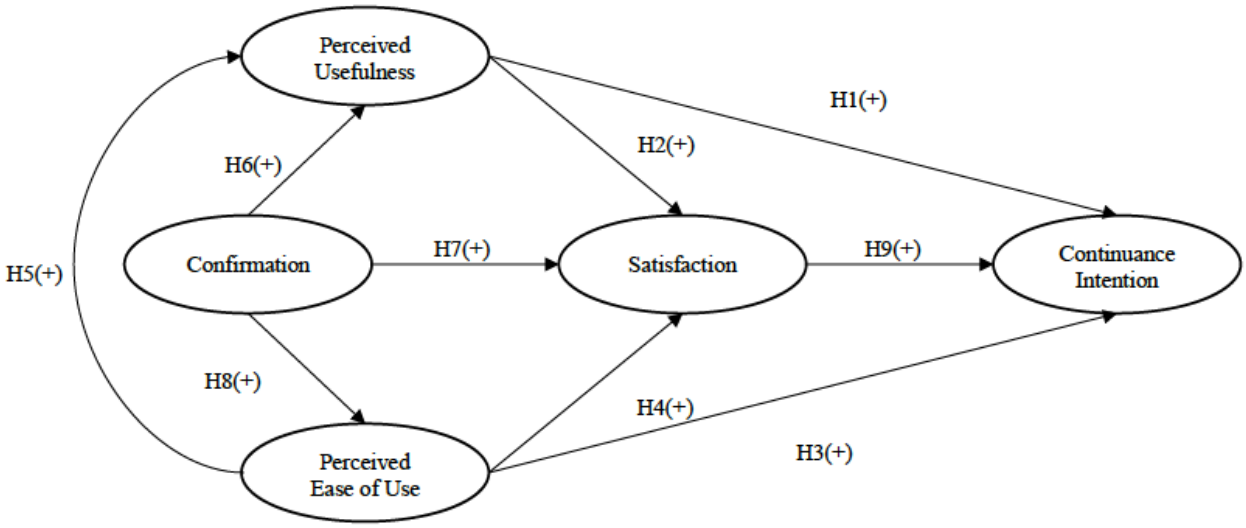
To understand users' continuance behavior in an information system context, an expectation-confirmation model (ECM) of information system continuance was proposed by Bhattacherjee [11]. This ECM has been empirically tested in a variety of contexts, including health services such as e-appointment systems and teleconsultations. The model predicts users' continuance intention via the following three antecedent constructs: perceived usefulness, confirmation, and satisfaction.

The ECM only incorporated perceived usefulness from the TAM, as Bhattacherjee [11] considered it to be the more salient and consistent predictor of information system use intention. However, both perceived usefulness and perceived ease of use are the primary motivators of information system acceptance in the TAM [18,19]. The significant impact of perceived ease of use on both perceived usefulness and usage intention has been verified in previous research (eg, studies on electronic health record acceptance by physicians and self-management technology acceptance by patients [20,21]). Given the particularities of patients and medical professionalism, perceived ease of use might also have the potential to influence patients' continuance intention. Therefore, this study extended existing ECM constructs by integrating perceived ease of use, hoping to provide a better understanding of patients' continuance intention in the context of CAs.

Based on the abovementioned theoretical reasoning and the results of previous research, we propose the following theoretical model (Figure 1) and hypotheses: (1) perceived usefulness positively affects continuance intention (hypothesis 1), (2) perceived usefulness positively affects satisfaction (hypothesis 2), (3) perceived ease of use positively affects continuance intention (hypothesis 3), (4) perceived ease of use positively affects satisfaction (hypothesis 4), (5) perceived ease of use positively affects perceived usefulness (hypothesis 5), (6) the confirmation of initial expectations positively affects perceived usefulness (hypothesis 6), (7) the confirmation of initial expectations positively affects satisfaction (hypothesis 7), (8) the confirmation of initial expectations positively affects perceived ease of use (hypothesis 8), and (9) satisfaction positively affects continuance intention (hypothesis 9).

XSL•FO
RenderX

Figure 1. Research model. H: hypothesis; +: positive effect.



Methods

Study Design and Setting

Tertiary hospitals in Shanghai established CAs in early 2021 to improve medical service efficiency and alleviate the overload in outpatient departments. Before face-to-face consultations, patients can provide their symptoms and medical histories on mobile devices via a contextual question–answering agent. Afterward, physicians can rapidly gain an understanding of patients' general conditions with the structured information delivered by CAs.

Shanghai Eye and ENT (ear, nose, and throat) Hospital has been a pioneer during the CA implementation process. Empirical data for this research were collected via a cross-sectional field survey that was conducted in the outpatient department of Shanghai Eye and ENT Hospital. The duration of this study was 3 months (November 2021 to January 2022). We invited patients and their companions who had used CAs. The survey was conducted near the pharmacy to make sure that patients finished their face-to-face consultations and minimize possible inconveniences.

Sample Size and Sampling

The minimum sample size for this research was 124, and according to Marcoulides and Saunders [22], the minimum sample size depends on the maximum number of arrows pointing at a latent variable. Hoyle [23] recommended a sample size of 100 to 200 when performing path modeling. The convenience sampling technique was used, and a total of 172 questionnaires were completed. This sample size met the requirements for obtaining sufficient statistical power.

Measurement Tools

The questionnaire included 3 parts—demographic characteristics, multiple-item scales, and the following optional open-ended question: “What are your expectations that CAs failed to meet?” The five constructs in the proposed model were measured by using multiple-item scales that were adapted from Davis [17] and Bhattacherjee [11], and the items were reworded to accommodate the context of CA use. Satisfaction items were scored on 5-point semantic differential scales. The remaining items were scored on 5-point Likert scales that ranged from 1 (strongly disagree) to 5 (strongly agree). Table 1 provides definitions and sources for the five constructs. The scale items were translated from English to Chinese because the survey was conducted in China. To avoid wording-related misapprehension, we used a back-translation process [24]. Multimedia Appendix 1 presents the items for each construct and their sources. A pretest was conducted among 25 patients to ensure the reliability and validity of the questionnaire.

Table 1. Definitions of constructs.

Construct	Operational definition	Reference
Continuance intention	Patients’ intention to continue using conversational agents	Bhattacherjee [11]
Satisfaction	Patients’ affects (feelings) prior to using conversational agents	Bhattacherjee [11]
Perceived usefulness	Patients’ perceptions of the expected benefits of conversational agents	Bhattacherjee [11]
Perceived ease of use	The degree to which patients believe that using conversational agents would be free from effort	Davis [17]
Confirmation	Patients’ perceptions of the congruence between expectations for conversational agents and their actual performance	Bhattacherjee [11]

XSL•FO
RenderX

Data Collection

To help patients better understand their choices and remain focused, two postgraduates from Shanghai Jiao Tong University School of Medicine conducted the in-person survey, using paper questionnaires. During this process, the investigators read all of the questions aloud to the patients and filled in the questionnaire with their answers, which saved patients the trouble of reading the items themselves. The questionnaires ended with an optional open-ended question ("What are your expectations that CAs failed to meet?"). The answers were collected through brief interviews and written down in the form of detailed summaries by the investigators. A total of 172 valid questionnaires were collected, with a 100% (172/172) response rate, and 74 participants answered the optional open-ended question.

Data Analysis

Descriptive statistics were performed by using SPSS 25.0 (IBM Corporation). A partial least squares structural equation model (PLS-SEM) analysis was performed in SmartPLS 3.3.3 (SmartPLS GmbH) to validate the research model and test the research hypotheses. A PLS-SEM was chosen because it is capable of producing robust results with restricted sample sizes and data lacking normality [25].

The implementation of this method was performed in 2 steps [26]. The first step consisted of assessing the reliability and validity of the measurement model using the partial least squares algorithm, while the second step focused on assessing the fit of the structural model and the significance of the hypotheses by using bootstrapping (5000 bootstrap samples) [27].

The qualitative data were analyzed via thematic analysis. Thematic analysis is a method for analyzing qualitative data that entails searching across a data set to identify, analyze, and report repeated patterns [28]. The initial codes were generated by deductively reading the manuscripts. This was done by a single coder, and the codes were reviewed by a second analyst [29]. After they reached consensus on the initial codes, the themes were extracted from and defined on the basis of the codes through group discussions.

Ethics Approval

This study was approved by Shanghai Children's Hospital (approval number: 2022R092-E01). All respondents participated in this study voluntarily and anonymously on the basis of informed consent.

Results

Demographic Information

A total of 172 questionnaires were complete and valid, with a 100% (172/172) response rate. The demographic information of CA users is listed in Table 2. Notably, 54.1% of the respondents were in the 25 to 35 years age group, and 63.4% (109/172) were women.

XSL•FO
RenderX

Table 2. Demographic information.

Participant characteristics	Participants (N=172), n (%)
Gender	
Men	63 (36.6)
Women	109 (63.4)
Age (years)	
<25	12 (7)
25-35	93 (54.1)
36-45	46 (26.7)
>45	21 (12.2)
Relationship with the patient	
Patients themselves	106 (61.6)
Patients' children	22 (12.8)
Patients' parents	44 (25.6)
Visit type	
First visit	112 (65.1)
Return visit	60 (34.9)
Number of visits over the past half year	
1	101 (58.7)
2-3	47 (27.3)
>3	24 (14)
Number of times that a participant used a conversational agent	
1	136 (79.1)
2-3	31 (18)
>3	5 (2.9)

Measurement Model Assessment

The measurement model was assessed in terms of construct reliability, convergent validity, and discriminative validity by performing a confirmatory composite analysis. The results are displayed in Table 3 and Table 4.

Reliability can be evaluated with Cronbach α and composite reliability values [30]. Convergent validity can be accessed with factor loading and average variance extracted (AVE) values [27,31]. As shown in Table 2, all of the Cronbach α and composite reliability values were above 0.7, the AVE for each construct was above 0.5, and the factor loadings for each item were above 0.7, indicating good reliability and convergent validity [27].

Discriminant validity reflects the extent to which constructs are significantly different from each other. To achieve discriminant validity, the square root of the AVE for a given construct must be higher than that construct's correlation with other constructs, and this must hold true for all constructs [31]. As shown in Table 4, the results indicated that discriminate validity was achieved. Therefore, we concluded that the quality of the measurement model was sufficient for testing the hypotheses in the model.

XSL•FO
RenderX

Table 3. Construct reliability and convergent validity.

Constructs and items	Factor loadings	Cronbach α	CR[a]	AVE[b]
CONF[c]		.938	0.960	0.890
CONF1	0.947			
CONF2	0.940			
CONF3	0.943			
CI[d]		.993	0.996	0.993
CI1	0.996			
CI2	0.996			
PEOU[e]		.792	0.871	0.694
PEOU1	0.866			
PEOU2	0.763			
PEOU3	0.865			
PU[f]		.793	0.879	0.710
PU1	0.825			
PU2	0.761			
PU3	0.933			
SAT[g]		.959	0.980	0.960
SAT1	0.980			
SAT2	0.981			

[a]CR: composite reliability.
[b]AVE: average variance extracted.
[c]CONF: confirmation.
[d]CI: continuance intention.
[e]PEOU: perceived ease of use.
[f]PU: perceived usefulness.
[g]SAT: satisfaction.

Table 4. Discriminant validity.

	Confirmation	Continuance intention	Perceived ease of use	Perceived usefulness	Satisfaction
Confirmation	0.943[a]	—[b]	—	—	—
Continuance intention	0.853	0.996[a]	—	—	—
Perceived ease of use	0.177	0.175	0.833[a]	—	—
Perceived usefulness	0.825	0.759	0.192	0.843[a]	—
Satisfaction	0.841	0.857	0.157	0.783	0.980[a]

[a]The square root of the average variance extracted for each construct.
[b]Not available.

Structure Model Assessment

The inner variance inflation factors were below 5, indicating that we were able to avoid construct collinearity in the model [27]. The research model was assessed by evaluating the path coefficients (β) and the coefficients of determination (R^2). The path coefficients and their significance levels, as well as hypothesis outcomes and R^2 values, are shown in Figure 2 and Table 5.

β represents the direct effects of independent variables on dependent variables. The hypotheses that were based on the original ECM (hypotheses 1, 2, 6, 7, and 9) were all supported, while the hypotheses regarding the newly integrated construct—perceived ease of use (hypotheses 3, 4, and 5)—were

XSL•FO
RenderX

rejected, except for hypothesis 8. R^2 refers to the amount of explained variance for each endogenous latent variable. The entire model explained 75.5% of the variance in continuance intention and 73.2% of the variance in satisfaction, which was considered substantial.

Figure 2. Results of the structure model. *:*P*<.05; **:*P*<.01; *** *P*<.001.

Table 5. Hypothesis test results.

Hypothesis	Path	Coefficient (β)	*P* value	Outcome
Hypothesis 1	Perceived usefulness → continuance intention	.221	.004	Supported
Hypothesis 2	Perceived usefulness → satisfaction	.280	<.001	Supported
Hypothesis 3	Perceived ease of use → continuance intention	−.004	.63	Rejected
Hypothesis 4	Perceived ease of use → satisfaction	.026	.91	Rejected
Hypothesis 5	Perceived ease of use → perceived usefulness	.048	.37	Rejected
Hypothesis 6	Confirmation → perceived usefulness	.817	<.001	Supported
Hypothesis 7	Confirmation → satisfaction	.610	<.001	Supported
Hypothesis 8	Confirmation → perceived ease of use	.177	.007	Supported
Hypothesis 9	Satisfaction → continuance intention	.680	<.001	Supported

Qualitative Data on Patients' Expectations

A total of 3 themes were extracted from the 74 answers regarding patients' specific expectations that CAs failed to meet. The first theme was *personalized interaction* (mentioned 50 times). Instead of the same interaction content and forms of interaction, patients expected to see more personalized conversations that were based on their previous medical histories and visit types. Older participants asked for a voice recognition function and larger font sizes. The second theme was *effective utilization* (mentioned 37 times). Patients expected CAs to have more useful functions, mainly focusing on self-assessments for referrals and self-management for follow-up treatments. The third theme was *clear illustrations on the use and promises of CAs* (mentioned 15 times). In some cases, CAs were easily mistaken as replacements for face-to-face consultations.

Discussion

Summary of Findings

This study identified key factors influencing patients' continuance intention toward CAs through an extended ECM. Satisfaction (β=.68; *P*<.001) and perceived usefulness (β=.221; *P*=.004) were significant predictors of continuance intention, with satisfaction being the stronger predictor. Patients' extent of confirmation significantly affected both perceived usefulness (β=.817; *P*<.001) and satisfaction (β=.61; *P*<.001). These findings are consistent with the original ECM as well as the findings of previous research on information system usage (eg, telemedicine and health data reporting platform usage) among patients [32-35]. The confirmation of patients' expectations has a positive effect on perceived usefulness and satisfaction, and the improvement of perceived usefulness and satisfaction can further enhance patients' enthusiasm for continuing to use a system.

XSL•FO
RenderX

Our qualitative data shed light on patients' specific expectations that CAs failed to meet, including personalized interactions, effective utilization, and clear illustrations. Our findings can help administrators and researchers better understand low CA usage rates. After using CAs, if these expectations have not been positively confirmed, perceived usefulness and satisfaction among patients will drop accordingly and result in their unwillingness to continue using CAs.

Although accumulated evidence has shown the significant impact of perceived ease of use on both perceived usefulness and usage intention [15-17], in this study, perceived ease of use turned out to be trivial in the context of CAs. Not coincidentally, some studies on information system usage in hospitals have also shown the insignificant relationship between perceived ease of use and usage intention [36-38]. This result indicates that once CAs prove to be useful and effective, patients will consider it worth their time and effort to learn how to use CAs. However, if CAs are easy to use but cannot collect useful medical histories from patients, patients' continuance intention will not improve anyway [37].

Managerial and Public Health Implications

Our findings have important managerial implications. The proposed model provides a feedback channel that administrators can use to gain insight into patients' actual experiences and expectations. To maximize patients' satisfaction and continuance intention, efforts should be made toward improving the aspects that patients reasonably expect CAs to have. Offering personalized interactions based on patients' histories and adding more functions can increase perceived usefulness among patients, while providing clear illustrations on the use and promises of CAs can result in patients having appropriate expectations, which allow for positive postuse confirmation.

Contributions of This Study

This study contributes to the body of knowledge about the determinants of CA continuance usage. Almost half of the existing literature on CA acceptance, adoption, and usage evaluates a specific CA artifact, while only 21% of studies put the user in the center of attention when investigating the determinants of their acceptance and usage of CAs [39]. Most of these user-focused empirical studies did not draw on specific concepts from theory for their evaluations [40-43], which makes the results hard to compare. The contribution of this paper is 2-fold. From a theoretical point of view, we identified key factors influencing users' continuance usage intention through a theoretical model. The applicability and validity of the model was empirically tested via a cross-sectional field survey. From a practical point of view, corresponding managerial implications based on the results were proposed to facilitate the successful and continuous development of CAs.

Study Limitations

This study has several limitations that should be addressed. The digital transformation of CA systems started less than 1 year ago, and the progress of this transformation varies dramatically from hospital to hospital. Therefore, this study was conducted at a hospital with a relatively well-designed system and a larger user base. Further research is needed to confirm our findings in the context of different hospitals and different CAs. Additionally, the sample was not normally distributed in terms of gender and age. However, the data analysis was trustworthy, since a PLS-SEM is capable of producing robust results with restricted sample sizes and data lacking normality. Furthermore, a successful digital transformation in health care is a joint effort, and in terms of CAs, this effort depends not only on patients' continuance but also on physicians' utilization and administrators' management of CAs. A multisource model is required to explore the relationships among the constructs.

Conclusions

This study intended to identify key factors influencing patients' continuance intention toward CAs. Satisfaction and perceived usefulness were significant predictors of continuance intention (P<.001 and P<.004, respectively) and were significantly affected by patients' extent of confirmation (P<.001 and P<.001, respectively). Developing a better understanding of patients' continuance intention can help administrators figure out how to facilitate the effective implementation of CAs. Efforts should be made toward improving the aspects that patients reasonably expect CAs to have, which include personalized interactions, effective utilization, and clear illustrations.

Acknowledgments

This study was supported by the General Program of the National Natural Science Foundation (grants 71874110, 72074146) and the Major Research Plan Project of the National Natural Science Foundation (grant 91846302).

Authors' Contributions

We confirm that this paper has been read and approved by all named authors. XL conceived and designed this study. SX and ZY provided technical support. XL and SM acquired and analyzed the data. GY supervised this study. XL drafted this paper. All authors critically revised this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Operationalization of the research variables.

XSL•FO
RenderX

[DOCX File , 16 KB - jmir_v24i11e40681_app1.docx]

References

1. Toscano F, O'Donnell E, Broderick JE, May M, Tucker P, Unruh MA, et al. How physicians spend their work time: an ecological momentary assessment. J Gen Intern Med 2020 Nov;35(11):3166-3172 [FREE Full text] [doi: 10.1007/s11606-020-06087-4] [Medline: 32808212]
2. Jabour AM. The impact of longer consultation time: A simulation-based approach. Appl Clin Inform 2020 Oct;11(5):857-864 [FREE Full text] [doi: 10.1055/s-0040-1721320] [Medline: 33368059]
3. Barber CEH, Barnabe C, Benseler S, Chin R, Johnson N, Luca N, et al. Patient factors associated with waiting time to pediatric rheumatologist consultation for patients with juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2020 Mar 06;18(1):22 [FREE Full text] [doi: 10.1186/s12969-020-0413-7] [Medline: 32143720]
4. Stadler PC, Senner S, Frey S, Clanner-Engelshofen BM, Frommherz LH, French LE, et al. Teledermatology in times of COVID-19. J Dermatol 2021 May;48(5):620-624 [FREE Full text] [doi: 10.1111/1346-8138.15812] [Medline: 33715181]
5. McGreevey JD3, Hanson CW3, Koppel R. Clinical, legal, and ethical aspects of artificial intelligence-assisted conversational agents in health care. JAMA 2020 Aug 11;324(6):552-553. [doi: 10.1001/jama.2020.2724] [Medline: 32706386]
6. Qian H, Dong B, Yuan JJ, Yin F, Wang Z, Wang HN, et al. Pre-consultation system based on the artificial intelligence has a better diagnostic performance than the physicians in the outpatient department of pediatrics. Front Med (Lausanne) 2021 Nov 08;8:695185 [FREE Full text] [doi: 10.3389/fmed.2021.695185] [Medline: 34820391]
7. de Cock C, Milne-Ives M, van Velthoven MH, Alturkistani A, Lam C, Meinert E. Effectiveness of conversational agents (virtual assistants) in health care: Protocol for a systematic review. JMIR Res Protoc 2020 Mar 09;9(3):e16934 [FREE Full text] [doi: 10.2196/16934] [Medline: 32149717]
8. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: Systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
9. Zakim D, Brandberg H, El Amrani S, Hultgren A, Stathakarou N, Nifakos S, et al. Computerized history-taking improves data quality for clinical decision-making-Comparison of EHR and computer-acquired history data in patients with chest pain. PLoS One 2021 Sep 27;16(9):e0257677 [FREE Full text] [doi: 10.1371/journal.pone.0257677] [Medline: 34570811]
10. Balaji D, He L, Giani S, Bosse T, Wiers R, de Bruijn GJ. Effectiveness and acceptability of conversational agents for sexual health promotion: a systematic review and meta-analysis. Sex Health 2022 Oct;19(5):391-405 [FREE Full text] [doi: 10.1071/SH22016] [Medline: 35863761]
11. Bhattacherjee A. Understanding information systems continuance: An expectation-confirmation model. MIS Q 2001 Sep;25(3):351-370. [doi: 10.2307/3250921]
12. Fazio RH, Zanna MP. Direct experience and attitude-behavior consistency. Adv Exp Soc Psychol 1981;14:161-202. [doi: 10.1016/s0065-2601(08)60372-x]
13. Mao CM, Hovick SR. Adding affordances and communication efficacy to the Technology Acceptance Model to study the messaging features of online patient portals among young adults. Health Commun 2022 Mar;37(3):307-315. [doi: 10.1080/10410236.2020.1838106] [Medline: 33243017]
14. McAlearney AS, Gaughan A, MacEwan SR, Fareed N, Huerta TR. Improving acceptance of inpatient portals: Patients' and care team members' perspectives. Telemed J E Health 2020 Mar;26(3):310-326 [FREE Full text] [doi: 10.1089/tmj.2019.0026] [Medline: 31081723]
15. Tavares J, Oliveira T. New integrated model approach to understand the factors that drive electronic health record portal adoption: Cross-sectional national survey. J Med Internet Res 2018 Nov 19;20(11):e11032 [FREE Full text] [doi: 10.2196/11032] [Medline: 30455169]
16. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the Technology Acceptance Model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: Descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
17. 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]
18. Liu K, Or CK, So M, Cheung B, Chan B, Tiwari A, et al. A longitudinal examination of tablet self-management technology acceptance by patients with chronic diseases: Integrating perceived hand function, perceived visual function, and perceived home space adequacy with the TAM and TPB. Appl Ergon 2022 Apr;100:103667. [doi: 10.1016/j.apergo.2021.103667] [Medline: 34920356]
19. Knox L, Gemine R, Rees S, Bowen S, Groom P, Taylor D, et al. Using the Technology Acceptance Model to conceptualise experiences of the usability and acceptability of a self-management app (COPD.Pal®) for chronic obstructive pulmonary disease. Health Technol (Berl) 2021;11(1):111-117 [FREE Full text] [doi: 10.1007/s12553-020-00494-7] [Medline: 33262925]
20. Beglaryan M, Petrosyan V, Bunker E. Development of a tripolar model of technology acceptance: Hospital-based physicians' perspective on EHR. Int J Med Inform 2017 Jun;102:50-61. [doi: 10.1016/j.ijmedinf.2017.02.013] [Medline: 28495348]

XSL•FO
RenderX

21. Gagnon MP, Ghandour EK, Talla PK, Simonyan D, Godin G, Labrecque M, et al. Electronic health record acceptance by physicians: testing an integrated theoretical model. J Biomed Inform 2014 Apr;48:17-27 [FREE Full text] [doi: 10.1016/j.jbi.2013.10.010] [Medline: 24184678]
22. Marcoulides GA, Saunders C. PLS: A silver bullet? MIS Q 2006 Jun;30(2):iii-iix. [doi: 10.2307/25148727]
23. Hoyle RH. The structural equation modeling approach: Basic concepts and fundamental issues. In: Hoyle RH, editor. Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks, CA: Sage Publications Inc; 1995:1-15.
24. Brislin RW. Back-translation for cross-cultural research. J Cross Cult Psychol 1970;1(3):185-216. [doi: 10.1177/135910457000100301]
25. Chin WW, Newsted PR. Structural equation modeling analysis with small samples using partial least squares. In: Hoyle RH, editor. Statistical Strategies for Small Sample Research. Thousand Oaks, CA: Sage Publications Inc; 1999:307-337.
26. Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. Psychol Bull 1988;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
27. Hair JFJ, Hult GTM, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM), Second Edition. Thousand Oaks, CA: Sage Publications Inc; 2017.
28. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2016 Dec 09;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
29. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. Med Teach 2020 Aug;42(8):846-854. [doi: 10.1080/0142159X.2020.1755030] [Medline: 32356468]
30. Nunnally JC, Bernstein IH. The assessment of reliability. In: Psychometric Theory, Third Edition. New York, NY: McGraw-Hill; 1994:248-292.
31. Chin WW. How to write up and report PLS analyses. In: Wang H, Henseler J, Chin WW, Vinzi VE, editors. Handbook of Partial Least Squares: Concepts, Methods and Applications. Berlin, Germany: Springer; 2010:655-690.
32. Reychav I, Arora A, Sabherwal R, Polyak K, Sun J, Azuri J. Reporting health data in waiting rooms with mobile technology: Patient expectation and confirmation. Int J Med Inform 2021 Apr;148:104376. [doi: 10.1016/j.ijmedinf.2021.104376] [Medline: 33453635]
33. Ouimet AG, Wagner G, Raymond L, Pare G. Investigating patients' intention to continue using teleconsultation to anticipate postcrisis momentum: Survey study. J Med Internet Res 2020 Nov 26;22(11):e22081 [FREE Full text] [doi: 10.2196/22081] [Medline: 33152685]
34. Liu J, Wang J. Users' intention to continue using online mental health communities: Empowerment theory perspective. Int J Environ Res Public Health 2021 Sep 07;18(18):9427 [FREE Full text] [doi: 10.3390/ijerph18189427] [Medline: 34574361]
35. Almutairi ILFH, Alazemi BF, Almutairi FLFH. Kuwaiti hospital patients' continuance intention to use telemedical systems in the wake of the COVID19 pandemic. Healthc Technol Lett 2021 Dec 08;8(6):159-168 [FREE Full text] [doi: 10.1049/htl2.12019] [Medline: 34938572]
36. Kao HY, Yang YC, Hung YH, Wu YJ. When does Da Vanci robotic surgical systems come into play? Front Public Health 2022 Jan 31;10:828542 [FREE Full text] [doi: 10.3389/fpubh.2022.828542] [Medline: 35174126]
37. Sun SL, Hwang HG, Dutta B, Peng MH. Exploring critical factors influencing nurses' intention to use tablet PC in patients' care using an integrated theoretical model. Libyan J Med 2019 Dec;14(1):1648963 [FREE Full text] [doi: 10.1080/19932820.2019.1648963] [Medline: 31357919]
38. Alam MZ, Khanam L. Comparison of the young aged and elderly female users' adoption of mHealth services. Health Care Women Int. Epub ahead of print 2022 Feb 23:1-25. [doi: 10.1080/07399332.2022.2039149] [Medline: 35195507]
39. Riefle L, Benz C. User-specific determinants of conversational agent usage: A review and potential for future research. In: Innovation Through Information Systems. 2021 Presented at: WI 2021: International Conference on Wirtschaftsinformatik; March 9-11, 2021; Germany p. 115-129. [doi: 10.1007/978-3-030-86797-3_8]
40. Gong E, Baptista S, Russell A, Scuffham P, Riddell M, Speight J, et al. My Diabetes Coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: Randomized effectiveness-implementation trial. J Med Internet Res 2020 Nov 05;22(11):e20322 [FREE Full text] [doi: 10.2196/20322] [Medline: 33151154]
41. Ponathil A, Ozkan F, Bertrand J, Agnisarman S, Narasimha S, Welch B, et al. An empirical study investigating the user acceptance of a virtual conversational agent interface for family health history collection among the geriatric population. Health Informatics J 2020 Dec;26(4):2946-2966 [FREE Full text] [doi: 10.1177/1460458220955104] [Medline: 32938275]
42. Hurmuz MZM, Jansen-Kosterink SM, den Akker HO, Hermens HJ. User experience and potential health effects of a conversational agent-based electronic health intervention: Protocol for an observational cohort study. JMIR Res Protoc 2020 Apr 03;9(4):e16641 [FREE Full text] [doi: 10.2196/16641] [Medline: 32242517]
43. Spinazze P, Aardoom J, Chavannes N, Kasteleyn M. The computer will see you now: Overcoming barriers to adoption of computer-assisted history taking (CAHT) in primary care. J Med Internet Res 2021 Feb 24;23(2):e19306 [FREE Full text] [doi: 10.2196/19306] [Medline: 33625360]

Abbreviations

AVE: average variance extracted
CA: conversational agent

XSL•FO
RenderX

ECM: expectation-confirmation model
ENT: ear, nose, and throat
PLS-SEM: partial least squares structural equation model
TAM: Technology Acceptance Model

Edited by G Eysenbach; submitted 30.06.22; peer-reviewed by M Rampioni, M Ozan-Rafferty, M Mbwogge; comments to author 23.07.22; revised version received 31.08.22; accepted 20.10.22; published 07.11.22.

Please cite as:
Li X, Xie S, Ye Z, Ma S, Yu G
Investigating Patients' Continuance Intention Toward Conversational Agents in Outpatient Departments: Cross-sectional Field Survey
J Med Internet Res 2022;24(11):e40681
URL: https://www.jmir.org/2022/11/e40681
doi:10.2196/40681
PMID:36342768

©Xingyi Li, Shirong Xie, Zhengqiang Ye, Shishi Ma, Guangjun Yu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Design and Evaluation Challenges of Conversational Agents in Health Care and Well-being: Selective Review Study

Ahmet Baki Kocaballi[1,2*], MSc, PhD; Emre Sezgin[3,4*], PhD; Leigh Clark[5*], PhD; John M Carroll[6], PhD; Yungui Huang[4], MBA, PhD; Jina Huh-Yoo[7], MHCI, PhD; Junhan Kim[8], MEng; Rafal Kocielnik[9], PhD; Yi-Chieh Lee[10], PhD; Lena Mamykina[11], PhD; Elliot G Mitchell[11,12], PhD; Robert J Moore[13], PhD; Prasanth Murali[14], MS; Elizabeth D Mynatt[14], PhD; Sun Young Park[8,15], PhD; Alessandro Pasta[16,17], MSc, PhD; Deborah Richards[18], BBus, MAppSc, PhD; Lucas M Silva[19], MSc; Diva Smriti[7], MSc; Brendan Spillane[20], PhD; Zhan Zhang[21], PhD; Tamara Zubatiy[22]

[1]School of Computer Science, University of Technology Sydney, Sydney, New South Wales, Australia
[2]Centre for Health Informatics, Macquarie University, Sydney, New South Wales, Australia
[3]Center for Biobehavioral Health, The Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus, OH, United States
[4]IT Research & Innovation, The Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus, OH, United States
[5]Bold Insight UK, London, United Kingdom
[6]College of Information Sciences and Technology, Pennsylvania State University, University Park, PA, United States
[7]College of Computing and Informatics, Drexel University, Philadelphia, PA, United States
[8]School of Information, University of Michigan, Ann Arbor, MI, United States
[9]California Institute of Technology, Pasadena, CA, United States
[10]Department of Computer Science, National University of Singapore, Singapore City, Singapore
[11]Department of Biomedical Informatics, Columbia University, New York City, NY, United States
[12]Steele Institute for Health Innovation, Geisinger, Danville, PA, United States
[13]IBM Research - Almaden, San Jose, CA, United States
[14]Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States
[15]School of Art and Design, University of Michigan, Ann Arbor, MI, United States
[16]Cognitive Systems Section, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark
[17]Demant A/S, Smørum, Denmark
[18]School of Computing, Macquarie University, Sydney, New South Wales, Australia
[19]Department of Informatics, University of California, Irvine, Irvine, CA, United States
[20]ADAPT Centre, University College Dublin, Dublin, Ireland
[21]School of Computer Science and Information Systems, Pace University, New York City, NY, United States
[22]Georgia Institute of Technology, Atlanta, GA, United States
[*]these authors contributed equally

Corresponding Author:
Ahmet Baki Kocaballi, MSc, PhD
School of Computer Science
University of Technology Sydney
Room CB11.7.103
81 Broadway
Sydney, New South Wales, 2007
Australia
Phone: 61 295141103
Email: baki.kocaballi@uts.edu.au

Abstract

Background: Health care and well-being are 2 main interconnected application areas of conversational agents (CAs). There is a significant increase in research, development, and commercial implementations in this area. In parallel to the increasing interest, new challenges in designing and evaluating CAs have emerged.

XSL•FO
RenderX

Objective: This study aims to identify key design, development, and evaluation challenges of CAs in health care and well-being research. The focus is on the very recent projects with their emerging challenges.

Methods: A review study was conducted with 17 invited studies, most of which were presented at the ACM (Association for Computing Machinery) CHI 2020 conference workshop on CAs for health and well-being. Eligibility criteria required the studies to involve a CA applied to a health or well-being project (ongoing or recently finished). The participating studies were asked to report on their projects' design and evaluation challenges. We used thematic analysis to review the studies.

Results: The findings include a range of topics from primary care to caring for older adults to health coaching. We identified 4 major themes: (1) Domain Information and Integration, (2) User-System Interaction and Partnership, (3) Evaluation, and (4) Conversational Competence.

Conclusions: CAs proved their worth during the pandemic as health screening tools, and are expected to stay to further support various health care domains, especially personal health care. Growth in investment in CAs also shows the value as a personal assistant. Our study shows that while some challenges are shared with other CA application areas, safety and privacy remain the major challenges in the health care and well-being domains. An increased level of collaboration across different institutions and entities may be a promising direction to address some of the major challenges that otherwise would be too complex to be addressed by the projects with their limited scope and budget.

(J Med Internet Res 2022;24(11):e38525) doi:10.2196/38525

KEYWORDS

conversational interfaces; conversational agents; dialog systems; health care; well-being

Introduction

Conversational agents (CAs) are applications that facilitate human-computer interaction through natural language. Automatic speech recognition (ASR) and natural language processing (NLP) models help to interpret human language and produce appropriate responses [1]. CAs (also widely known as chatbots, virtual assistants, dialog systems, or voice assistants) are used in several domains such as e-commerce, scheduling services, and question-answer systems [2]. The user-system interaction could be over text (eg, SMS text messaging over an app or web service), voice (eg, interactive voice response via phone calls, voice assistants via a smartphone or smart speaker), or multimodal (eg, visual, text, and audio feedback and interaction via a smartphone, smart speaker, or any other smart and internet of things devices) [3].

CAs have been variously used and studied in health care for supporting behavioral health and healthy living [3,4]; health information seeking [5-9]; appointment, medication, symptom tracking and chronic condition management [10-12]; and facilitating COVID-19 screening and information sharing [13,14]. Mobile phone ownership enables and increases the potential applications, availability, and access to CAs in practice. Mobile phone ownership is around 15 billion worldwide as of 2021 [15], and 97% of US adults own a mobile device [16]. Current studies and randomized trials showed that CAs could be effectively used in health care delivery and improving health outcomes, such as improving mental health [17], maternal health [18], and healthy behaviors [19]. In addition, there is an increasing investment in chatbots in the health care industry; some examples include Woebot, Babylon, and ADA Health [20].

Despite the increasing interest in using CAs in supporting health care and well-being, there are many challenges in the development, deployment, and use of CAs. Recent review studies have highlighted some of the challenges including NLP [21,22], patient safety [1,23,24], integration with other technologies [22], information dissemination [25,26], medico-legal issues [27], and ethics [28]. In response to that, recent workshops have explored the challenges and opportunities of conversational user interfaces [29-31] and the design and evaluation of CAs in health care [32]. Here we contribute to this developing literature by reporting a self-assessment of 17 such projects.

In May 2020, a workshop entitled "Conversational Agents for Health and Well-being" was held at the *ACM (Association for Computing Machinery) Conference on Human Factors in Computing Systems (CHI 2020)* [32]. The aim was to understand the most current challenges recent research projects face and devise potential directions for future research to address those challenges. The workshop included completed or ongoing projects from 30 participants in 5 countries, covering various topics from supporting older adults to mental health and coaching to supporting everyday health. Following the workshop, participants were invited to collectively report the design and evaluation challenges of CAs in health care to provide researchers, designers, and health care professionals practical perspectives on these challenges. This paper aims to present the challenges of designing and evaluating CAs derived from recent health care projects conducted in the last 2 years.

Methods

We followed a selective review study design focusing on the challenges of recent studies on CAs in health and well-being. Coauthors were invited to report their original CA research in health care and well-being, outlining major challenges in the design and evaluation of the CA being used in their research. Coauthors were asked to report on design and evaluation challenges they faced in their project. In their written report, each coauthor or author group (1) described their research, (2) explained challenges (limited to 3 major design challenges and 3 major evaluation challenges), (3) explained how they

XSL•FO
RenderX

addressed the challenges or how they plan to address, and (4) support their findings and suggestions with prior literature. Only the information provided through the written report was analyzed as a case study. Each case study went through an open peer review process among authors and was revised. Finalized cases were analyzed by 3 coauthors (ABK, ES, and LC). We used thematic analysis to identify, assess, and analyze the patterns in the cases [33]. The following steps were used in the analysis process: (1) familiarizing with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing and refining themes, (5) defining and naming themes, and (6) reporting the findings (Textbox 1).

Textbox 1. Steps used in the analysis process.

1. Familiarizing with data
   - To gain familiarity with the data and understand the depth and breadth of the content, coauthors (ABK, ES, and LC) read and re-read the case studies.
2. Generating initial code
   - Following an open-coding approach (without predefined codes, developed, and modified during the coding process), coauthors (ABK, ES, and LC) created initial codes independently. They reviewed the codes iteratively. The codes were compared, and group decisions and consensus created the finalized codes. Coauthors used Google Sheets to create codebooks.
3. Searching for themes
   - The codes were sorted at first to understand the frequency of occurrence. They were reviewed to find patterns and grouped into the themes collectively by the coauthors (ABK, ES, and LC). Each theme was labeled to guide the grouping and reviewed by the coauthors (ABK, ES, and LC) iteratively. Similar to the coding, coauthors reached a consensus to finalize themes.
4. Reviewing and refining themes
   - All themes were reviewed. The relationships between codes and themes were discussed by the coauthors (ABK, ES, and LC). Some of the themes were combined that were found to be related; for example, obtaining domain information and training data were combined into Domain Information and Training. Themes were reviewed and finalized for consistency regarding their content, by consensus.
5. Defining and naming themes
   - Definitions of the themes were created regarding codes, subthemes, and corresponding cases. In case of disagreements, coauthors reviewed themes to ensure consensus in theme content, definition, and labeling.
6. Reporting the findings
   - Thematic analysis results were reported through a chart with themes, subthemes, definitions, associated cases, and frequency of occurrences.

Results

Overview

The review included 17 studies covering many domains from primary care to caring for older adults to health coaching. A summary of the projects is presented in Table 1, with a unique project number for each project to be referenced in the presentation of results. The thematic analysis identified 4 major themes: (1) Domain Information and Integration, (2) User-System Interaction and Partnership, (3) Evaluation, and (4) Conversational Competence (Table 2).

XSL•FO
RenderX

Table 1. A summary of participating projects, including titles, domain, CA[a] purpose, and CA input and output modalities.

Project number	Project title	Health/well-being domain	CA purpose	CA input/output	Project status
1	Digital Scribe: A Wizard of Oz Study	Primary care	Work with general practitioners to document patient information in a consultation	Written and spoken/written, spoken, and visual	Ongoing
2	Speech Diversity and Speech Interfaces – Considering an Inclusive Future Through Stammering [34]	Accessibility and inclusivity/speech diversity	All speech-based CAs (nonspecific projects)	Spoken/written, spoken, and visual	Ongoing
3	ADELE: An Artificial Conversational Care Agent for the Elderly [35-37]	Care of the elderly in their homes	Provide health and well-being care, advice, and monitoring	Written/written	Completed
4	Talk the Talk: How Human Conversational Agents Build Trust [38-41]	People with visual impairments	Support navigation and other activities by having the person with visual impairment share their smartphone camera feed. Based on the camera feed and conversational interaction, plus use of online tools such as Google Maps, a remotely located sighted person provides guidance to the person with visual impairment.	Spoken/spoken	Ongoing
5	Empowering Older Adults With Mild Cognitive Impairment and Their Caregivers Using Conversational Agents	Older adults with MCI[b]	Empower the person with MCI as well as their caregiver; amplify caregiver	Spoken/spoken and visual	Ongoing
6	Adaptive Conversational Agents for the Health Care Enterprise	Health care enterprise, payer	Help users answer analytical questions about health care enterprises	Written and spoken/written, spoken, and visual	Completed
7	Encouraging Users' Self-disclosure With a Chatbot Mediator [42-44]	Mental health care	Collect truthful self-disclosure and deliver guidance	Written/written	Ongoing
8	Motivational Interviewing Conversational Agent (MICA) [45,46]	Family eating habits	Deliver a counseling method named motivational interviewing in an automated manner to help parents eat healthier along with their children	Spoken/spoken	Completed
9	Conversational Agent for African Americans With Chronic Illnesses [47,48]	Chronic illness	Deliver health information on COVID-19 to African Americans with chronic illnesses	Written/written	Completed
10	Exploring Voice Assistants in Multimodal Food Journaling [49,50]	Food journaling	Entries in food journal	Spoken/written and spoken	Completed
11	t2.coach: A Chatbot Health Coach for Diabetes Self-management [51,52]	Health coaching	Health coaching and goal setting for type 2 diabetes self-management	Written/written and visual	Ongoing
12	Designing Audiologist Bots Fusing Soundscapes and User Feedback	Hearing health care	Recommending personalized hearing aid settings by gathering user feedback in real-world environments	Written and spoken/written, spoken, and visual	Ongoing
13	Utilization of Self-Diagnosis Health Chatbots in the Wild: A Case Study [53]	Self-diagnosis	Offering medical advice (eg, diagnostic suggestions) to patients based on their input (eg, symptoms)	Written and spoken/written	Completed
14	eADVICE: Providing Specialist Treatment Advice to Patients on Waiting List [54]	Children and their families referred to incontinence and sleep clinic awaiting appointment	Discuss treatments to encourage adherence	Written/written and spoken	Ongoing

XSL•FO
RenderX

Project number	Project title	Health/well-being domain	CA purpose	CA input/output	Project status
15	Symptom and Health Events Tracking at Home for Children With Special Health Care Needs (CSHCN) Using Conversational Agents [55,56]	Documentation and care coordination support for children with special health care needs	Assisting caregivers and patients with tracking and communicating symptoms and health events outside of clinical settings to reduce documentation burden and facilitate care coordination	Written and spoken/written and spoken	Completed
16	Palliative Care With Spiritual Support by Conversational Agents	Elderly care, spiritual support, human-agent/robot interaction	Development of a CA that provides end of life planning and spiritual counseling	Written/spoken and visual	Ongoing
17	HarborBot: A Chatbot for Social Needs Screening [57,58]	Social needs screening in public hospital emergency departments	Collect high-quality social needs data from vulnerable populations while increasing engagement	Written/written and spoken	Completed

[a]CA: conversational agent.

[b]MCI: mild cognitive impairment.

Table 2. A summary of themes and subthemes and the studies reporting them.

Themes and subthemes	Reported by
Domain Information and Integration	
Domain Information and Training	P2[a], P6, P10, P11, P13, P14, P17
Integration and Infrastructure	P11, P14, P15
User-System Interaction and Partnership	
Personalization	P3, P5, P6, P8, P9, P11, P12, P13, P14
Relationship Building	P3, P4, P7, P12, P13, P16, P17
Safety and Privacy	P3, P5, P7, P8, P9, P10, P13
User Engagement	P4, P7, P9, P10, P13, P14
Evaluation	
Methodological Limitations	P1, P8, P11, P12
Experimental Limitations	P1, P3, P5, P12, P13, P14, P17
Lack of Guidance on Evaluation	P1, P2, P9, 13, P15
Conversational Competence	
Topic Detection and ASR[b]	P1, P2, P3, P8, P10, P12, P15
Discoverability and Conversational Interaction Model	P10, P11, P1, P2, P17
Accessibility and Inclusivity	P2, P5, P8, P9, P14, P16, P17

[a]P: project.

[b]ASR: automatic speech recognition.

Domain Information and Integration

Overview and Subthemes

Health care CAs often operate within particular domains of health care that require the integration of domain-specific information and language. This theme is concerned with the challenges of obtaining the required problem domain information, collecting data to train CAs, and integrating CAs with the existing systems and infrastructures. For example, a CA to be designed for helping general practitioners in the primary care domain needs to be trained by a large number of doctor-patient conversations. Obtaining this kind of information is challenging and resource intensive. In addition, the expert medical knowledge needs to be translated into conversational format. Similarly, CAs cannot be developed as isolated applications: they need to be integrated with the existing systems and infrastructure.

Domain Information and Training

Two projects (P6 and P14) reported on the difficulties of obtaining domain information, in these cases, for health care enterprise and medical adherence communications. These difficulties are associated with the lack of time of domain experts (P6), the domain information being not in a conversational format (P6) or being distributed across many

XSL•FO
RenderX

subdomains (P6), and the knowledge acquisition bottleneck (P14). While P6 recruited subject matter experts to review conversation flows and response frames to format health care data for conversation-like interactions, P14 implemented incremental acquisition from the expert or user to alleviate the problem of obtaining domain-specific information.

Integration and Infrastructure

Three studies reported some challenges associated with integrating CAs into existing infrastructures. One project (P15) using CAs to support health care tracking of children with special needs at home explained the security and privacy challenges of integrating CA-collected information into clinical systems and workflows. Using interoperability standards (eg, Fast Healthcare Interoperability Resources) to share caregiver notes with health care professionals was suggested as one of the fundamental approaches for integration (P15). Many CAs are developed using underlying development platforms such as Amazon Alexa Skills or Google Dialogflow. P14 highlighted the difficulty in adapting to the rapidly evolving platforms that can potentially make the CAs designed according to the previous platform functionality stop working. Finally, P11 discussed the tension between the ubiquity and richness of the underlying platforms. For example, while SMS-based text messaging is ubiquitous, it is purely text based and does not offer some of the multimodal interaction options of other messaging platforms such as Facebook Messenger or WhatsApp (eg, 1-click suggested response buttons or carousel menus). P11 used SMS text message–based messaging to support a higher degree of accessibility and ubiquity; however, to expand the richness of the user experience they included more immersive education content with multimedia messages including infographics to elaborate on each health goal option.

User-System Interaction and Partnership

Overview and Subthemes

This was one of the major themes that captured several challenges related to the characteristics and qualities of interaction between users and CAs, and the ways in which users and CAs work together. The subthemes included Personalization of CAs, Relationship Building Between CAs and Users, Safety and Privacy, and User Engagement. Some common challenges were supporting users' trust in CAs, enabling CAs to show empathy, and ensuring users' privacy.

Personalization

Personalization emerged as an essential design feature as part of the projects. Personalization was reported as a challenge in 10 projects. The personalization challenges were providing appropriate responses based on users' context (P8 and P3), tailoring conversations for different user groups with a broad age range and different health literacy levels (P14, P6, P9, P11, and P12), and minimizing the question overload in surveys (P13). While 1 project reported on the difficulties in evaluating the effects on adaptive features (P6), another project presented the challenges associated with designing for dyads (P5), which include a patient and a caregiver. P5 explained that dyads using the same device pose several challenges, including dyads' difficulties managing technology, dyads' overall wide range of technology literacy, and how to support both members of the dyad through the functionality of the CA.

Relationship Building

Humans and CAs may have different forms of relationships ranging from very short-term relationships typically characterized by one-off task-based conversational exchanges to long-term relationships in which CAs have longer conversational interactions across different topics over a longer period. Relationship building—how a human-CA relationship is established and maintained—was reported as a challenge in 13 projects. Trust (P3, P4, P7, P13, and P16), empathy (P4, P17, and P16), self-disclosure (P7), and transparency (P12) were presented as important dimensions of relationship building. One project focusing on designing a CA for people with visual impairment (P4) explains the importance of CAs to show empathy in establishing and strengthening trust between the sighted Aira agent and the client with visual impairment. According to P4, to incorporate empathy into conversational interaction, an agent must represent not only the situation itself (eg, what objects are present, their spatial relationships, movement vectors) but also the other agent's experience and interpretation of the situation. A project designing a CA for older adults' care (P3) describes the difficulties of building trust due to concerns about security, legal issues, the sharing and storing of personal and sensitive information, and privacy and ethical concerns, among others [59]. Issues with any of these and other concerns have the potential to damage the relationship critically or terminally between the CA and the patient. To address this challenge, P3 incorporated trust-building and repair strategies from the onset. Previous research has shown that reliability is likely the preeminent factor in building trust between the patient and the artificial care agent [60,61]. By contrast, other factors such as competence, benevolence, and integrity are also likely to significantly influence [62]. Transferring users' self-disclosure was reported as a challenge in human-in-the-loop artificial intelligence systems where CAs mediate between a user and a domain expert (P7).

Safety and Privacy

Safety of user data and privacy of personal information shared with CAs have influenced the user decision and perception toward CAs. Six projects reported users' safety (P3 and P13) and privacy (P5, P7, P8, and P10) as challenges. P10 explained that while some users were worried that using the voice assistant would disturb others, other users felt discomfort or concerned with privacy when other people could hear them track their food. To tackle this issue, some suggested solutions in P10 included adopting other devices or modalities for input (eg, taking picture or text input using phone or web), as well as implementing food "template" features for quick commands (eg, saying "Alexa, journal number 1" or "Google, journal same breakfast as yesterday"). In the cases of CAs as mediators between 2 users (P7) or multiple users (P5), privacy challenges may become more critical. As many health care CAs deal with safety-critical user information and decision making, P3 and P13 noted the need for more strict and standardized evaluation measures for CAs.

XSL•FO
RenderX

User Engagement

Establishing user engagement strategies for CAs is fundamental to improving user experience and sustained use. However, 6 projects reported enabling and maintaining user engagement as a challenge (P4, P7, P9, P10, P13, and P14). P9 explained the importance of culturally sensitive CAs to support increased trust and adoption in the context of helping African Americans with chronic diseases and the difficulties of understanding and incorporating cultural aspects into CAs. P10's focus on creating a multimodal food journal mentioned their participants' problem of remembering or discovering the voice commands to track their food, and the need to have a better mapping of commands to multiple utterance styles and advanced intent recognition. P13, a project using a self-diagnosis chatbot, reported that their users tended to drop out of the consultation with their chatbot, especially during the early stages, and pointed out the importance of examining and evaluating the mechanisms and approaches that can increase the uptake and utilization of health chatbots. P14 incorporated various strategies into the chatbot to support higher treatment adherence for pediatric patients awaiting a specialist appointment. These included developing a working alliance; having face-to-face communication; using everyday conversational language; and empathic language strategies such as choices, consequences, and nonjudgmental affirmations.

Evaluation

Overview and Subthemes

The Evaluation theme encapsulates 3 limitations that authors have encountered during their experiments with CAs: (1) methodological limitations that show challenges in evaluating interactions and performance of CAs; (2) experimental limitations related to the challenges with data collection and analysis and study environment; and (3) lack of guidance in evaluation describing challenges to navigate assessment of CAs without guidance or prior evidence.

Methodological Limitations

By nature, CA interactions are designed to communicate with humans and are dependent on intellectually crafted bidirectional conversations that are hard to generate or replicate. Early efforts toward testing the CAs included scripted conversations to measure the performance of CAs objectively. Yet, scripted conversations fall short in mimicking the actual number of conversations and iterations that may occur in the real world (P1). In addition, a limited number of interactions create a barrier to evaluating the performance effectively. Creating complex scripts or testing with actual patients are some of the solutions proposed (P1).

To provide standard measures to evaluate the performance of CAs, user-CA interactions can be tested in a controlled environment using scripts, such as scenarios, role-playing, or Wizard of Oz testing. However, these scripted or simulated interactions might be limited to providing organically flowing conversations (P1) and accurate assessments due to the simulated nature of system functionalities and user scenarios (P8). Such evaluation yields the results only in a controlled environment, impacting the end user's judgment (P8). The observations on user-CA interactions may explain a CA's efficacy to a degree yet not its effectiveness, which can be observed in the real-world environment. Using actual end users with different health and well-being needs in real-world settings and observing them longitudinally may improve the evidence of CA interactions (P1, P11, and P14).

Experimental Limitations

Effects of Training Materials/Unplanned Events

Training data are core requirements in developing and assessing CAs performance. However, it could be hard to gauge the effects of training materials (P5). Developers must track the features used in training and compare them with the actual CA usage. Similarly, unplanned or unforeseen events can affect the outcomes of CA interaction evaluations. The COVID-19 pandemic shifted the norms toward remote management of experiments, causing miscommunications and inefficiencies in simple user training and troubleshooting (P5).

Accessing Vulnerable Populations

It is important to design for diverse populations, including vulnerable populations, and people with various social ills, such as homelessness, poverty, and hunger [58]. However, recruiting and engaging them with limited resources and connections with community partners can be difficult. In terms of improving CA interaction with such populations and enhancing access to the CA, inputs from heterogeneous groups and different users are needed through alternative platforms and technologies (P12 and P14). For instance, CA use evaluation in an emergency department with low-literacy users may require trained personnel to guide users and understand their experiences (P17).

Challenges of Testing in Real-World Settings

In research, test settings and iterations often occur in controlled environments, without real-world interactions to objectively assess CAs. This results in decontextualization of testing and yields limited results. Some significant difficulties are involved in performing CA testing in real-world, authentic settings. For instance, P1, in a study with CA-supported automated documentation, reported that without a real-world electronic health record interaction, many work routines cannot be tested. Even if a CA gets integrated into electronic health record, technically, logistically, and legally it is hard to roll out. However, such efforts are necessary for understanding all end users' perceptions (nurses, doctors, patients) and the complexity of medical workflow (P17). P3 suggested a staged evaluation approach through a research platform, which will allow evaluations in mock care settings. There are still very few user-CA evaluations in real-world settings. Further efforts are required to promote real-world testing (in uncontrolled authentic environments). P13 explained that without an in-depth understanding of contextual elements in a problem domain, it is challenging for CA designers and developers to figure out how to improve the user experience and how to overcome the challenges in the actual use of health care CAs.

XSL•FO
RenderX

Lack of Guidance in Evaluation

Lack of Evaluation Data for Special Population Groups

As with vulnerable populations, there is a lack of evaluation data for special population groups. This issue may lead to having no real knowledge of user experiences with CAs for people with diverse speech patterns (P2). Similarly, there is a lack of evaluation material designed (P13) specifically for marginalized or minor user groups, such as African Americans with low technology literacy (P9). Development of population-specific evaluation methods and promotion of participatory design and interactive sessions are necessary (P2 and P9).

Lack of Evaluation Guidelines and Metrics

The lack of evaluation guidelines for CAs leads to creating robust frameworks for effectively and uniformly measuring impact and outcomes. In addition, the CA-based interventions' safety, efficacy, and effectiveness are lacking due to no clear guidance provided in the literature yet (P13). Metrics to measure the effects of interaction, engagement, and measures for health outcomes are necessary. One solution could be longitudinally observing interaction to identify key indicators to be used as success metrics (P15).

Theoretical approaches to evaluating the CAs may not be aligned with the standard measures. In P11, behavior change techniques showed a mismatch with usability metrics, potentially showing an inverse relationship between user engagement (quality of communication in user experience and CA usage patterns) and behavior change techniques (eg, goal setting). Combining theories and evaluation approaches might be necessary to create multifaceted evaluation metrics and triangulate use patterns.

Difficulties in Multimodal Testing

CA could be potentially provided in multiple platforms and different modalities (eg, text-based chatbot, voice assistant with avatar). Evaluating individual modalities in a multimodal system is necessary, yet it is hard to evaluate the graphical and conversational interfaces separately. However, it might still be possible to design visual elements and layouts in a minimal way to reduce their effects on users' perception and assessment of the CA interface (P1).

Conversational Competence

Overview and Subthemes

Several projects discussed challenges related to the conversational competence of CAs and the impacts these might have on people's interactions. Competence here refers to accurately understanding user input and responding appropriately, whether the conversation is an appropriate metaphor for design and how user interactions can be best facilitated, and how CA interactions can be made more accessible and inclusive.

Topic Detection and Automated Speech Recognition

Detecting the topic being discussed can be difficult in several CA scenarios. The unscripted nature of the social talk, often undertaken as part of caregiving interactions, makes it difficult to follow the discussed topics (P3). Even more scripted interactions such as primary care consultations may be nonlinear and fragmented, creating further difficulties in detecting the current topic being discussed (P1). Proposed solutions include advanced topic detection methods (P3) and collaboratively built data sets to support these methods (P1). Using specific phrases to highlight topic shifts during interactions (P1) and personalized systems (P3) may also improve CAs.

Topic detection in CA in health care settings also faces challenges related to specific contexts and users. For example, tracking health care outcomes for different patients and treatments requires an understanding of specific medical terms (P15). Using services such as Amazon Comprehend Medical [63] alongside manual interventions could help expand appropriate vocabulary to improve CA comprehension. Similarly, P12 describes using CAs for people with hearing aids that may require them to use an agreed set of terms or additional supervised training to recognize "audiological intents."

More generally, there is an ongoing challenge of understanding when someone has finished an utterance while communicating with a CA (endpoint detection). This may require additional research for people with diverse speech patterns such as stammering (P2), to collect the necessary audio data and understand their interactions.

In addition to detecting topics in interaction, CAs in health care interactions face difficulties with ASR. This is also common in CA interactions outside of health care [64]. Audio may be impacted by recording quality ambient noise (eg, other devices such as televisions or other people talking), which could be combated with the use of directional microphones (P1 and P15) [21]. Again, in more nuanced interactions such as using CAs for people with hearing aids, optimal CA responses may require the processing of environmental information such as loudness and signal-to-noise ratio (P12). Inserting such information into the dialog may improve CA performance for this type of interaction.

In addition to audio quality, the language a user produces can also be a limiting factor for CAs. Active error correction by a CA may help in the misunderstanding and nonrecognition of specific terms (P10). For privacy-first interactions, processing of speech data may be performed locally on the device rather than requiring any interaction with servers, though this may reduce the performance of the CA's speech processing capabilities (P8).

Discoverability and Conversational Interaction Model

A key challenge in CA design is making the set of possible actions or commands discoverable for users [65]. Understanding and remembering how to interact with CAs in different contexts can create difficulties for people using these systems (P10 and P11). Making CAs open ended and lightweight may allow people to explore systems' capabilities and adapt them for their purposes (P10). For people with low levels of technology literacy, CA-initiated dialogs can be implemented at consistent intervals to counteract the lack of discoverability of user-initiated features (P11).

CAs rely on turn-taking–based conversational communication and interaction models; however, they may not always be appropriate or require additional scaffolding. General

XSL•FO
RenderX

practitioners taking notes using a CA, for example, may need to interrupt interactions with their patients. By contrast, screen-based technologies for the same task can support multimodal, more continuous interactions, and less intrusive data entry (P1). Continuing the multimodal use of information entry for general practitioners may improve or resolve interruptions. Furthermore, monitoring utterances in user-CA interactions may not be as smooth as human-human interactions. Consequently, enabling a CA to detect when someone has finished speaking is critical to more seamless interactions (P2). For diverse speech patterns, such as stammering, we may be able to draw on advice for interacting with various demographics, though crucially, we must understand the nuances of these interactions and design CAs with inclusivity in mind. For scenarios such as filling in forms and surveys with CAs, audio may be an inefficient modality and create design tension with sequential questions and wait times (P17). Optimizing delay usage to limit unnecessary wait times, using shorter phrases, and allowing people to opt-out of audio in multimodal CAs can be used to work against these limitations.

Accessibility and Inclusivity

Several projects discuss the need to make CAs accessible to a broad set of demographics or to focus on improving systems to make them inclusive for specific types of people. P14 identifies the need to provide equitable and easy access to CAs that improve health and well-being. They propose allowing access through a web browser and downloadable software, as well as offering technical support and downloadable fact sheets to work with or replace their CA. Off-the-shelf CAs may present difficulties for people in lower-resource areas (eg, limited internet access, financial constraints), which could be improved with systems that use more offline resources (P8), though this may present its own performance challenges that need to be addressed.

Levels of education, experience, and technological literacy can also impact users' interactions with CAs (P5, P16, and P17). These obstacles may be overcome by providing multiple modalities to compensate for people's preferences and abilities (P5 and P17) and engaging in participatory design with target demographics (P16). Evaluating such systems may require redeveloping protocols to include people with lower literacy and ensure they are able to understand the questions being posed to them (P17). Specific communities may also require additional thought in designing and evaluating CAs and building on the available sparse research literature (P2 and P9). Working with people who stammer, for example, has seen little research with CA interactions and requires a fundamental understanding of the barriers to successful interactions and how these can be overcome (P2). P9 highlights the need to consider African American communities and how their perceptions towards CA and interactions with them cannot be assumed to be the same as other communities. Consequently, novel methods and evaluation techniques may need to be developed with diverse populations in mind.

Discussion

Principal Findings

This review highlights numerous challenges of CA interactions in the health care and well-being fields, including 4 major themes on Domain Information and Integration, Conversational Competence, User-System Interaction and Partnership, and Evaluation. Many challenges reported in this review echo those discussed in related CA work. ASR errors are long-standing concerns with speech-based CAs [66] that can lead users to alter their speech patterns to increase comprehension [67]. Difficulties in fostering user engagement have been highlighted when CAs cease to perform their intended utility [68]. Existing research has addressed this issue, for example, by explaining the cause of interaction errors to users or allowing users to halt interactions when they desire [69]. Accessibility concerns discussed in this review also map to those discussed in prior work on both text-based [70] and speech-based CAs [71]. Research continues to identify how CA accessibility can be improved [72] and design recommendations such as the Web Content Accessibility Guidelines [73] have been suggested as a means of addressing these. Prior studies also reported challenges associated with involving and engaging patients in their home environments [74], limitations in short-duration laboratory studies and challenges in longitudinal assessments [75], and difficulties with multicomponent system evaluation [76].

In addition to the similar challenges identified in prior CA literature, this review identifies some challenges that are specific or more critical to health care and well-being domains, including empathy, safety, recruitment of vulnerable populations, and challenges in testing authentic settings. Prior work has brought attention to open challenges across the broader CA field [77] and highlights the need to create responsible CAs that focus on fairness, transparency, and ethics. This is critical for CAs discussed in this review, particularly given the sensitive nature of interactions and the underlying data collected from these. The fragmented nature of speech-based CA work, in particular, has recently been addressed [29,78]. This prior work notes a scarcity of robust evaluation metrics and research involving real-world testing. Our review shows that there are similar problems for application of CAs in health care and well-being. However, the nature of these domains means there can be additional obstacles when entering research phases that require strict considerations of medical ethics, laws, and standardized practices. Improving consistency in CA research and implementation may benefit from reaching out to other disciplines (eg, cognitive sciences, linguistics) [78], combining existing theories and evaluation approaches (see the "Lack of Guidance in Evaluation" section), and examining guidelines from both academia and industry [79]. While broad challenges around CAs must be considered within health care and well-being contexts, this review also identifies more nuanced challenges for these domains. Future work should consider how these different challenges overlap, depending on both the expected CA scenarios and user demographics.

XSL•FO
RenderX

Data Collection in Health Care Domains

Obtaining domain-specific information and collecting data are typically challenging. There are 2 major reasons: first, health care professionals who can provide critical domain information are very time-poor and their contribution to the projects is limited; and second, health care data are extremely sensitive, and there are a lot of privacy and safety concerns. As recruiting a full-time health care professional is not financially feasible, an incremental acquisition method might be useful (P14). Data privacy and safety are significant concerns for health care professionals, and consumers are more hesitant to share information compared with other application areas (P1). For example, the Digital Scribe project needed to collect thousands of doctor-patient conversations in primary care settings to train their NLP algorithm [21]. However, managing the audio recording system, informing patients and getting their consent, and privacy and safety concerns of not only patients but also doctors made the data collection process extremely challenging. In this case, multi-institution collaborations are needed to organize and manage data collection to reduce the frictions in the process such as employing a dedicated technical person to provide support, more automated and easy-to-control audio recording system, and streamlined patient consent gathering. Such data collection challenges are not specific to primary care as many health care settings have similar situations and requirements. Therefore, incentivizing collaborations and data sharing between institutions and creating ethical frameworks to facilitate data sharing are needed, with some already emerging in the CA community [80]. In addition, proven data safety protocols and certifications can be made available to ensure safety and increase user trust. Integrating CAs into existing systems and enabling data sharing across systems were other challenges reported. Fast Healthcare Interoperability Resources was suggested as a promising interoperability standard (P15).

Developing Empathy in CA Interactions

Concerning the theme User-System Interaction and Partnership, empathy, safety, and privacy emerged as important challenges. Especially for CAs used in mental health applications, being able to show empathy is a valuable characteristic. Challenges related to empathy can be grouped into 2 categories: challenges of detecting the users' current emotional sensitivity to a topic and crafting an appropriate response to the user according to their situation and preferences. For example, P4, using their CA interface Aira, used client profiles (typical interaction contexts, preferred measurement units) to reference and incorporate these details in the interaction to signal empathy. P17 aimed their CAs to show empathy through more minimal means. This contrasts with other successful but more costly implementations of embodied CAs [81,82] in general medical contexts. P17 designed a series of neutral and empathetic reactions to user answers as well as other social utterances to augment the question administration dialog. They included phrases that help the user anticipate a sensitive topic before it is introduced (eg, "The next questions are about your personal safety and may be tough to answer."), provide acknowledgments for answering neutral questions (eg, "Okay, I'm getting a better idea of where you are at," "Got it"), and empathetic reactions to sensitive questions (eg, "That must be stressful, I'm sorry to hear that."). They also pointed out some users' different preferences, such as a lower level of socialization in chat than others.

Establishing Safety and Privacy in CA Applications

Safety and privacy concerns are more critical in health and well-being applications. A systematic review found that patient safety concerns were rarely addressed in the majority of the papers reviewed [1]. Similarly, Bickmore et al [24] found that CAs with unconstrained natural language input pose serious safety risks. Prior studies also found that commonly available voice assistants fail to answer appropriately to safety-critical user prompts [4,6]. Safety risks may occur due to misrecognized prompts, the inability to detect the severity of users' prompts (when there are no ASR errors), or gaps in clinical reasoning. A 5-stage framework for evaluating different aspects of symptom checkers might be adapted to be used for other CA applications [83]. Privacy of personal user data is also of utmost importance. Personal health data are considered one of the most sensitive types of information. Therefore, protecting the privacy of users' data becomes even more critical in health care CAs. Data privacy is particularly more relevant to CAs as many CAs process user prompts through third-party services in the cloud. Thus, there is a valid concern about the security of such information. Local NLP engines might be a solution to this (P8).

Limitations

This review included a limited number of studies in the emerging domain of health care and well-being CAs, which were conducted within a specific period and reported during a conference workshop. The main purpose is to provide a snapshot of some of the major challenges faced by the recent projects in this area rather than providing a comprehensive overview of the challenges. Although the reported challenges provide a useful overview of some significant challenges, they should not be considered the complete set of challenges in this domain. They represent the challenges faced by some recent projects conducted by the researchers actively working in this area across the globe.

Conclusion

This paper examined 17 recent studies of CAs in health care and well-being to identify design and evaluation challenges. While many challenges, including accessibility, personalization, and empathy provision, are shared with other application areas of CAs, safety and privacy remain the major challenges that are more critical in the health care domain.

CAs proved their worth during the pandemic as health screening tools, and are here to stay to further assist personal health care. Growth in investment in CAs also shows the value as a personal assistant. An increased level of collaboration across different institutions and entities may be a promising direction to address some of the major challenges that otherwise would be too complex to be addressed by the projects with their limited scope and budget.

XSL•FO
RenderX

Authors' Contributions

ABK contributed to the study design. ABK, ES, LC, JMC, JH-Y, YH, JK, RK, Y-CL, LM, EGM, RJM, PM, EDM, SYP, AP, DR, LMS, DS, BS, ZZ, and TZ contributed to data reporting. ABK, ES, and LC performed thematic analysis. ABK, ES, and LC proposed the first draft. ABK, ES, LC, JMC, JH-Y, YH, JK, RK, Y-CL, LM, EGM, RJM, PM, EDM, SYP, AP, DR, LMS, DS, BS, ZZ, and TZ performed revisions and finalized subsequent drafts.

Conflicts of Interest

None declared.

References

1. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
2. Comes S, Schatsky D, Chauhan R. Conversational AI. Deloitte Insights. 2021. URL: https://wwwdeloitte.com/us/en/insights/focus/signals-for-strategists/the-future-of-conversational-ai.html [accessed 2022-10-19]
3. Sezgin E, Militello LK, Huang Y, Lin S. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. Transl Behav Med 2020 Aug 07;10(3):606-628. [doi: 10.1093/tbm/ibz141] [Medline: 32766865]
4. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of Conversational Agents to Health and Lifestyle Prompts: Investigation of Appropriateness and Presentation Structures. J Med Internet Res 2020 Feb 09;22(2):e15823 [FREE Full text] [doi: 10.2196/15823] [Medline: 32039810]
5. Yang S, Lee J, Sezgin E, Bridge J, Lin S. Clinical Advice by Voice Assistants on Postpartum Depression: Cross-Sectional Investigation Using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e24045 [FREE Full text] [doi: 10.2196/24045] [Medline: 33427680]
6. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
7. Alagha EC, Helbing RR. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform 2019 Nov;26(1):e100075 [FREE Full text] [doi: 10.1136/bmjhci-2019-100075] [Medline: 31767629]
8. Nobles AL, Leas EC, Caputi TL, Zhu S, Strathdee SA, Ayers JW. Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. NPJ Digit Med 2020;3:11 [FREE Full text] [doi: 10.1038/s41746-019-0215-9] [Medline: 32025572]
9. Wilson N, MacDonald EJ, Mansoor OD, Morgan J. In bed with Siri and Google Assistant: a comparison of sexual health advice. BMJ 2017 Dec 13;359:j5635. [doi: 10.1136/bmj.j5635] [Medline: 29237603]
10. Zand A, Sharma A, Stokes Z, Reynolds C, Montilla A, Sauk J, et al. An Exploration Into the Use of a Chatbot for Patients With Inflammatory Bowel Diseases: Retrospective Cohort Study. J Med Internet Res 2020 May 26;22(5):e15589 [FREE Full text] [doi: 10.2196/15589] [Medline: 32452808]
11. Bérubé C, Kovacs, Fleisch, Kowatsch. Reliability of Commercial Voice Assistants' Responses to Health-Related Questions in Noncommunicable Disease Management: Factorial Experiment Assessing Response Rate and Source of Information. J Med Internet Res 2021 Dec 20;23(12):e32161 [FREE Full text] [doi: 10.2196/32161] [Medline: 34932003]
12. Hong G, Folcarelli A, Less J, Wang C, Erbasi N, Lin S. Voice Assistants and Cancer Screening: A Comparison of Alexa, Siri, Google Assistant, and Cortana. Ann Fam Med 2021;19(5):447-449 [FREE Full text] [doi: 10.1370/afm.2713] [Medline: 34546951]
13. Centers for Disease Control and Prevention (CDC). Symptoms of COVID-19. CDC. 2022. URL: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html [accessed 2022-10-19]
14. Apple releases new COVID-19 app and website based on CDC guidance Internet. Apple Inc. 2020. URL: https://www.apple.com/newsroom/2020/03/apple-releases-new-covid-19-app-and-website-based-on-CDC-guidance/ [accessed 2022-10-19]
15. Number of mobile devices worldwide 2020-2025. Statista. 2020. URL: https://www.statista.com/statistics/245501/multiple-mobile-device-ownership-worldwide/ [accessed 2022-10-19]
16. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-10-19]
17. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
18. Maeda E, Miyata A, Boivin J, Nomura K, Kumazawa Y, Shirasawa H, et al. Promoting fertility awareness and preconception health using a chatbot: a randomized controlled trial. Reprod Biomed Online 2020 Dec;41(6):1133-1143. [doi: 10.1016/j.rbmo.2020.09.006] [Medline: 33039321]

XSL•FO
RenderX

19. Piao M, Ryu H, Lee H, Kim J. Use of the Healthy Lifestyle Coaching Chatbot App to Promote Stair-Climbing Habits Among Office Workers: Exploratory Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 May 19;8(5):e15085 [FREE Full text] [doi: 10.2196/15085] [Medline: 32427114]
20. Health chatbots are here, heavily funded, and we'll have to get used to them. Crunchbase News. 2021. URL: https://news.crunchbase.com/news/health-chatbots-startups-vc-funding/ [accessed 2022-10-19]
21. Quiroz JC, Laranjo L, Kocaballi AB, Berkovsky S, Rezazadegan D, Coiera E. Challenges of developing a digital scribe to reduce clinical documentation burden. NPJ Digit Med 2019;2:114 [FREE Full text] [doi: 10.1038/s41746-019-0190-1] [Medline: 31799422]
22. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Systems with Applications 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
23. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
24. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
25. 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]
26. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020;3:122 [FREE Full text] [doi: 10.1038/s41746-020-00332-0] [Medline: 33015374]
27. Kocaballi AB, Ijaz K, Laranjo L, Quiroz JC, Rezazadegan D, Tong HL, et al. Envisioning an artificial intelligence documentation assistant for future primary care consultations: A co-design study with general practitioners. J Am Med Inform Assoc 2020 Nov 01;27(11):1695-1704 [FREE Full text] [doi: 10.1093/jamia/ocaa131] [Medline: 32845984]
28. Luxton DD. Ethical implications of conversational agents in global public health. Bull World Health Organ 2020 Apr 01;98(4):285-287 [FREE Full text] [doi: 10.2471/BLT.19.237636] [Medline: 32284654]
29. Clark L, Cowan B, Edwards J, Munteanu C, Murad C, Aylett M, et al. Mapping Theoretical and Methodological Perspectives for Understanding Speech Interface Interactions. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2019 Presented at: 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 1-8. [doi: 10.1145/3290607.3299009]
30. Moore RJ, Arar R, Ren GJ, Szymanski MH. Conversational UX Design. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2017 Presented at: 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. A. [doi: 10.1145/3027063.3027077]
31. Reeves S, Porcheron M, Fischer J, Candello H, McMillan D, McGregor M, et al. Voice-based Conversational UX Studies and Design. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2018 Presented at: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montréal, QC, Canada p. 1-8. [doi: 10.1145/3170427.3170619]
32. Kocaballi A, Quiroz J, Laranjo L, Rezazadegan D, Kocielnik R, Clark L, et al. Conversational Agents for Health and Wellbeing. In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-8. [doi: 10.1145/3334480.3375154]
33. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Clark L, Cowan B, Roper A, Lindsay S, Sheers O. Speech diversity and speech interfaces: considering an inclusive future through stammering. In: Proceedings of the 2nd Conference on Conversational User Interfaces. New York, NY: Association for Computing Machinery; 2020 Jul Presented at: CUI '20: 2nd Conference on Conversational User Interfaces; July 22-24, 2020; Bilbao, Spain p. 1-3. [doi: 10.1145/3405755.3406139]
35. Spillane B, Gilmartin E, Saam C, Su K, Cowan B, Lawless S, et al. Introducing ADELE: a personalized intelligent companion. In: ISIAA 2017: Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents. New York, NY: Association for Computing Machinery; 2017 Presented at: ICMI '17: International Conference on Multimodal Interaction; November 13, 2017; Glasgow, UK p. 43-44. [doi: 10.1145/3139491.3139492]
36. Gilmartin E, Saam C, Spillane B, O'Reilly M, Su K, Calvo A, et al. The ADELE Corpus of Dyadic Social Text Conversations: Dialog Act Annotation with ISO 24617-2. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Paris, France: European Language Resources Association; 2018 Presented at: Eleventh International Conference on Language Resources and Evaluation; May 7-12, 2018; Miyazaki, Japan p. e1 URL: https://aclanthology.org/L18-1633
37. Spillane B, Gilmartin E, Saam C, Cowan B, Wade V. ADELE: Care and Companionship for Independent Aging. CEUR Workshop Proceedings. 2018. URL: http://ceur-ws.org/Vol-2338/paper2.pdf [accessed 2022-10-19]

XSL•FO
RenderX

38. Lee S, Yu R, Xie J, Billah S, Carroll J. Opportunities for Human-AI Collaboration in Remote Sighted Assistance. New York, NY: Association for Computing Machinery; 2022 Presented at: IUI '22: 27th International Conference on Intelligent User Interfaces; March 22-25, 2022; Helsinki, Finland p. 63-78. [doi: 10.1145/3490099.3511113]
39. Lee S, Reddie M, Tsai C, Beck J, Rosson M, Carroll J. The Emerging Professional Practice of Remote Sighted Assistance for People with Visual Impairments. In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-12. [doi: 10.1145/3313831.3376591]
40. Xie J, Reddie M, Lee S, Billah SM, Zhou Z, Tsai C, et al. Iterative Design and Prototyping of Computer Vision Mediated Remote Sighted Assistance. ACM Trans. Comput.-Hum. Interact 2022 Aug 31;29(4):1-40. [doi: 10.1145/3501298]
41. Xie J, Yu R, Lee S, Lyu Y, Billah S, Carroll J. Helping Helpers: Supporting Volunteers in Remote Sighted Assistance with Augmented Reality Maps. In: DIS '22: Designing Interactive Systems Conference. New York, NY: Association for Computing Machinery; 2022 Jun Presented at: DIS '22: Designing Interactive Systems Conference; June 13-17, 2022; Virtual Event Australia p. 881-897. [doi: 10.1145/3532106.3533560]
42. Lee Y, Yamashita N, Huang Y. Exploring the Effects of Incorporating Human Experts to Deliver Journaling Guidance through a Chatbot. Proc. ACM Hum.-Comput. Interact 2021 Apr 13;5(CSCW1):1-27 [FREE Full text] [doi: 10.1145/3449196]
43. Lee Y, Yamashita N, Huang Y. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. Proc. ACM Hum.-Comput. Interact 2020 May 28;4(CSCW1):1-27. [doi: 10.1145/3392836]
44. Lee Y, Yamashita N, Huang Y, Fu W. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2020 Apr Presented at: CHI '20: 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-12. [doi: 10.1145/3313831.3376175]
45. Smriti D, Shin J, Mujib M, Colosimo M, Kao T, Williams J, et al. TAMICA: Tailorable Autonomous Motivational Interviewing Conversational Agent. In: Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare. New York, NY: Association for Computing Machinery; 2020 Presented at: 14th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 18-20, 2020; Virtual. [doi: 10.1145/3421937.3421963]
46. Smriti D, Kao TA, Rathod R, Shin JY, Peng W, Williams J, et al. Motivational Interviewing Conversational Agent for Parents as Proxies for Their Children in Healthy Eating: Development and User Testing. JMIR Hum Factors 2022 Oct 07;9(4):e38908. [doi: 10.2196/38908] [Medline: 36206036]
47. Kim J, Park S. Conversational Agents for Health and Wellbeing: Review and Future Agendas. In: CHI EA '20: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2019 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-8 URL: https://deepblue.lib.umich.edu/handle/2027.42/151800
48. Kim J, Muhic J, Robert LP, Park SY. Designing Chatbots with Black Americans with Chronic Conditions: Overcoming Challenges against COVID-19. In: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2022 Apr Presented at: CHI '22: 2022 CHI Conference on Human Factors in Computing Systems; April 29, 2022 to May 5, 2022; New Orleans LA p. 1-17. [doi: 10.1145/3491102.3502116]
49. Silva LM, Epstein DA. Investigating Preferred Food Description Practices in Digital Food Journaling. New York, NY: Association for Computing Machinery; 2021 Presented at: DIS '21: Designing Interactive Systems Conference 2021; June 28, 2021 to July 2, 2021; Virtual Event p. 589-605. [doi: 10.1145/3461778.3462145]
50. Silva LM, Cibrian FL, Epstein DA, Bhattacharya A, Ankrah EA, Monteiro E, et al. Adapting Multidevice Deployments During a Pandemic: Lessons Learned From Two Studies. IEEE Pervasive Comput 2022 Jan 1;21(1):48-56 [FREE Full text] [doi: 10.1109/MPRV.2021.3104262]
51. Mitchell EG, Maimone R, Cassells A, Tobin JN, Davidson P, Smaldone AM, et al. Automated vs. Human Health Coaching. Proc. ACM Hum.-Comput. Interact 2021 Apr 13;5(CSCW1):1-37. [doi: 10.1145/3449173]
52. Mitchell E, Mamykina L. From the Curtain to Kansas: Conducting Wizard-of-Oz Studies in the Wild. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2021 Presented at: 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 1-6. [doi: 10.1145/3411763.3443446]
53. Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study. J Med Internet Res 2021 Jan 06;23(1):e19928 [FREE Full text] [doi: 10.2196/19928] [Medline: 33404508]
54. Richards D, Caldwell P. Improving Health Outcomes Sooner Rather Than Later via an Interactive Website and Virtual Specialist. IEEE J Biomed Health Inform 2018 Sep;22(5):1699-1706. [doi: 10.1109/JBHI.2017.2782210] [Medline: 29990229]
55. Sezgin E, Noritz G, Lin S, Huang Y. Feasibility of a Voice-Enabled Medical Diary App (SpeakHealth) for Caregivers of Children With Special Health Care Needs and Health Care Providers: Mixed Methods Study. JMIR Form Res 2021 May 11;5(5):e25503 [FREE Full text] [doi: 10.2196/25503] [Medline: 33865233]

XSL•FO
RenderX

56. Sezgin E, Oiler B, Abbott B, Noritz G, Huang Y. "Hey Siri, Help Me Take Care of My Child": A Feasibility Study With Caregivers of Children With Special Healthcare Needs Using Voice Interaction and Automatic Speech Recognition in Remote Care Management. Front Public Health 2022;10:849322 [FREE Full text] [doi: 10.3389/fpubh.2022.849322] [Medline: 35309210]
57. Kocielnik R, Langevin R, George J, Akenaga S, Wang A, Jones D, et al. Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In: CUI '21: Proceedings of the 3rd Conference on Conversational User Interfaces. New York, NY: Association for Computing Machinery; 2021 Presented at: CUI '21: CUI 2021 - 3rd Conference on Conversational User Interfaces; July 27-29, 2021; Bilbao (online), Spain p. 1-10. [doi: 10.1145/3469595.3469599]
58. Kocielnik R, Agapie E, Argyle A, Hsieh DT, Yadav K, Taira B, et al. HarborBot: A Chatbot for Social Needs Screening. AMIA Annu Symp Proc 2019;2019:552-561 [FREE Full text] [Medline: 32308849]
59. Spillane B, Gilmartin E, Saam C, Wade V. Issues relating to trust in care agents for the elderly. New York, NY: Association for Computing Machinery; 2019 Presented at: CUI 2019: 1st International Conference on Conversational User Interfaces; August 22-23, 2019; Dublin, Ireland p. 1-3. [doi: 10.1145/3342775.3342808]
60. Yuksel BF, Collisson P, Czerwinski M. Brains or Beauty. ACM Trans. Internet Technol 2017 Mar 06;17(1):1-20 [FREE Full text] [doi: 10.1145/2998572]
61. Fan X, Oh S, McNeese M, Yen J, Cuevas H, Strater L, et al. The influence of agent reliability on trust in human-agent collaboration. New York, NY: Association for Computing Machinery; 2008 Presented at: ECCE08: European Conference on Cognitive Ergonomics; September 16-19, 2008; Funchal, Portugal p. 1-8. [doi: 10.1145/1473018.1473028]
62. Benbasat I, Wang W. Trust In and Adoption of Online Recommendation Agents. JAIS 2005 Mar;6(3):72-101. [doi: 10.17705/1jais.00065]
63. Amazon Inc.. Amazon Web Services (AWS) Comprehend Medical. Amazon Inc. 2022. URL: https://aws.amazon.com/comprehend/medical/ [accessed 2022-10-19]
64. Errattahi R, El Hannani A, Ouahmane H. Automatic Speech Recognition Errors Detection and Correction: A Review. Procedia Computer Science 2018;128:32-37 [FREE Full text] [doi: 10.1016/j.procs.2018.03.005]
65. Corbett E, Weber A. What can I say? addressing user experience challenges of a mobile voice user interface for accessibility. In: MobileHCI '16: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. New York, NY: Association for Computing Machinery; 2016 Presented at: MobileHCI '16: 18th International Conference on Human-Computer Interaction with Mobile Devices and Services; September 6-9, 2016; Florence, Italy p. 72-82. [doi: 10.1145/2935334.2935386]
66. Oviatt S, Bernard J, Levow GA. Linguistic adaptations during spoken and multimodal error resolution. Lang Speech 1998;41 (Pt 3-4):419-442. [doi: 10.1177/002383099804100409] [Medline: 10746365]
67. Branigan HP, Pickering MJ, Pearson J, McLean JF, Brown A. The role of beliefs in lexical alignment: evidence from dialogs with humans and computers. Cognition 2011 Oct;121(1):41-57. [doi: 10.1016/j.cognition.2011.05.011] [Medline: 21723549]
68. Luger E, Sellen A. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2016 May Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 5286-5297. [doi: 10.1145/2858036.2858288]
69. Fischer J, Reeves S, Porcheron M, Sikveland R. Progressivity for voice interface design. In: CUI '19: Proceedings of the 1st International Conference on Conversational User Interfaces. New York, NY: Association for Computing Machinery; 2019 Presented at: CUI 2019: 1st International Conference on Conversational User Interfaces; August 22-23, 2019; Dublin, Ireland p. 1-8. [doi: 10.1145/3342775.3342788]
70. Torres C, Franklin W, Martins L. Accessibility in Chatbots: The State of the Art in Favor of Users with Visual Impairment. In: Advances in Usability, User Experience and Assistive Technology. Berlin, Germany: Springer; 2019:623-635.
71. Brewer R, Findlater L, Kaye, Lasecki W, Munteanu C, Weber A. Accessible Voice Interfaces. In: CSCW '18: Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing. New York, NY: Association for Computing Machinery; 2018 Presented at: CSCW '18: Computer Supported Cooperative Work and Social Computing; November 3-7, 2018; Jersey City, NJ p. 44-446. [doi: 10.1145/3272973.3273006]
72. Lister K, Coughlan T, Iniesto F, Freear N, Devine P. Accessible conversational user interfaces: considerations for design. In: W4A '20: Proceedings of the 17th International Web for All Conference. New York, NY: Association for Computing Machinery; 2020 Presented at: W4A '20: 17th Web for All Conference; April 20-21, 2020; Taipei, Taiwan p. 1-11. [doi: 10.1145/3371300.3383343]
73. WCAG 2 Overview. W3C Web Accessibility Initiative (WAI). URL: https://www.w3.org/WAI/standards-guidelines/wcag/ [accessed 2022-10-19]
74. Stara V, Vera B, Bolliger D, Rossi L, Felici E, Di Rosa M, et al. Usability and Acceptance of the Embodied Conversational Agent Anne by People With Dementia and Their Caregivers: Exploratory Study in Home Environment Settings. JMIR Mhealth Uhealth 2021 Jun 25;9(6):e25891 [FREE Full text] [doi: 10.2196/25891] [Medline: 34170256]

XSL•FO
RenderX

75. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design Features of Embodied Conversational Agents in eHealth: a Literature Review. International Journal of Human-Computer Studies 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
76. Kramer LL, Ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review. J Med Internet Res 2020 Feb 06;22(2):e14058 [FREE Full text] [doi: 10.2196/14058] [Medline: 32022693]
77. Lee M, Sin J, Laban G, Kraus M, Clark L, Porcheron M, et al. Ethics of Conversational User Interfaces. In: CHI EA '22: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2022 Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; April 29, 2022 to May 5, 2022; New Orleans, LA p. 1-7. [doi: 10.1145/3491101.3503699]
78. Clark L, Doyle P, Garaialde D, Gilmartin E, Schlögl S, Edlund J, et al. The State of Speech in HCI: Trends, Themes and Challenges. Interact Comput 2019;31(4):349-371. [doi: 10.1093/iwc/iwz016]
79. Murad C, Munteanu C, Cowan B, Clark L, Porcheron M, Candello H, et al. Let's talk about CUIs: Putting conversational user interface design into practice. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2021 Presented at: 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 1-6. [doi: 10.1145/3411763.3441336]
80. Ethical guidelines for voice experiences a case for inclusivity and trustworthiness. openvoicenetwork. 2022. URL: https://openvoicenetwork.org/documents/ovn_ethical_guidlines_voice_experiences.pdf [accessed 2022-08-01]
81. Seehausen M, Kazzer P, Bajbouj M, Prehn K. Effects of empathic paraphrasing - extrinsic emotion regulation in social conflict. Front Psychol 2012;3:482 [FREE Full text] [doi: 10.3389/fpsyg.2012.00482] [Medline: 23162516]
82. Jain M, Kumar P, Kota R, Patel S. Evaluating and Informing the Design of Chatbots. In: Proceedings of the 2018 Designing Interactive Systems Conference. New York, NY: Association for Computing Machinery; 2018 Presented at: 2018 Designing Interactive Systems Conference; June 9-13, 2018; Hong Kong, China p. 895-906. [doi: 10.1145/3196709.3196735]
83. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018 Nov 24;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]

Abbreviations

ASR: automatic speech recognition
CA: conversational agent
MCI: mild cognitive impairment
NLP: natural language processing

Edited by R Kukafka; submitted 06.04.22; peer-reviewed by R Ahmad, V Stara, M Rampioni, J Kim, C Bérubé; comments to author 03.06.22; revised version received 17.09.22; accepted 26.09.22; published 15.11.22.

Please cite as:
Kocaballi AB, Sezgin E, Clark L, Carroll JM, Huang Y, Huh-Yoo J, Kim J, Kocielnik R, Lee YC, Mamykina L, Mitchell EG, Moore RJ, Murali P, Mynatt ED, Park SY, Pasta A, Richards D, Silva LM, Smriti D, Spillane B, Zhang Z, Zubatiy T
Design and Evaluation Challenges of Conversational Agents in Health Care and Well-being: Selective Review Study
J Med Internet Res 2022;24(11):e38525
URL: https://www.jmir.org/2022/11/e38525
doi:10.2196/38525
PMID:36378515

©Ahmet Baki Kocaballi, Emre Sezgin, Leigh Clark, John M Carroll, Yungui Huang, Jina Huh-Yoo, Junhan Kim, Rafal Kocielnik, Yi-Chieh Lee, Lena Mamykina, Elliot G Mitchell, Robert J Moore, Prasanth Murali, Elizabeth D Mynatt, Sun Young Park, Alessandro Pasta, Deborah Richards, Lucas M Silva, Diva Smriti, Brendan Spillane, Zhan Zhang, Tamara Zubatiy. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Determinants of e-Mental Health Use During COVID-19: Cross-sectional Canadian Study

Ellie Yu[1], MPH, PhD; Bowen Xu[1], BESc, MEng; Lydia Sequeira[1,2], MHI, PhD

[1]Canada Health Infoway, Toronto, ON, Canada

[2]Centre for Addiction and Mental Health, Campbell Family Mental Health Research Institute, Toronto, ON, Canada

Corresponding Author:
Lydia Sequeira, MHI, PhD
Canada Health Infoway
150 King St W
Ste 1300
Toronto, ON, M5H 1J9
Canada
Phone: 1 4163039181
Email: lsequeira@infoway-inforoute.ca

Abstract

Background: Access to mental health treatment across Canada remains a challenge, with many reporting unmet care needs. National and provincial e-Mental health (eMH) programs have been developed over the past decade across Canada, with many more emerging during COVID-19 in an attempt to reduce barriers related to geography, isolation, transportation, physical disability, and availability.

Objective: The aim of this study was to identify factors associated with the utilization of eMH services across Canada during the COVID-19 pandemic using Andersen and Newman's framework of health service utilization.

Methods: This study used data gathered from the 2021 Canadian Digital Health Survey, a cross-sectional, web-based survey of 12,052 Canadians aged 16 years and older with internet access. Bivariate associations between the use of eMH services and health service utilization factors (predisposing, enabling, illness level) of survey respondents were assessed using χ^2 tests for categorical variables and *t* tests for the continuous variable. Logistic regression was used to predict the probability of using eMH services given the respondents' predisposing, enabling, and illness-level factors while adjusting for respondents' age and gender.

Results: The proportion of eMH service users among survey respondents was small (883/12,052, 7.33%). Results from the logistic regression suggest that users of eMH services were likely to be those with regular family physician access (odds ratio [OR] 1.57, *P*=.02), living in nonrural communities (OR 1.08, *P*<.001), having undergraduate (OR 1.40, *P*=.001) or postgraduate (OR 1.48, *P*=.003) education, and being eHealth literate (OR 1.05, *P*<.001). Those with lower eMH usage were less likely to speak English at home (OR 0.06, *P*<.001).

Conclusions: Our study provides empirical evidence on the impact of individual health utilization factors on the use of eMH among Canadians during the COVID-19 pandemic. Given the opportunities and promise of eMH services in increasing access to care, future digital interventions should both tailor themselves toward users of these services and consider awareness campaigns to reach nonusers. Future research should also focus on understanding the reasons behind the use and nonuse of eMH services.

(J Med Internet Res 2022;24(11):e39662) doi:10.2196/39662

KEYWORDS

digital health; mental health; e-Mental health; user profile; determinants; health service; use; utilization; COVID-19; pandemic; Canada; users; factors

XSL•FO
RenderX

Introduction

Challenges Within Mental Health Treatment in Canada

Mental health is defined as "a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her own community" [1]. Mental health problems and illnesses refer to the "range of behaviors, thoughts and emotions that can result in some level of distress or impairment in areas such as school, work, social and family interactions and the ability to live independently" [2]. These problems can be present in many forms, including mood disorders (eg, depression, bipolar disorder), anxiety disorders, schizophrenia, personality disorders, or eating disorders [3]. The prevalence of mental illness across Canada is high, with 1 in 5 Canadians experiencing a mental health or addiction problem in any given year [4]. The COVID-19 pandemic further impacted the mental health of Canadians, with an increase seen in the prevalence of certain mental health disorders such as depression and anxiety [5], as well as psychological problems such as insomnia, as seen in multiple nations across Europe, North America, and Asia [6].

In terms of treatment, the continuum of mental health care delivery across Canada includes community-based care such as primary care clinics, social services, mental health and addictions service clinics, and residential services, along with acute care such as the emergency department, inpatient, or psychiatric services, with publicly funded treatment focusing on those deemed to be "medically necessary" [7]. Low-intensity care for mild-to-moderate mental illnesses, including private-practice psychotherapists, still requires Canadians to pay out of pocket or access services through private insurance plans [8].

As such, access to mental health treatment remains a challenge, with 2.3 million Canadians reporting that their mental health care needs were either partially or entirely unmet, with the most frequently reported barriers being awareness and navigation of services, the time required for accessing services, or not being able to pay for services [9]. Other barriers to accessing mental health treatment across Canada include long wait times, shortage of mental health professionals, lack of mental health service integration, cultural and language barriers, concerns about stigma, and inequities due to geography or demographics [10].

e-Mental Health and the Canadian Landscape

e-Mental health (eMH), which refers to the use of internet and related technologies to deliver timely, effective mental health services [11], has emerged over the past decade across Canada as a method that increases the accessibility of services, including broadening the reach of services for individuals in rural and remote locations, while also allowing individuals in urban and semiurban locations to overcome barriers related to transportation, physical disability, or availability [12]. Shortcomings of eMH services also exist, as there is a "digital divide," defined as the separation between those who readily have access to a computer or device (eg, smartphone) and the internet, and those who do not [13].

Within Canada, there exist national eMH programs (eg, Kids Help Phone, Wellness Together Canada) and provincially focused programs (eg, Bridge the gApp, Tel-jeunes, Togetherall, Bounce Back). These programs operate as part of mental health service delivery through hospital- or community-based providers, or as a component of a government's mental health strategy [13]. Moreover, beyond these programs, there has been a proliferation of online self-management resources (including websites and mobile apps) developed by academic or private organizations, with many of these eMH interventions being used to support population mental health during the COVID-19 pandemic [14]. A recent systematic review also demonstrated strong support for the effectiveness of digital cognitive behavior therapy to treat insomnia [15].

COVID-19 has also accelerated the pace of adoption of digital technologies for publicly funded mental health service delivery through the implementation of virtual billing codes and an increase in access to virtual assessments [16]. Toolkits to accelerate the awareness, uptake, and implementation of services across practitioners have emerged [17,18], aiding with the awareness and adoption of eMH services.

Despite this vast variety of services available for both population-level mental health management and focused-condition or diagnosis-specific care, the utility of eMH services across Canada is not well understood. Understanding help-seeking behaviors and service utilization is critical to assessing if and how the current eMH landscape addresses current mental health needs. Andersen and Newman's [19,20] framework states that an individual's health service utilization is dependent on the predisposition of the individual to use services, the individual's ability to secure services, and the individual's illness level. Predisposing factors include demographics (eg, age, gender), social structure (eg, education, ethnicity), and beliefs, whereby some individuals have a propensity to use services more than others. Enabling factors include family factors (eg, income, insurance) and community factors (eg, urban-rural), which are conditions that make health service resources more available to an individual. Finally, illness levels include an individual's symptoms and diagnoses, which usually represent the most immediate causes of health service use [19,20].

Study Aims

The objective of this study was to identify predisposing, enabling, and illness-level factors associated with the utilization of eMH services across Canada during the COVID-19 pandemic, using information collected through the 2021 Canadian Digital Health Survey.

Methods

Recruitment and Data Collection

The study used data gathered from the 2021 Canadian Digital Health Survey, a cross-sectional, web-based survey of 12,052 Canadians aged 16 years and older with internet access. The survey was commissioned by Canada Health Infoway (Infoway)

XSL•FO
RenderX

and conducted by Léger. Data collection took place from July 14 to August 6, 2021.

The survey collected information on the use of digital health services, use of health services, and socioeconomic and demographic factors of selected Canadians. Survey participants were selected from the Léger Opinion panel—the largest Canadian panel with approximately 500,000 representative panelists from all regions of Canada—using random digital dial sampling.

Respondents from hard-to-reach target groups (eg, cancer patients) were added to the panel through targeted recruitment campaigns. Administration of the survey was conducted by Léger; however, testing of the online survey was conducted by both Infoway and Léger staff. Small monetary incentives were offered to survey participants as part of the data collection process, administered through Léger. Based on respondents' default language of choice, the survey was presented in either English or French.

Ethics Considerations

Informed consent was collected at the beginning of the survey and no personal identifiers were collected as part of the survey. Data were coded by Léger and then transferred to Infoway for analysis. Survey data and interactive visualizations are publicly available [21]. Due to the use of publicly available data, the requirement for ethics approval was waived.

Independent and Dependent Variables

Overview

Table 1 provides an overview of the independent and dependent variables used for analysis, categorized according to Andersen and Newman's [19,20] framework for individual determinants of health service utilization.

Table 1. Dependent and independent variables categorized according to Andersen and Newman's [19,20] framework for individual determinants of health service utilization.

Framework category	Variables
Independent variables	
Predisposing factors	
Demographics	Age, gender
Social structure	Ethnicity, education level, immigrant status, language, employment
Beliefs	eHealth literacy
Enabling factors	
Family	Household income, health insurance coverage
Community	Community size, access to family doctor
Illness levels	Self-rated mental health status, self-rated health status, diagnosed mental health condition, caregiving status
Dependent variable	Use of e-Mental health services

Independent Variables

Predisposing Factors

Predisposing factors indicate the propensity for an individual to use services more than other individuals based on their characteristics that exist prior to onset of any specific illness [19,20]. As per Andersen and Newman [19,20], demographics and social structures are characteristics that might predict use of health services, and are closely linked to the third component of predisposing conditions: attitudes or beliefs about care and illness.

Demographics included age and gender. Age was calculated as the difference between a respondent's year of birth and the survey date. Gender was measured with the question, "How would you describe your gender identity?" with responses collapsed into female, male, and other.

Social structure included ethnicity, education level, immigrant status, language, and employment. Ethnicity was collected based on the question, "Which ancestry category best describes you?" and responses were dichotomized into white and nonwhite. Education level was assessed with the self-reported highest level of education obtained, including qualifications obtained outside of Canada. The categories were collapsed into high school or equivalent, college or trades, undergraduate degree, postgraduate degree, and other/none/prefer not to say. Immigrant status was based on the question, "Are you a Canadian citizen?" Language at home was based on the question "Which language do you speak on a regular basis at home?" Employment status was based on the question "What is your current employment status?" and responses were collapsed into working, unemployed, retired, disabled, student, and other/prefer not to say. Respondents who were employed either full or part time were classified as employed.

eHealth literacy was measured using the eHealth Literacy Scale (eHEALS), an 8-item self-assessment tool designed to measure respondents' knowledge, comfort, and perceived skills at finding, evaluating, and applying electronic health information to health problems [13]. Originally developed to assess eHealth literacy levels among youth and youth workers by Skinner and Norman [13], the scale has since been adapted to a variety of settings, population groups, and multiple languages [14]. Each question measures an aspect of perceived eHealth literacy and is scored on a Likert scale ranging from 1 to 5. Scores are summed to derive an overall eHealth literacy score that ranges

XSL•FO
RenderX

from 8 to 40 for each respondent. A higher eHEALS score represents higher self-perceived eHealth literacy.

Enabling Factors

Despite individuals being predisposed to using health services, it is necessary to have the means available for them to access these services. Enabling conditions are those that make health service resources available to individuals [19,20], which were broken down into family and community variables in this study.

With regard to family, household income was based on the total self-reported household income before tax in the past year. The categories were collapsed into $24,999 or less, $25,000-$80,000, $80,000 or more, and prefer not to answer (in CAD, in which CAD $1=US $0.78 at the time of the survey). Insurance coverage was based on the question "Which of the following best describes the type of health insurance coverage you currently have?" and the categories were collapsed into public coverage only, private coverage, no coverage, and don't know/prefer not to say. Private coverage includes insurance plans paid for by the respondent, a family member, an employer, or an association.

With regard to community, access to a family doctor was assessed through the question "Do you have a family doctor or regular place of care, such as a health center or a family medical/medicine group?" The responses were dichotomized into yes and no/don't know.

Community size was based on the question "How would you describe the community you live in?" and responses were collapsed into rural, small to large population centers, and urban center. The population size for a rural community was defined as less than 1000 people. The population size for a small, medium, and large population center was defined as 1000-29,999 people, 20,000-99,999 people, and 100,000-999,999 people, respectively. Urban centers were defined as 1 million people and over.

Illness Levels

Illness-level factors represent the most immediate cause of utilization of health services and can include perspectives of illness as well as clinical diagnoses [19,20]. Self-reported mental health (SRMH) status and self-reported health (SRH) status were measured by asking respondents, "In general, how would you rate your overall physical/mental health?" The responses were collapsed into fair/poor, good, very good/excellent, and prefer not to say. Self-reported diagnosed mental health condition was assessed with the question "Do you have emotional, psychological, or mental health conditions (eg, anxiety, depression, bipolar disorder, substance abuse, anorexia, etc) diagnosed by a health professional?" The responses were dichotomized into yes and no.

Caregiver status was based on the question, "Do you have primary or joint responsibility for providing care and/or assistance to someone?" Respondents were given the prompt that assistance refers to voluntary assistance, excluding employment or work done for payment.

Outcome (Dependent) Variable

The dependent variable was the use of eMH services, which was measured with the question, "Did you access websites, mobile applications (apps) or interactive online tools and services to help or support you with mental health issues you may be dealing with, such as depression, anxiety, or substance abuse in the last 12 months?" Responses were dichotomized into yes and no/don't know.

Statistical Analysis

Participant Profile

SPSS version 24 (IBM SPSS Statistics) was used for descriptive analyses. SAS version 9.4 (SAS Institute Inc) was used for logistic regression analyses. All estimates reported are based on weighted data that reflect the age, gender, and geographic distribution of Canadians aged 16 years and above in the 2016 census. Descriptive statistics were calculated for respondents who had used eMH services during the past 12 months and those who had not. Cross-tabulations were used to estimate the prevalence of eMH service utilization within our sample as well as characteristics associated with users and nonusers of eMH services.

Bivariate Associations for Use and Nonuse of eMH Services

Bivariate associations between the use of eMH services and predisposing, enabling, and illness-level factors of survey respondents were assessed using χ^2 tests for categorical variables and *t* tests for the continuous variable.

Unadjusted Logistic Regression Model to Assess Determinants of Use of eMH Services

Two independent adjusted logistic regression models were performed. The first model was adjusted for predisposing, enabling, and illness-level factors. Predisposing factors included in the model were ethnicity, education level, immigrant status, home language, employment status, and eHealth literacy. Enabling factors included in the model were household income, health insurance coverage, access to family doctors, and community size. Illness-level factors included in the model were SRMH, SRH, diagnosed mental health status, and caregiver status.

Adjusted Logistic Regression Model to Assess Determinants of Use of eMH Services

The second model was also adjusted for predisposing, enabling, and illness-level factors along with demographic factors, including age and gender. The adjusted multivariable logistic regressions were performed to assess associations between predisposing, enabling, and illness-level factors and use of eMH services controlling for age and gender.

We tested for multicollinearity by assessing the bivariate correlation between two predictor variables. No interactions were found between access to a family physician, SRH, SRMH, diagnosed mental health condition, household income, education, immigrant status, language at home, employment status, insurance coverage, age, gender, ethnicity, community size, and caregiver status.

XSL•FO
RenderX

Results

Participant Profile

A total of 12,052 Canadians aged 16 years or older were surveyed. The proportion of respondents who self-reported using an eMH service in the past 12 months (ie, users) was 883 out of 12,052 (7.33%) and the proportion of respondents who did not use any eMH service in the past 12 months (ie, nonuser) was 11,169 out of 12,052 (92.67%). Table 2 compares the predisposing, enabling, and illness-level factors of users and nonusers of eMH services.

The average age of eMH users was 40.4 (SD 15.97) years and the average age of nonusers was 47.61 (SD 17.72) years. The proportion of women within our sample was higher among users of eMH services as compared to nonusers of eMH services. The proportion of those who identified in the gender category "other" varied significantly between users and nonusers, with a higher proportion falling within the eMH-user group (Table 2). In our sample, the prevalence of white respondents among users of eMH services was lower than that of nonusers. A higher percentage of users of eMH services identified as immigrants and noncitizens, as employed full or part time, disabled, and a student when compared to nonusers. The difference in the distribution of education level among users and nonusers was statistically significant, with a higher percentage of eMH users having obtained at least an undergraduate degree compared to nonusers. The percentage of eMH users living in rural communities was lower than that of nonusers, with a higher percentage of eMH users living in urban centers. A higher percentage of eMH users self-reported to be caregivers compared to nonusers. A higher percentage of eMH service users reported their annual household income to be CAD $80,000 or more compared to nonusers. A higher percentage of users of eMH services identified as having private insurance coverage when compared to nonusers.

XSL•FO
RenderX

Table 2. Characteristics of users and nonusers of e-Mental health (eMH) services, and associations with demographic, health-related, and socioeconomic factors for Canadians aged 16 years or older (2021 Canadian Digital Health Survey; N=12,052).

Predictor variables	Used eMH services in the past 12 months (n=883 unweighted, n=897 weighted)	Did not use eMH services in the past 12 months (n=11,169 unweighted, n=11,155 weighted)	χ^2 or *F (df)*	*P* value
Predisposing factors				
Age, mean (SD)	40.40 (15.97)	47.61 (17.72)	35.06 (1)	<.001
Gender, n (%)[a]			8.78 (2)	.01
Man (ref)	412 (45.9)	5384 (48.3)		
Woman	467 (52.0)	5658 (50.7)		
Other[b]	18 (2.0)	113 (1.0)		
Ethnicity, n (%)			24.24 (1)	<.001
White	608 (67.8)	8390 (75.2)		
Nonwhite[b,c] (ref)	289 (32.30)	2765 (24.8)		
Education level, n (%)			43.22 (3)	<.001
High school or equivalent (ref)	181 (20.2)	2537 (22.7)		
College or trades	197 (21.9)	3231 (29.0)		
Undergraduate degree[b]	365 (40.7)	3668 (32.9)		
Postgraduate degree[b]	136 (15.2)	1342 (12.0)		
Other/none of above/prefer not to answer	18 (2.5)	377 (3.4)		
Immigrant status, n (%)			2.21 (2)	.33
Born in Canada (ref)	713 (79.5)	9016 (80.8)		
Immigrant[d]	137 (15.2)	1667 (14.9)		
Not a citizen	47 (5.3)	472 (4.2)		
Language at home, n (%)			55.21 (2)	<.001
English (ref)	765 (85.3)	8267 (74.1)		
French[b]	107 (11.9)	2,373 (21.3)		
Other[b]	25 (2.8)	515 (4.6)		
Employment status, n (%)			91.06 (5)	<.001
Employed (full or part time) (ref)	593 (66.1)	6255 (56.1)		
Unemployed	74 (8.3)	888 (8.0)		
Retired[b]	92 (10.3)	2644 (23.7)		
Disabled	30 (3.4)	291 (2.6)		
Student	92 (10.2)	839 (7.5)		
Other/prefer not to say	16 (1.7)	238 (2.1)		
Enabling factors, n (%)				
Household income[e]			5.73 (3)	.13
24,999 or less (ref)	79 (8.8)	1051 (9.4)		
25,000-80,000	322 (35.9)	4253 (38.1)		
80,000 or more	413 (46.0)	4690 (42.1)		
Prefer not to answer	83 (9.3)	1161 (10.4)		
Insurance coverage			10.67 (3)	.01
Public only (ref)	285 (31.8)	3844 (34.5)		

XSL•FO
RenderX

Predictor variables	Used eMH services in the past 12 months (n=883 unweighted, n=897 weighted)	Did not use eMH services in the past 12 months (n=11,169 unweighted, n=11,155 weighted)	χ^2 or *F (df)*	*P* value
Private	488 (54.4)	5,465 (49.0)		
No coverage	64 (7.1)	997 (8.9)		
I don't know/prefer not to say	60 (6.7)	849 (7.6)		
Community size			6.03 (2)	.05
Rural (ref)	58 (6.4)	909 (8.2)		
Small-large population centers	585 (65.2)	7420 (66.5)		
Urban centers	254 (28.3)	2826 (25.3)		
Access to a family doctor			17.60 (1)	<.001
Yes (ref)	825 (91.9)	9722 (87.2)		
No/don't know[b]	72 (8.1)	1433 (12.8)		
Illness levels, n (%)				
SRMH[f]			154.21 (3)	<.001
Excellent/very good (ref)	293 (32.7)	5164 (46.3)		
Good	237 (26.4)	3442 (30.9)		
Poor/fair[b]	364 (40.6)	2504 (22.4)		
Prefer not to say	3 (0.3)	45 (0.4)		
SRH[g]			6.73 (2)	.08
Excellent/very good (ref)	374 (41.7)	4811 (43.1)		
Good	324 (36.1)	4204 (37.7)		
Poor/fair	198 (22.1)	2103 (18.9)		
Prefer not to say	1 (0.1)	37 (0.3)		
Diagnosed mental health condition			241.19 (1)	<.001
Yes (ref)	312 (34.7)	1655 (14.8)		
No[b]	585 (65.3)	9500 (85.2)		
Caregiver status			114.04 (1)	<.001
Yes (ref)	331 (36.9)	2388 (21.4)		
No[b]	566 (63.1)	8766 (78.6)		

[a]Percentages are weighted and have been rounded, thus may not total 100.

[b]Significantly different from estimate for reference category (*P*<.05), Bonferroni-adjusted pairwise *Z*-test.

[c]Including respondents who selected "prefer not to answer" and "other."

[d]Referring to the proportion of respondents who are immigrant and granted citizenship of Canada under the *Citizenship Act.*

[e]In Canadian dollars (CAD $1=US $0.78).

[f]SRMH: self-rated mental health status.

[g]SRH: self-rated health status.

Bivariate Associations for Use and Nonuse of eMH Services

Table 2 outlines the bivariate associations between use of eMH service and predisposing, enabling, and illness-level factors. The average age of eMH service users was significantly younger than that of nonusers of eMH services. The association between gender and use of eMH services was also statistically significant. Education, employment, language, and insurance coverage were all significantly associated with the use of eMH services. Ethnicity was also significantly associated with the use of eMH services. The prevalence of nonwhite individuals significantly differed from that of white individuals for both users and nonusers, although the proportional difference between the prevalence of white and nonwhite respondents was more pronounced among nonusers.

Similarly, the prevalence of caregivers significantly differed from that of noncaregivers for both users and nonusers, and the

XSL•FO
RenderX

proportional difference was more pronounced among nonusers. Access to a family physician was significantly associated with the use of eMH services, with a higher proportion of caregivers among those reporting using eMH services when compared to nonusers. SRMH status was also significantly associated with the use of eMH services, with significant differences between respondents who reported a fair/poor mental health status and those who reported an excellent/very good mental health status. The prevalence of a diagnosed mental health condition was significantly associated with use of eMH services, with a greater proportion of users having a self-reported mental health condition.

For education, those with an undergraduate or postgraduate degree significantly differed from those with only high school or equivalent diplomas, for both users and nonusers. For language, the prevalence of English speakers was significantly different from that of French or other-language speakers, for both users and nonusers. The only group that differed significantly from the reference group of full-time or part-time employees was retired individuals for both users and nonusers. Post hoc tests did not reveal pairwise differences for insurance coverage.

Unadjusted Logistic Regression Model to Assess Determinants of Use of eMH Services

Table 3 shows the estimates and odds ratios from the unadjusted regression model of predisposing, enabling, and illness-level factors and the associations with use of eMH services.

For predisposing factors, those with an undergraduate or postgraduate degree had higher odds of using eMH services when compared to those with a high school diploma or equivalent. Survey respondents who did not speak English at home had lower odds of using eMH services. Compared to respondents who were employed, those who were unemployed, retired, or disabled had lower odds of using eMH services. Finally, those with a higher eHEALS score had higher odds of using eMH services.

The only enabling factor significantly associated with eMH use was family physician access. Having a regular family doctor was positively associated with use of eMH services.

Except for SRH, all illness-level factors were significantly associated with the use of eMH services. For SRMH status, those with fair or poor SRMH were more likely to use eMH services. Having a diagnosed mental health condition and being a caregiver were both positively associated with greater odds of using eMH services.

XSL•FO
RenderX

Table 3. Logistic regression adjusted for predisposing, enabling, and illness-level factors and their association with e-Mental health service use among Canadians aged 16 years or older (2021 Canadian Digital Health Survey).

Variables	Odds ratio	95% CI	*P* value
Predisposing factors (vs reference group)			
Ethnicity (nonwhite vs white)	0.80	0.67-0.95	.01
Education (college, trades vs high school)	0.91	0.73-1.14	.43
Education (undergraduate degree vs high school)	1.34	1.10-1.65	.005
Education (postgraduate degree vs high school)	1.35	1.05-1.75	.02
Education (prefer not to say, none of the above, other vs high school)	0.86	0.52-1.43	.56
Immigrant status (immigrant, not a citizen vs born in Canada)	1.02	0.84-1.24	.85
Language (other than English vs English)	0.58	0.48-0.70	<.001
Employment (unemployed[a] vs employed)	0.66	0.55-0.79	<.001
Employment (student vs employed)	1.19	0.92-1.54	.18
Employment (prefer not to say, other vs employed)	0.66	0.38-1.14	.14
eHEALS[b]	1.05	1.04-1.06	<.001
Enabling factors (vs reference group)			
Income[c] (25,000-79,000 vs <25,000)	0.92	0.70-1.21	.56
Income (80,000 or more vs <25,000)	0.94	0.71-1.25	.67
Income (prefer not to say vs <25,000)	0.99	0.71-1.39	.96
Insurance (public only vs has private insurance)	0.98	0.83-1.15	.79
Insurance (no coverage, don’t know, prefer not to say vs has private insurance)	0.82	0.66-1.02	.08
Has a family physician (yes vs no, don’t know)	1.47	1.14-1.90	<.001
Community size (small to large population centers vs rural)	1.09	0.81-1.45	.58
Community size (urban center vs rural)	1.04	0.76-1.42	.82
Illness-level factors (vs reference group)			
SRMH[d] (fair, poor vs excellent, very good, good, prefer not to say)	1.79	1.51-2.13	<.001
SRH[e] (fair, poor vs excellent, very good, good, prefer not to say)	0.90	0.74-1.08	.26
Diagnosed mental health condition (yes vs no)	2.35	1.98-2.79	<.001
Caregiver status (yes vs no)	1.87	1.61-2.17	<.001

[a]Unemployed includes disabled and retired respondents.

[b]eHEALS: eHealth Literacy Scale.

[c]Income given in Canadian dollars (CAD $1=US $0.78).

[d]SRMH: self-rated mental health.

[e]SRH: self-rated health.

Adjusted Logistic Regression Model to Assess Determinants of Use of eMH Services

Table 4 shows the estimates of the adjusted logistic regression model of predisposing, enabling, and illness-level factors, and the associations with use of eMH services.

Education, language, and eHealth literacy were predisposing factors that significantly predicted use of eMH services. Both those with undergraduate and postgraduate education had higher odds of using eMH services compared to those with high school education. Respondents who did not speak English at home had lower odds of using eMH services compared to those who did. Again, those with a higher eHEALS score had higher odds of using eMH services.

Access to a family physician, community size, and income were enabling factors significantly associated with use of eMH services. Those making CAD $25,000-$79,000 were less likely to use eMH services than those making less than CAD $25,000; however, for those making CAD $80,000 or more or for those who did not report their income, this relationship was not significant. Those with a regular family physician and those living in small, medium, or large population centers were more likely to use eMH services than those without a family physician and those living in rural areas, respectively.

XSL•FO
RenderX

For illness-level factors, in the model adjusted for age and gender, SRMH was not significantly associated with use of eMH. Those who rated their SRH as poor or fair had lower odds of using eMH services (Table 3).

Table 4. Logistic regression adjusted for predisposing, enabling, illness-level, and demographic factors and their association with e-Mental health service use among Canadians aged 16 years or older (2021 Canadian Digital Health Survey).

Variables	Odds ratio	95% CI	*P* value
Predisposing factors (vs reference group)			
Ethnicity (nonwhite vs white)	0.89	0.75-1.07	.22
Education (college, trades vs high school)	1.00	0.80-1.25	.99
Education (undergraduate degree vs high school)	1.40	1.14-1.72	.001
Education (postgraduate degree vs high school)	1.48	1.14-1.92	.003
Education (prefer not to say, none of the above, other vs high school)	0.84	0.50-1.40	.50
Immigrant status (immigrant, not a citizen vs born in Canada)	1.07	0.88-1.30	.49
Language (other than English vs English)	0.56	0.46-0.69	<.001
Employment (unemployed[a] vs employed)	0.87	0.71-1.07	.20
Employment (student vs employed)	0.95	0.73-1.24	.70
Employment (prefer not to say, other vs employed)	0.74	0.43-1.27	.27
eHEALS[b]	1.05	1.03-1.06	<.001
Enabling factors (vs reference group)			
Income[c] (25,000-79,000 vs <25,000)	0.97	0.73-1.27	<.001
Income (80,000 or more vs <25,000)	1.00	0.75-1.33	.80
Income (prefer not to say vs <25,000)	1.06	0.76-1.49	>.99
Insurance (public only vs has private insurance)	1.02	0.86-1.20	.73
Insurance (no coverage, don’t know, prefer not to say vs has private insurance)	0.77	0.62-0.96	.85
Has a family physician (yes vs no, don’t know)	1.57	1.22-2.03	.02
Community size (small to large population centers vs rural)	1.08	0.81-1.44	<.001
Community size (urban center vs rural)	1.05	0.77-1.43	.62
Illness-level factors (vs reference group)			
SRMH[d] (fair, poor vs excellent, very good, good, prefer not to say)	1.68	1.41-2.00	.77
SRH[e] (fair, poor vs excellent, very good, good, prefer not to say)	0.95	0.78-1.15	<.001
Diagnosed mental health condition (yes vs no)	2.20	1.85-2.62	.58
Caregiver status (yes vs no)	1.85	1.59-2.15	<.001
Demographic factors			
Gender	1.19	0.70-2.02	.52
Age	0.98	0.98-0.99	<.001

[a]Unemployed includes disabled and retired respondents.

[b]eHEALS: eHealth Literacy Scale.

[c]Income given in Canadian dollars (CAD $1=US $0.78).

[d]SRHM: self-rated mental health.

[e]SRH: self-rated health.

Discussion

Principal Results and Comparison With Prior Work

Our results demonstrate that the adoption of eMH across Canada is limited, with only 883 out of 12,052 (7.33%) survey respondents reporting use of these services within the last 12 months. Lifetime usage of eMH was only slightly higher within this sample, at 1217 out of 12,052 (10.10%). This low usage of eMH services has been observed in previous research from Germany [22], which contrasts with data from the United States in which up to 55% of respondents reported using specific digital mental health tools and technologies to manage mental health

XSL•FO
RenderX

during the COVID-19 pandemic. The difference between eMH use across Canada and the United States could in part be due to the interpretation of the term "e-Mental health," as some individuals exclude the use of telephones within this definition and others include it. Additionally, there have been several digital health policies implemented within the United States (eg, HITECH Act, 21st Century Cures Act) that have improved the adoption of digital health across the country [23], with digital health (including mental health) across Canada inching behind. Funding models for health care also differ between the two countries [24], with many more opportunities for growth of digital health companies in the United States with private-insurance payers when compared to public-insurance payers across Canadian provinces.

Benefits of our study sample include its representativeness of the Canadian population. Moreover, within our study, 35% of individuals who reported being in very good mental health sought eMH services within the prior 12 months, demonstrating that the survey respondents were also using these tools for maintaining mental health rather than solely for treating mental illness. Previous research has identified older adults' motivation to use technology to improve mood through mechanisms of distraction, normalization, and facilitated expression of mental states [25].

Profile of eMH Service Users

The predisposing demographic factors that had a significant association with use of eMH services included age, where younger participants were more likely to be users of eMH services. This result is consistent with population-level trends on the use of digital health technology seen in the United States, where younger people were found to be more likely to use these forms of health care [26]. A systematic review of mental health help-seeking behaviors among young people found that facilitators for online options included greater anonymity and confidentiality (which could lower concerns related to stigma), timely access through the ability to access care 24 hours per day, and empowerment through improved information access [27].

With regard to variables related to an individual's social structure, ethnicity and immigrant status were associated with eMH use, with the prevalence of white respondents among users of eMH services lower than that of nonusers. This higher proportion of individuals within the eMH user group could speak to the increase in access to providers who share cultural and linguistic backgrounds with immigrant participants or with participants of a particular ethnicity [28].

Finally, individuals reporting poor SRMH were 1.7 times more likely to use eMH services, and those with a diagnosis of a mental health condition were 2.5 times more likely to use these services, demonstrating Andersen and Newman's [19,20] individual determinant of illness level, which represents the most immediate cause of health service use. However, some of these significant associations were not translated to our logistic regression analyses, a finding that requires further investigation.

Determinants of Use of eMH Services

Through our logistic regression analyses, three of the predisposing factors that predicted eMH use included education, language, and eHealth literacy. Similar trends have been identified within research studying help-seeking behaviors of ethnic minorities with a mental health diagnosis [29], due to both attitudinal barriers and structural barriers such as the cultural inappropriateness of interventions [29,30]. Another predisposing factor, eHealth literacy, was also a significant predictor of the use of eMH services. eHealth literacy has been defined as a metaliteracy comprised of traditional literacy and numeracy, health literacy, computer literacy, science literacy, media literacy, and information literacy [31]. Similar trends have borne out in previous research on the use of Web 2.0 websites such as Facebook and Twitter for searching for and sharing health information [32]. Research has also identified the impact of eHealth literacy on differing levels of trust with digital channels and sources, with higher eHealth literacy in certain populations leading to a high perceived trust in online government organizations [33].

With regard to enabling features, those with access to a family physician were more likely to use eMH services, which could potentially be due to an increase in awareness of these services. Past research has shown that the actions of health care professionals can influence patient activation and engagement, with engagement predicting digital information–seeking behavior [34]. Primary care is a majority of Canadians' first point of contact with the health system for mental health and addiction challenges, and also where the majority of mental health care is delivered [35,36]. With efforts to educate health professionals (including primary care) about the use and implementation of eMH resources [18], such education and awareness are likely to be passed onto patients.

Factors such as having a higher income and living within nonrural areas have proven to be enablers to accessing eMH services, furthering the evidence of the existence of a digital divide. Challenges with improving rural and remote communities' access to broadband internet persist, and until regulatory changes happen across Canada, this gap will continue to hinder access among rural communities [37]. Beyond broadband access, the need for devices (eg, computers, tablets, smartphones) is necessary for engagement in eMH services, along with appropriate digital literacy [38].

Finally, caregiver status was a predictor of eMH use within our sample. The COVID-19 pandemic increased caregiving intensity and caregiving burden [39], which had downstream impacts on caregivers' mental health [40].

It is evident that while there has been a proliferation of eMH services by governmental offerings and beyond, expanded by the COVID-19 pandemic, the range of eMH services is vast and service utilization remains low. In addition, empirical evidence on the effectiveness of eMH service remains limited. With rapid expansions of eMH service options and technologies, it is imperative that there will be a coordinated national approach to direct eMH policies, research, and best-practice guidelines moving forward. Nonetheless, results from our survey were able to identify a persona of eMH service users during the

XSL•FO
RenderX

COVID-19 pandemic, allowing for key population-level insights. Expanding eMH care across Canada will require raising awareness about available technologies and integrating "proven" technologies within the model of care [41].

Our findings suggest that SRMH was significantly associated with use of eMH, but when controlled for age and gender there was no significance. This suggests that age and gender have a stronger effect on the use of eMH services when compared to the effect of SRMH alone. Previous literature has shown that individuals who identified as male showed significantly lower recognition of symptoms associated with mental illness [42]. Research has also demonstrated that younger individuals tended to have more positive mental health perceptions, where they are able to identify and acknowledge their mental health better than older individuals [43].

Limitations

A limitation of our study is the imbalance in responses among participants for our dependent variable (ie, use of eMH services), with eMH users making up a very small percentage (7.33%) of the sample. Moreover, at a constant threshold of $P<.05$, large sample sizes are more likely to find a significant relationship if one exists [44]. This is why we reported the odds ratios for both unadjusted and adjusted models. Additionally, the measure that identified use of eMH services asked about individuals' access of "websites, mobile applications, or interactive online tools," without specifying telephone as a modality (follow-up questions listed telephone services as options). This could have caused a potential decrease in the number of individuals who identified as using these services within the last 12 months. In addition, there were missing data for approximately 10% of the questions on income levels, where individuals had chosen the "prefer not to answer" option. This response rate is comparable to the literature, whereby questions on income are often unanswered by a small percentage of respondents [45].

We also recognize that the Canadian Digital Health Survey is an online survey and therefore may limit participation of populations with limited access to technological equipment and internet, as well as certain ethnic/culture groups overrepresented in these populations. In addition, the survey did not collect information on the duration or completion of the eMH encounter, and therefore the visit could have varied in duration, quality, and completeness. As the survey relies on self-reported data, data collected may be impacted by recall error, although past research has shown that bias and variance of recall error of health care usage were minimized for the 12-month recall period [30].

Conclusion

Our study provides an overview of the individual determinants of eMH use across Canada. The proportion of eMH service users was small, and users were likely to be those with regular family physician access, living in nonrural communities, more educated, eHealth-literate, and English-speaking. Given the opportunities and promise of eMH services in increasing access to care, future digital interventions should both tailor themselves toward users of these services, while also considering awareness campaigns to reach nonusers. Understanding the reasons behind use and nonuse is also important.

Acknowledgments

The authors would like to thank Katie Bryski for her thorough proofreading of the article, and Canada Health Infoway for their in-kind contributions to this work.

Authors' Contributions

EY, LS, and BX conceptualized the study. EY and LS drafted the manuscript. EY and BX performed the data analysis. All authors edited and provided feedback on the manuscript.

Conflicts of Interest

None declared.

References

1. Health and well-being. World Health Organization. URL: https://www.who.int/data/gho/data/major-themes/health-and-well-being [accessed 2022-10-22]
2. Making the case for investing in mental health in Canada. Mental Health Commission of Canada. 2013 Sep. URL: https://mentalhealthcommission.ca/resource/making-the-case-for-investing-in-mental-health-in-canada/ [accessed 2022-10-05]
3. What are the types of mental illness? Government of Canada. 2022. URL: https://www.canada.ca/en/public-health/services/about-mental-illness.html#a2 [accessed 2022-10-05]
4. Smetanin P, Briante C, Stiff D, Ahmad S, Khan M. The life and economic impact of major mental illnesses in Canada: 2011 to 2041. RiskAnalytica. Mental Health Commission of Canada. 2011. URL: https://www.mentalhealthcommission.ca/wp-content/uploads/drupal/MHCC_Report_Base_Case_FINAL_ENG_0_0.pdf [accessed 2022-10-05]
5. Survey on COVID-19 and Mental Health, February to May 2021. Statistics Canada. 2021 Sep 27. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/210927/dq210927a-eng.htm [accessed 2022-10-05]

XSL•FO
RenderX

6. Morin CM, Bjorvatn B, Chung F, Holzinger B, Partinen M, Penzel T, et al. Insomnia, anxiety, and depression during the COVID-19 pandemic: an international collaborative study. Sleep Med 2021 Nov;87:38-45 [FREE Full text] [doi: 10.1016/j.sleep.2021.07.035] [Medline: 34508986]
7. Health system resources for mental health and addictions care in Canada. Canadian Institute for Health Information. 2019. URL: https://www.cihi.ca/sites/default/files/document/mental-health-chartbook-report-2019-en-web.pdf [accessed 2022-10-05]
8. Mental health in the balance: ending the health care disparity in Canada. Canadian Mental Health Association. 2018 Sep 17. URL: https://cmha.ca/brochure/mental-health-in-the-balance-ending-the-health-care-disparity-in-canada/ [accessed 2022-10-05]
9. Mental health care needs, 2018. Health Fact Sheets. Statistics Canada. 2018. URL: https://www150.statcan.gc.ca/n1/en/catalogue/82-625-X201900100011 [accessed 2022-10-05]
10. Moroz N, Moroz I, D'Angelo MS. Mental health services in Canada: barriers and cost-effective solutions to increase access. Healthc Manage Forum 2020 Nov;33(6):282-287. [doi: 10.1177/0840470420933911] [Medline: 32613867]
11. E-Mental Health. Mental Health Commission of Canada. URL: https://mentalhealthcommission.ca/what-we-do/e-mental-health/ [accessed 2022-10-05]
12. Lal S. E-mental health: Promising advancements in policy, research, and practice. Healthc Manage Forum 2019 Mar 10;32(2):56-62 [FREE Full text] [doi: 10.1177/0840470418818583] [Medline: 30739487]
13. Advancing the evolution: insights into the state of e-Mental health services in Canada. Mental Health Commission of Canada. 2017. URL: https://mentalhealthcommission.ca/resource/advancing-the-evolution-insights-into-the-state-of-e-mental-health-services-in-canada/ [accessed 2022-10-05]
14. Strudwick G, Sockalingam S, Kassam I, Sequeira L, Bonato S, Youssef A, et al. Digital interventions to support population mental health in Canada during the COVID-19 pandemic: rapid review. JMIR Ment Health 2021 Mar 02;8(3):e26550 [FREE Full text] [doi: 10.2196/26550] [Medline: 33650985]
15. Soh HL, Ho RC, Ho CS, Tam WW. Efficacy of digital cognitive behavioural therapy for insomnia: a meta-analysis of randomised controlled trials. Sleep Med 2020 Nov;75:315-325. [doi: 10.1016/j.sleep.2020.08.020] [Medline: 32950013]
16. Strudwick G, Impey D, Torous J, Krausz RM, Wiljer D. Advancing e-Mental health in Canada: report from a multistakeholder meeting. JMIR Ment Health 2020 Apr 30;7(4):e19360 [FREE Full text] [doi: 10.2196/19360] [Medline: 32330114]
17. McGrath P, Wozney L, Bishop A, Curran J, Chorney J, Rathore SS. Toolkit for e-Mental Health Implementation. Mental Health Commission of Canada. 2018. URL: https://mentalhealthcommission.ca/resource/toolkit-for-e-mental-health- implementation/ [accessed 2022-10-05]
18. Strudwick G, McLay D, Lo B, Shin HD, Currie L, Thomson N, et al. Development of a resource guide to support the engagement of mental health providers and patients with digital health tools: multimethod study. J Med Internet Res 2021 Apr 22;23(4):e25773 [FREE Full text] [doi: 10.2196/25773] [Medline: 33885374]
19. Andersen R, Newman J. Societal and individual determinants of medical care utilization in the United States. Milbank Q 2005;83(4):1-28. [doi: 10.1111/j.1468-0009.2005.00428.x]
20. Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. Milbank Mem Fund Q Health Soc 1973;51(1):95-124. [Medline: 4198894]
21. Canadian Digital Health Survey: Understanding Canadians' experiences with digital health. Canada Health Infoway. URL: https://insights.infoway-inforoute.ca/digital-health-survey [accessed 2022-10-05]
22. Webelhorst C, Jepsen L, Rummel-Kluge C. Utilization of e-mental-health and online self-management interventions of patients with mental disorders-A cross-sectional analysis. PLoS One 2020 Apr 20;15(4):e0231373 [FREE Full text] [doi: 10.1371/journal.pone.0231373] [Medline: 32310991]
23. Adler-Milstein J. From digitization to digital transformation: policy priorities for closing the gap. JAMA 2021 Feb 23;325(8):717-718. [doi: 10.1001/jama.2020.27014] [Medline: 33620414]
24. Ronis ST, Slaunwhite AK, Malcom KE. Comparing strategies for providing child and youth mental health care services in Canada, the United States, and the Netherlands. Adm Policy Ment Health 2017 Nov 13;44(6):955-966. [doi: 10.1007/s10488-017-0808-z] [Medline: 28612298]
25. Andrews JA, Brown LJ, Hawley MS, Astell AJ. Older adults' perspectives on using digital technology to maintain good mental health: interactive group study. J Med Internet Res 2019 Feb 13;21(2):e11694 [FREE Full text] [doi: 10.2196/11694] [Medline: 30758292]
26. Mahajan S, Lu Y, Spatz ES, Nasir K, Krumholz HM. Trends and predictors of use of digital health technology in the United States. Am J Med 2021 Jan;134(1):129-134. [doi: 10.1016/j.amjmed.2020.06.033] [Medline: 32717188]
27. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 04;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
28. Grieco-Page H, Black CJ, Berent JM, Gautam B, Betancourt TS. Beyond the pandemic: leveraging rapid expansions in U.S. telemental health and digital platforms to address disparities and resolve the digital divide. Front Psychiatry 2021 Aug 6;12:671502. [doi: 10.3389/fpsyt.2021.671502] [Medline: 34421669]
29. Magaard JL, Seeralan T, Schulz H, Brütt AL. Factors associated with help-seeking behaviour among individuals with major depression: a systematic review. PLoS One 2017 May 11;12(5):e0176730 [FREE Full text] [doi: 10.1371/journal.pone.0176730] [Medline: 28493904]

XSL•FO
RenderX

30. Doblyte S, Jiménez-Mejías E. Understanding help-seeking behavior in depression: a qualitative synthesis of patients' experiences. Qual Health Res 2017 Jan 12;27(1):100-113. [doi: 10.1177/1049732316681282] [Medline: 27956660]
31. Griebel L, Enwald H, Gilstad H, Pohl A, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2018 Dec 18;43(4):427-442. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
32. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
33. Paige SR, Krieger JL, Stellefson ML. The influence of eHealth literacy on perceived trust in online health communication channels and sources. J Health Commun 2017 Jan 21;22(1):53-65 [FREE Full text] [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
34. Graffigna G, Barello S, Bonanomi A, Riva G. Factors affecting patients' online health information-seeking behaviours: the role of the Patient Health Engagement (PHE) Model. Patient Educ Couns 2017 Oct;100(10):1918-1927. [doi: 10.1016/j.pec.2017.05.033] [Medline: 28583722]
35. Kates N. Mental health and primary care: contributing to mental health system transformation in Canada. Can J Commun Mental Health 2017 Dec 01;36(4):33-67. [doi: 10.7870/cjcmh-2017-033]
36. Thota AB, Sipe TA, Byard GJ, Zometa CS, Hahn RA, McKnight-Eily LR, Community Preventive Services Task Force. Collaborative care to improve the management of depressive disorders: a community guide systematic review and meta-analysis. Am J Prev Med 2012 May;42(5):525-538. [doi: 10.1016/j.amepre.2012.01.019] [Medline: 22516495]
37. Ruimy D. Broadband connectivity in rural Canada: overcoming the digital divide. House of Commons. 2018. URL: https://www.ourcommons.ca/Committees/en/INDU/StudyActivity?studyActivityId=9604427 [accessed 2022-10-05]
38. 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]
39. Cohen SA, Kunicki ZJ, Drohan MM, Greaney ML. Exploring changes in caregiver burden and caregiving intensity due to COVID-19. Gerontol Geriatr Med 2021 Feb 26;7:2333721421999279 [FREE Full text] [doi: 10.1177/2333721421999279] [Medline: 33718523]
40. Park S. Caregivers' mental health and somatic symptoms during COVID-19. J Gerontol B Psychol Sci Soc Sci 2021 Mar 14;76(4):e235-e240 [FREE Full text] [doi: 10.1093/geronb/gbaa121] [Medline: 32738144]
41. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental Health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017 Jun 02;17(1):78 [FREE Full text] [doi: 10.1186/s12911-017-0474-9] [Medline: 28577543]
42. Cotton SM, Wright A, Harris MG, Jorm AF, McGorry PD. Influence of gender on mental health literacy in young Australians. Aust N Z J Psychiatry 2006 Sep 26;40(9):790-796. [doi: 10.1080/j.1440-1614.2006.01885.x] [Medline: 16911755]
43. Currin JB, Hayslip B, Temple JR. The relationship between age, gender, historical change, and adults' perceptions of mental health and mental health services. Int J Aging Hum Dev 2011;72(4):317-341. [doi: 10.2190/AG.72.4.c] [Medline: 21977677]
44. Thiese MS, Ronna B, Ott U. P value interpretations and considerations. J Thorac Dis 2016 Sep;8(9):E928-E931. [doi: 10.21037/jtd.2016.08.16] [Medline: 27747028]
45. Davern M, Rodin H, Beebe T, Call K. The effect of income question design in health surveys on family income, poverty and eligibility estimates. Health Serv Res 2005 Oct;40(5 Pt 1):1534-1552 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00416.x] [Medline: 16174146]

Abbreviations

eHEALS: eHealth Literacy Scale
eMH: e-Mental health
SRH: self-reported health
SRMH: self-reported mental health

Edited by G Eysenbach, S Hagens; submitted 17.05.22; peer-reviewed by R Ho, P Matthews, A AL-Asadi; comments to author 29.07.22; revised version received 30.08.22; accepted 01.10.22; published 16.11.22.

Please cite as:
Yu E, Xu B, Sequeira L
Determinants of e-Mental Health Use During COVID-19: Cross-sectional Canadian Study
J Med Internet Res 2022;24(11):e39662
URL: https://www.jmir.org/2022/11/e39662
doi:10.2196/39662
PMID:36191173

XSL•FO
RenderX

©Ellie Yu, Bowen Xu, Lydia Sequeira. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Optimizing an Emergency Medical Dispatch System to Improve Prehospital Diagnosis and Treatment of Acute Coronary Syndrome: Nationwide Retrospective Study in China

Xuejie Dong[1,2], MS; Fang Ding[3], MD; Shuduo Zhou[1,2], MD, PhD; Junxiong Ma[1,2], MS; Na Li[1,2], MS; Mailikezhati Maimaitiming[1,2], BS; Yawei Xu[4], MD, PhD; Zhigang Guo[5], MM; Shaobin Jia[6], MD, PhD; Chunjie Li[5], MD, PhD; Suxin Luo[7], MD, PhD; Huiping Bian[8], MD, PhD; Gesang Luobu[9], MD, PhD; Zuyi Yuan[10], MD, PhD; Hong Shi[11], MM; Zhi-jie Zheng[1,2], MD, PhD; Yinzi Jin[1,2*], PhD; Yong Huo[12*], MD, PhD

[1]Department of Global Health, School of Public Health, Peking University, Beijing, China
[2]Institute for Global Health and Development, Peking University, Beijing, China
[3]Emergency Medical Dispatcher Industries, Beijing, China
[4]Department of Cardiology, Shanghai Tenth People's Hospital, Shanghai, China
[5]Department of Cardiology Surgery, Tianjin Chest Hospital, Tianjin, China
[6]Department of Cardiology, General Hospital of Ningxia Medical University, Ningxia, China
[7]Department of Cardiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[8]Department of Cardiology, Qinghai Province Cardio-cerebrovascular Disease Specialist Hospital, Xining, China
[9]Department of Cardiology, Tibet Autonomous Region People's Hospital, Lasa, China
[10]Department of Cardiology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China
[11]Chinese Medical Association, Beijing, China
[12]Department of Cardiology, Peking University First Hospital, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Yong Huo, MD, PhD
Department of Cardiology
Peking University First Hospital
8 Xishiku Street, Xicheng District
Beijing, 100191
China
Phone: 86 010 83572211
Email: huoyong@263.net.cn

Abstract

Background: Acute coronary syndrome (ACS) is the most time-sensitive acute cardiac event that requires rapid dispatching and response. The medical priority dispatch system (MPDS), one of the most extensively used types of emergency dispatch systems, is hypothesized to provide better-quality prehospital emergency treatment. However, few studies have revealed the impact of MPDS use on the process of ACS care.

Objective: This study aimed to investigate whether the use of MPDS was associated with higher prehospital diagnosis accuracy and shorter prehospital delay for patients with ACS transferred by an emergency medical service (EMS), using a national database in China.

Methods: This retrospective analysis was based on an integrated database of China's MPDS and hospital registry. From January 1, 2016, to December 31, 2020, EMS-treated ACS cases were divided into before MPDS and after MPDS groups in accordance with the MPDS launch time at each EMS center. The primary outcomes included diagnosis consistency between hospital admission and discharge, and prehospital delay. Multivariable logistic regression and propensity score–matching analysis were performed to compare outcomes between the 2 groups for total ACS and subtypes.

Results: A total of 9806 ACS cases (3561 before MPDS and 6245 after MPDS) treated by 43 EMS centers were included. The overall diagnosis consistency of the after MPDS group (Cohen κ=0.918, P<.001) was higher than that of the before MPDS group

XSL•FO
RenderX

(Cohen κ=0.889, *P*<.001). After the use of the MPDS, the call-to-EMS arrival time was shortened in the matched ACS cases (20.0 vs 16.0 min, *P*<.001; adjusted difference: –1.67, 95% CI –2.33 to –1.02; *P*<.001) and in the subtype of ST-elevation myocardial infarction (adjusted difference: –3.81, 95% CI –4.63 to –2.98, *P*<.001), while the EMS arrival-to-door time (20.0 vs 20.0 min, *P*=.31) was not significantly different in all ACS cases and subtypes.

Conclusions: The optimized use of MPDS in China was associated with increased diagnosis consistency and a reduced call-to-EMS arrival time among EMS-treated patients with ACS. An emergency medical dispatch system should be designed specifically to fit into different prehospital modes in the EMS system on a regional basis.

(J Med Internet Res 2022;24(11):e36929) doi:10.2196/36929

KEYWORDS

medical priority dispatch system; acute coronary syndrome; prehospital care; emergency medical service; health service; healthcare; health care; coronary; cardiology; cardiovascular

Introduction

An emergency medical dispatch system is the principal link between the public caller requesting urgent medical care and the emergency medical service (EMS) system, and forms an integral part of EMS practice [1,2]. With proper training, administration, and supervision, an emergency medical dispatcher can accurately query the caller, select an appropriate method of response, provide patient information to responders, and provide appropriate medical direction for patients through the caller. Thus, emergency medical dispatch functions through rapid recognition, rapid dispatching based on priority, and prehospital instructions [3]. Acute coronary syndrome (ACS) is the most time-sensitive acute cardiovascular disease, which requires rapid dispatching and response to dispatching beginning at the time of symptom onset [4]. Timely reperfusion therapy for ACS can be highly effective if following a "chain of survival," which consists of 3 key components: (1) early symptom recognition and call for EMS, (2) early transportation and evaluation, and (3) early in-hospital treatment. Through appropriate application and reference to a written, medically approved, emergency medical dispatch protocol, an emergency medical dispatch system can lead to a higher diagnosis accuracy and a shorter prehospital delay, which modulates better outcomes for patients with ACS [5-7].

As one of the emergency medical dispatch systems, the medical priority dispatch system (MPDS) has been widely used in more than 50 countries covering more than 3500 EMS centers. MPDS is a scripted protocol designed to direct certified dispatchers to identify the presented symptoms and provide prehospital medical directions based on callers' responses to scripted questions [8]. MPDS were introduced in China in 2010 and were quickly developed and applied in more than 80 EMS centers after the National Health Commission implemented the Notice on Strengthening the Capacity of Healthcare Delivery for Acute Cardiovascular Diseases in 2015. However, in addition to dispatching and EMS responses, patients with ACS need coordinated care between EMS and hospitals at the regional level, in which EMS providers obtain prehospital electrocardiograms and activate cardiac catheterization laboratories before hospital arrival, bypass the emergency department when appropriate, and provide ongoing quality review and feedback [9]. Therefore, efforts should be focused on information sharing among dispatchers, EMS providers on ambulance, and health care professionals in hospitals.

In China, the EMS framework was designed specifically to fit into the local health care system. Based on the department in charge of dispatching, prehospital transport, and in-hospital treatment functions, there are at least 4 prehospital EMS system models varying across cities: independent, prehospital, dispatching, and dependent models [10] (Multimedia Appendix 1). The dependent model is the main one encompassing more than 80% of the EMS centers, and the dispatching and independent ones only exist in a few developed cities [11]. The MPDS of China has taken the lead in establishing an information sharing system by linking the EMS and the hospitals to facilitate the coordination of care at the time of entering the EMS system. The MPDS of China is optimized in that it has focused on the establishment of regional systems of ACS care by integrating health care among EMS providers, emergency departments physicians, cardiologists, and catheterization laboratory staff.

A number of studies have verified the accuracy of MPDS dispatch codes in regard to prehospital acuity [12-14]. Prior studies focused on the impact of MPDS use on patients' outcomes were limited to out-of-hospital cardiac arrest cases [15-17]. However, few studies have revealed the impact of MPDS use on the process of ACS care. Moreover, to our knowledge, no studies have focused on the effectiveness of MPDS use in China and other low- and middle-income countries. To fill the gaps, the objective of this study was to investigate whether the use of the optimized MPDS is associated with higher diagnostic accuracy and a shorter prehospital delay among EMS-treated patients with ACS, using a national database in China. In our patient cohort, ACS was further divided into 3 subtypes: ST-elevation myocardial infarction (STEMI), non–ST-elevation myocardial infarction (NSTEMI), and unstable angina pectoris (UA).

Methods

Study Design and Data Source

This retrospective analysis was based on the database of the China MPDS registry and its registered hospitals from January 1, 2016, to December 31, 2020. Data on registered hospitals were extracted from the Chinese Cardiovascular Association Database-Chest Pain Center—a nationwide, web-based, unified database that collects data of patients discharged from the

XSL•FO
RenderX

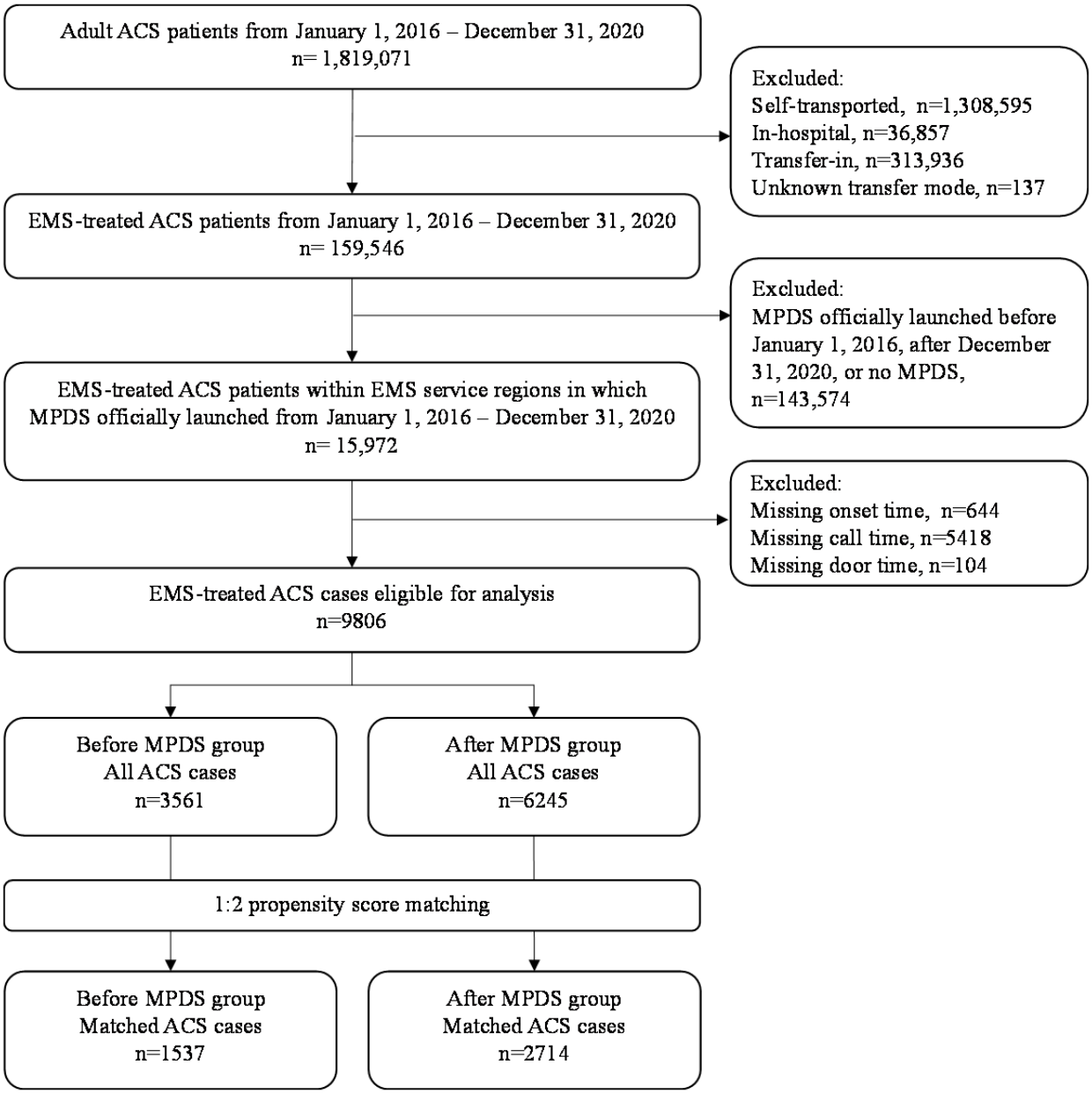
hospital-based chest pain centers [18]. The MPDS registry database collected information of all the EMS users of the MPDS across China, including the name of EMS centers, the date of their official launching of the MPDS, and the code of the administrative region covered by their service.

Based on the date of the official launch of the MPDS at each EMS center, enrolled cases within the EMS service regions were divided into the before MPDS and after MPDS groups. The before MPDS group included cases enrolled in the registered hospitals that had not implemented the MPDS, and the after MPDS group included cases enrolled in the registered hospitals' chest pain centers that had implemented the MPDS.

Study Population

From January 1, 2016, to December 31, 2020, a total of 15,972 patients with a discharge diagnosis of ACS (STEMI, NSTEMI, and UA subtypes) were enrolled in the registered hospitals and treated at a total of 43 EMS centers. A total of 6166 patients were excluded owing to missing data on analyzed indicators including onset time, call time, and door time. Finally, 9806 ACS cases were included in the final analyses and were divided into the before MPDS (n=3561) and after MPDS (n=6245) groups (Figure 1).

Figure 1. Flow diagram for study recruitment. ACS: acute coronary syndrome; EMS: emergency medical service; MPDS: medical priority dispatch system.



XSL•FO
RenderX

Measures

Primary outcomes included diagnosis consistency and prehospital delay. The diagnosis consistency between diagnosis upon hospital admission (the prehospital diagnosis by the EMS crew) and diagnosis at hospital discharge was indicated using the Cohen κ value. Cohen κ is one of the most common statistics to test interrater reliability and is used to measure the agreement of 2 raters or methods rating on categorical scales [19]. We computed the Cohen κ value to assess the agreement in diagnosing 3 specific subtypes of ACS (STEMI, NSTEMI, and UA) between hospital admission diagnosis and hospital discharge diagnosis.

The prehospital delay was measured by the call-to-EMS arrival time (the time interval from the EMS dispatcher receiving the emergency call from the patient or bystander to ambulance arrival at the scene), the EMS arrival-to-door time (the time interval from EMS arrival at the scene to EMS arrival at the hospital), and total EMS time (the time interval from the EMS dispatcher receiving the emergency call to EMS arrival at the hospital). Covariates for prehospital delay included patients' demographic characteristics (age and gender), onset environment (city level, call time of the day, and call time of the week), and event characteristics (onset-to-call time, precall chest pain symptoms, type of ACS, and Killip class).

Data Analysis

We compared the characteristics and outcomes of the study population between the before MPDS and after MPDS groups using a 2-tailed independent samples *t* test and Wilcoxon signed-rank test for continuous variables and the chi-square test for categorical variables. Continuous variables are reported as mean (SD) or median (IQR) values; categorical variables, as n (%) values. To examine the impact of the optimized MPDS on prehospital delay, we used 2 models including propensity score–matching analysis and multivariable logistic regression analysis. Both models were adjusted for precall covariates including patients' demographic characteristics (age and gender), onset environment (city level, call time of the day, and call time of the week), and event characteristics (onset-to-call time, precall chest pain symptoms, type of ACS, and Killip class), with $P<.05$ considered the threshold for statistical significance. In the propensity score–matching analyses, 1:2 matching was performed without replacement for each patient, using a nearest-neighbor matching algorithm with a caliper width of 0.02. Matched patients were compared to assess balance in covariates (ie, standardized differences for each covariate were <10%). In the multivariable logistic regression analysis, adjusted differences with 95% CIs are presented. All statistical analyses were performed using R (version 4.0.4; The R Foundation).

Ethical Considerations

This study was approved by the institutional review board of Peking University (IRB00001052-21020). Informed consent was obtained from all participants prior to questionnaire administration.

Results

Patient Characteristics

Compared to the before MPDS group (n=3561), the after MPDS group (N=6245) comprised younger patients (mean 65.6, SD 12.9 vs mean 66.2, SD 13.4 years, respectively, $P=.03$), had a higher proportion of cases from provincial capital cities (41.4% vs 32.4%, $P<.001$), and had a higher proportion of patients with STEMI (62.9% vs 57.8%, $P<.001$) and Killip class I myocardial infarction (77.8% vs 72.8%, $P<.001$). After propensity score–matching, 2715 patients in the before MPDS group and 5429 patients in the after MPDS group were matched (Table 1).

XSL•FO
RenderX

Table 1. Characteristics and outcomes of acute coronary syndrome cases, before and after dispatch with the optimized medical priority dispatch system (MPDS), total cases (N=9806), and propensity score–matched cases.

Characteristics and Outcomes	Total cases			Propensity score–matched cases[a] (n=8144)			Standardized mean difference
	Before MPDS (n=3561)	After MPDS (n=6245)	*P* value	Before MPDS (n=2715)	After MPDS (n=5429)	*P* value	
Age (years), mean (SD)	66.2 (13.4)	65.6 (12.9)	.03	62.6 (13.3)	65.2 (12.9)	.25	0.02
Male, n (%)	2588 (72.7)	4414 (70.7)	.04	2008 (74.0)	3888 (71.6)	.03	0.05
Living in a provincial capital city, n (%)	1168 (32.4)	2587 (41.4)	<.001	1027 (37.8)	2251 (41.5)	.002	0.20
Call time of the day, n (%)			.10			.07	0.06
12-5:59 AM	651 (18.3)	1252 (20.0)		491 (18.1)	1089 (20.1)		
6-11:59 AM	1154 (32.4)	2030 (32.5)		860 (31.7)	1765 (32.5)		
12-5:59 PM	875 (24.6)	1518 (24.3)		690 (25.4)	1324 (24.4)		
6-11:59 PM	881 (24.7)	1445 (23.1)		674 (24.8)	1251 (23.0)		
Call on weekday, n (%)	2516 (70.7)	4505 (72.1)	.12	1914 (70.5)	3925 (72.3)	.09	0.04
Precall chest pain, n (%)[b]			.004			.21	0.04
Persistent chest pain	2233 (69.1)	4221 (72.1)		1954 (72.0)	3960 (72.9)		
Intermittent chest pain	837 (25.9)	1336 (22.8)		644 (23.7)	1206 (22.2)		
Eased chest pain	163 (5.0)	299 (5.1)		117 (4.3)	263 (4.8)		
Type of acute coronary syndrome, n (%)			<.001			.21	0.06
ST-elevation myocardial infarction	2057 (57.8)	3928 (62.9)		1703 (62.7)	3508 (64.6)		
Non–ST-elevation myocardial infarction	627 (17.6)	1084 (17.4)		554 (20.4)	971 (17.9)		
Unstable angina pectoris	877 (24.6)	1233 (19.7)		458 (16.9)	950 (17.5)		
Killip class, n (%)[b]			<.001			.05	0.07
I	2333 (72.8)	4495 (77.8)		2073 (76.4)	4269 (78.6)		
II-III	625 (19.5)	876 (15.2)		440 (16.2)	814 (15.0)		
IV	245 (7.6)	408 (7.1)		202 (7.4)	346 (6.4)		
Onset-to-call time (minutes), median (IQR)	54.0 (18.0-124.0)	56.0 (20.0-124.0)	.48	56.0 (20.0-128.0)	56.0 (20.0-124.0)	.87	0.03
Call-to-EMS[c] arrival time (min), median (IQR)	18.0 (12.0-30.0)	16.0 (10.0-26.0)	<.001	20.0 (12.0-30.0)	16.0 (10.0-26.0)	<.001	N/A[d]
EMS arrival-to-door time (minutes), median (IQR)	20.0 (12.0-30.0)	20.0 (12.0-28.0)	<.001	20.0 (12.0-30.0)	20.0 (12.0-30.0)	.31	N/A
Total EMS time (call-to-door; minutes), median (IQR)	40.0 (28.0-56.0)	38.0 (28.0-52.0)	<.001	40.0 (29.0-58.0)	38.0 (28.0-52.0)	<.001	N/A

[a]Propensity score matched for age, gender, city level, call time of the day, call on weekday, precall chest pain symptoms, type of acute coronary syndrome, Killip class, and onset-to-call time.

[b]Missing cases were excluded when comparing the precall chest pain symptoms and Killip class between the 2 groups.

[c]EMS: emergency medical service.

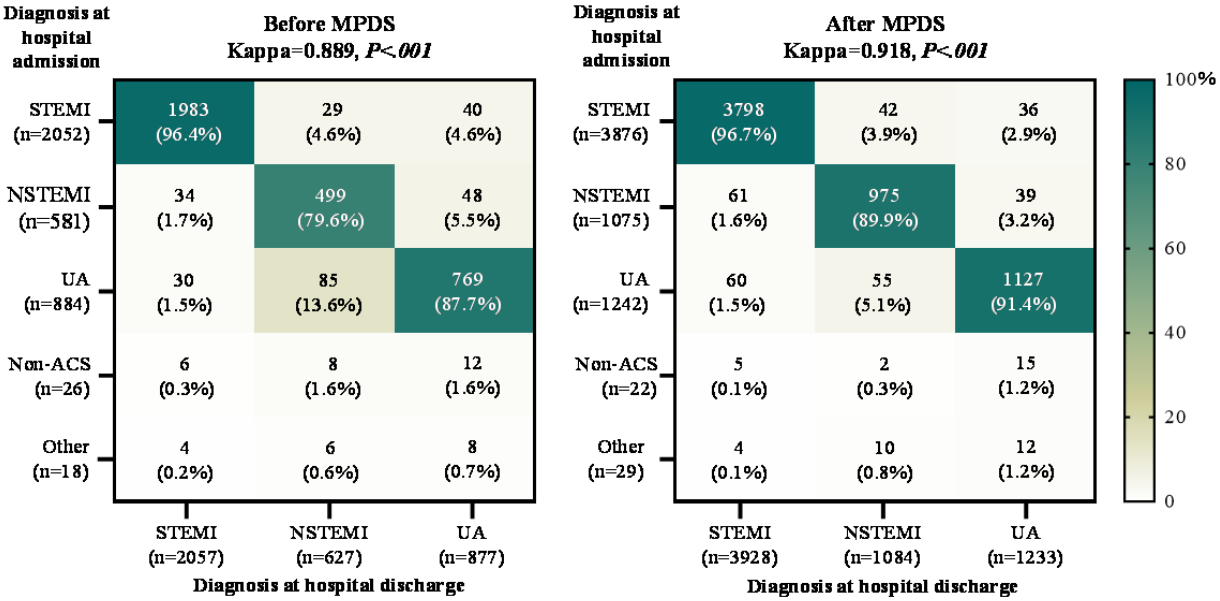
[d]N/A: not applicable.

Diagnosis Consistency

The Cohen κ of all ACS subtypes was higher in the after MPDS group (0.918, *P*<.001) than in the before MPDS group (0.889, *P*<.001). Specifically, diagnosis consistency of NSTEMI (79.6% vs 89.9%, *P*<.001) and UA (87.7% vs 91.4%, *P*=.001) were remarkably improved after the use of the optimized MPDS, while that of STEMI (96.4% vs 96.7%, *P*>.99) was not significantly changed (Figure 2). Moreover, 44 of 3561 (1.2%) of patients in the before MPDS group and 51 of 6245 (0.8%) patients in the after MPDS group with a discharge diagnosis of ACS chest pain were treated for non-ACS chest pain or other diseases upon admission.

XSL•FO
RenderX

Figure 2. Diagnosis consistency between hospital admission and discharge of EMS-treated patients with ACS before and after dispatch with the optimized MPDS. Data are reported as the proportion of each diagnosis upon admission among the cases with certain diagnosis at discharge. Cohen κ values were computed in 3 certain types of ACS (STEMI, NSTEMI, UA) diagnosis between admission and discharge. Non-ACS cases include pulmonary embolism, aortic dissection, and non-cardiogenic chest pain. Other cases include non–chest pain and unknown diagnosis. ACS: acute coronary syndrome; MPDS: medical priority dispatch system; STEMI: ST-elevation myocardial infarction; NSTEMI: non–ST-elevation myocardial infarction; UA: unstable angina pectoris.



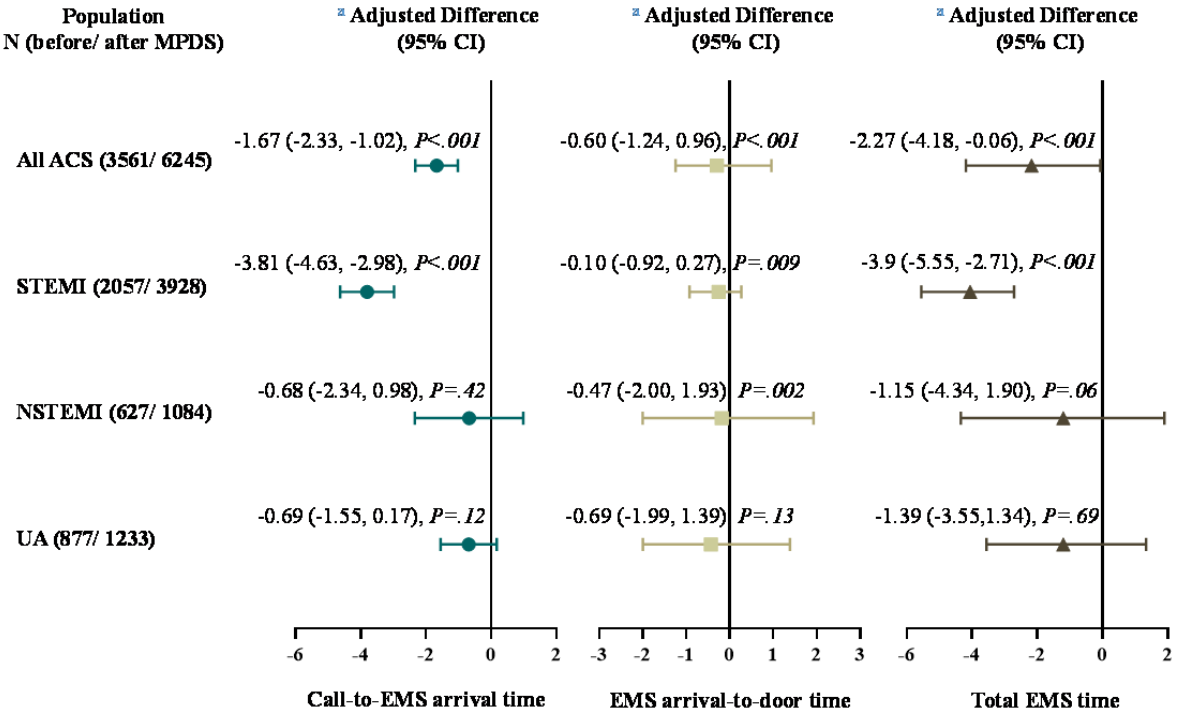
Prehospital Delay

In the propensity score–matched population, call-to-EMS arrival time (20.0 vs 16.0 minutes, *P*<.001) and the total EMS time (40.0 vs 38.0 minutes, *P*<.001) were significantly shorter after the use of MPDS, while the EMS arrival-to-door time (20.0 vs 20.0 minutes, *P*=.31) between the before and after MPDS groups were not significantly different (Table 1).

Patients in the after MPDS group had a significantly shorter call-to-EMS arrival time than those in the before MPDS group in all ACS cases (adjusted difference –1.67, 95% CI –2.33 to –1.01, *P*<.001), and those of the STEMI subtype (adjusted difference –3.81, 95% CI –4.63 to –2.98, *P*<.001). There were no significant differences between the 2 groups in EMS arrival-to-door time in all ACS cases or subtypes (Figure 3).

XSL•FO
RenderX

Figure 3. Multivariate analysis of outcomes using a generalized linear model in all ACS cases and for each subgroup with STEMI, NSTEMI, and UA. The model was adjusted for age, gender, city level, call time of the day, and calling on a weekday. Call-to-EMS arrival time was defined as the time interval from EMS dispatcher receiving the emergency call to ambulance arrival at the scene; the EMS arrival-to-door time was defined as the time interval from EMS arrival at the scene to EMS arrival at the hospital; total EMS time was defined as the time interval from the EMS dispatcher receiving the emergency call to EMS arrival at the hospital. ACS: acute coronary syndrome; EMS: emergency medical service; MPDS: medical priority dispatch system; NSTEMI: non–ST-elevation myocardial infarction; STEMI: ST-elevation myocardial infarction; UA: unstable angina pectoris.



Discussion

Principal Findings

In this retrospective study of EMS-treated patients based on a national database in China, we found that the use of the optimized MPDS was associated with a higher consistency between diagnosis at hospital admission and discharge and a shorter call-to-EMS arrival time; however, there were no significant differences in the EMS arrival-to-door time among patients with ACS. Our findings are consistent with those of prior studies in which the use of MPDS has been proven to be associated with high dispatching accuracy [12,13] and improved dispatch efficacy [14], which could potentially prove the general assumption that MPDS could provide higher diagnosis accuracy and lesser prehospital delay, thereby potentially resulting in better survival outcomes for ACS.

The first assumption was that the optimized MPDS could help rapidly identify and diagnose diseases, which theoretically led to a higher diagnostic accuracy of EMS. In this study, we observed an increase in overall diagnosis consistency between hospital admission and discharge after the use of the optimized MPDS, which was similar to the diagnostic accuracy of ACS in China that reported elsewhere [20,21]. This suggests that although the optimized MPDS might not provide a definite diagnosis for each case, it has the potential to allocate patients to the right priority levels in accordance with their symptom presentation. In fact, the MPDS was purposefully designed to be highly sensitive and to avoid undertriage by creating overtriage so as to ensure patient care and safety at the first place [12,22,23]. We also observed that a lower proportion of patients with ACS were treated for other diseases upon admission, which might lead to reduced wastage of resources and risk for personnel [24]. Nevertheless, our findings once again revealed the complexity of the diagnosis of ACS.

The second assumption was that the optimized MPDS could reduce prehospital delay through timely dispatch and appropriate EMS responses. In fact, the use of the optimized MPDS reduced the transportation delay in the call-to-EMS arrival time; however, it did not translate to a shorter EMS arrival-to-door time. On the one hand, although the MPDS of China has taken efforts to establishing the information sharing system to integrate health care between the EMS and the hospital-based chest pain centers, it was still only involved in the process from call receiving to EMS arrival at the scene. On the other hand, the reduced call-to-EMS arrival time indicated the adaptability of the optimized MPDS in China's EMS system at the local level. As indicated, the varied EMS systems in China could be classified into 4 main models. In spite of different characteristics, all 4 models could present prehospital delay. The independent model and prehospital model tend to have longer dispatching and ambulance returning times, especially within broad regions with limited health resources. In these cities, the optimized MPDS's priorities could help dispatchers mobilize health resources, which may avoid unnecessary wastage of health resources, thus shortening the time of dispatching and arriving at the scene. For the dispatching model and dependent model, the response speed of hospitals may be worse than expected owing to limited authority of the EMS, leading to low response to dispatching. The optimized MPDS follows standardized procedures and records detailed registration

XSL•FO
RenderX

of every emergency call and would empower the EMS with greater authority, which may improve the responsiveness of hospitals to dispatching, thus reducing the call-to-EMS arrival time. Therefore, to further improve the impact of the optimized MPDS, the optimized MPDS should be designed specifically to fit into different prehospital models of the EMS system on a regional basis.

The third assumption was that with a higher diagnostic accuracy and a shorter prehospital delay, the optimized MPDS could result in better survival outcomes for ACS. Though the outcome data could not be obtained and analyzed in this study, the onset-to-call time was still near 1 hour; thus, the optimized MPDS could hardly predict improved in-hospital mortality. For time-concerning emergencies such as ACS, the first link of the chain of survival would always be early symptom recognition and seeking for EMS by the public [25-27]. Any subsequent treatment will not be effective without timely activation of this first link. Given the fact that 1-year mortality for patients with ACS would increase by 7.5% with every additional 30 minutes of prehospital delay [4], this large period between symptom onset to call would always limit what the optimized MPDS can do. Therefore, what should be designed in a dispatching system and whether its implementation can result in satisfactory effects not only depend on the EMS but also require the joint efforts of the public, EMS, and hospitals.

Limitations

This study had some limitations. First, this was a retrospective study, which increased the risk of residual confounding. Although we eliminated imbalance between the groups through propensity score–matching analysis, unmeasured confounding factors may have influenced the outcomes. Second, our study population comprised EMS-treated patients enrolled at chest pain centers, and all patients were at least alive at the time of admission, which might limit the generalizability of our findings. However, our comparison between propensity score–matched groups was able to eliminate this bias. Third, we failed to classify our included EMS systems into specific types of EMS models because of a lack of an official classification, which may limit the certainty of our findings. Fourth, owing to limited variables in the database, we could not obtain the survival outcome; we failed to determine the call processing time, the ambulance dispatch time, or EMS on-scene time, which would affect the prehospital delay and could be impacted by the MPDS; for the measures of diagnosis accuracy of the optimized MPDS, we could only compute the Cohen κ using the disease diagnosis rather than priority levels, while the sensitivity and specificity for discriminative, positive, and negative predictive values could not be obtained.

Conclusions

The use of the optimized MPDS in China was associated with a higher diagnosis consistency and a shorter call-to-EMS arrival time; however, no potentially improved EMS-to-door time among EMS-treated patients with ACS. These benefits can be realized by the emergency medical dispatch system when coordinated care between the EMS and hospitals was delivered on the regional level.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (No. 71904004), and the Beijing Natural Science Foundation (No. 9204025).

Authors' Contributions

YJ and XD contributed to the conception of the study. YX, ZG, SJ, CL, SL, HB, GL, and ZY contributed to the acquisition of data. XD, JM, NL, and MM contributed significantly to data analysis and manuscript preparation. YJ, SZ, and HS finalized the manuscript. ZZ and YH provided administrative advice and consultations. All authors contributed substantially to the revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prehospital modes of China’s EMS system.
[DOCX File , 15 KB - jmir_v24i11e36929_app1.docx]

References

1. Al-Shaqsi S. Models of international emergency medical service (EMS) systems. Oman Med J 2010 Oct;25(4):320-323 [FREE Full text] [doi: 10.5001/omj.2010.92] [Medline: 22043368]
2. Kashani S, Sanko S, Eckstein M. The critical role of dispatch. Cardiol Clin 2018 Aug;36(3):343-350. [doi: 10.1016/j.ccl.2018.03.001] [Medline: 30293600]
3. Hettinger A, Cushman J, Shah M, Noyes K. Emergency medical dispatch codes association with emergency department outcomes. Prehosp Emerg Care 2013;17(1):29-37. [doi: 10.3109/10903127.2012.710716] [Medline: 23140195]
4. De Luca G, Suryapranata H, Ottervanger JP, Antman EM. Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction. Circulation 2004 Mar 16;109(10):1223-1225. [doi: 10.1161/01.cir.0000121424.76486.20]

XSL•FO
RenderX

5. Khraim FM, Carey MG. Predictors of pre-hospital delay among patients with acute myocardial infarction. Patient Educ Couns 2009 May;75(2):155-161. [doi: 10.1016/j.pec.2008.09.019] [Medline: 19036551]
6. Langabeer JR, Dellifraine J, Fowler R, Jollis JG, Stuart L, Segrest W, et al. Emergency medical services as a strategy for improving ST-elevation myocardial infarction system treatment times. J Emerg Med 2014 Mar;46(3):355-362. [doi: 10.1016/j.jemermed.2013.08.112] [Medline: 24268897]
7. Brokmann JC, Conrad C, Rossaint R, Bergrath S, Beckers SK, Tamm M, et al. Treatment of acute coronary syndrome by telemedically supported paramedics compared with physician-based treatment: a prospective, interventional, multicenter trial. J Med Internet Res 2016 Dec 01;18(12):e314 [FREE Full text] [doi: 10.2196/jmir.6358] [Medline: 27908843]
8. Sporer K, Johnson N, Yeh C, Youngblood GM. Can emergency medical dispatch codes predict prehospital interventions for common 9-1-1 call types? Prehosp Emerg Care 2008;12(4):470-478. [doi: 10.1080/10903120802290877] [Medline: 18924011]
9. Jollis J, Al-Khalidi H, Roettig M, Berger P, Corbett C, Doerfler S, et al. Impact of regionalization of ST-segment–elevation myocardial infarction care on treatment times and outcomes for emergency medical services–transported patients presenting to hospitals with percutaneous coronary intervention. Circulation 2018 Jan 23;137(4):376-387. [doi: 10.1161/circulationaha.117.032446]
10. Tan Z. Discussion on the current situation of pre-hospital first aid in domestic and overseas and the new model of grass-roots collaboration. Chinese Journal of Disaster Medicine 2020 Dec 29;8(12):692-694. [doi: 10.13919/j.issn.2095-6274.2020.12.010]
11. Shi X, Bao J, Zhang H, Wang H, Wang Y, Li L, et al. Emergency medicine in China: a review of the history of progress and current and future challenges after 40 years of reform. Am J Emerg Med 2020 Mar;38(3):662-669 [FREE Full text] [doi: 10.1016/j.ajem.2019.11.008] [Medline: 31902696]
12. Ball SJ, Williams TA, Smith K, Cameron P, Fatovich D, O'Halloran KL, et al. Association between ambulance dispatch priority and patient condition. Emerg Med Australas 2016 Dec 04;28(6):716-724. [doi: 10.1111/1742-6723.12656] [Medline: 27592247]
13. Clawson J, Olola C, Heward A, Patterson B, Scott G. The Medical Priority Dispatch System's ability to predict cardiac arrest outcomes and high acuity pre-hospital alerts in chest pain patients presenting to 9-9-9. Resuscitation 2008 Sep;78(3):298-306. [doi: 10.1016/j.resuscitation.2008.03.229] [Medline: 18562077]
14. Clawson J, Gardett I, Scott G, Fivaz C, Barron T, Broadbent M, et al. Hospital-confirmed acute myocardial infarction: prehospital identification using the medical priority dispatch system. Prehosp Disaster Med 2018 Feb;33(1):29-35. [doi: 10.1017/S1049023X1700704X] [Medline: 29223194]
15. Bohm K, Vaillancourt C, Charette M, Dunford J, Castrén M. In patients with out-of-hospital cardiac arrest, does the provision of dispatch cardiopulmonary resuscitation instructions as opposed to no instructions improve outcome: a systematic review of the literature. Resuscitation 2011 Dec;82(12):1490-1495. [doi: 10.1016/j.resuscitation.2011.09.004] [Medline: 21925129]
16. Hardeland C, Olasveengen TM, Lawrence R, Garrison D, Lorem T, Farstad G, et al. Comparison of medical priority dispatch (MPD) and criteria based dispatch (CBD) relating to cardiac arrest calls. Resuscitation 2014 May;85(5):612-616. [doi: 10.1016/j.resuscitation.2014.01.029] [Medline: 24525117]
17. Shah M, Bartram C, Irwin K, Vellano K, McNally B, Gallagher T, et al. Evaluating dispatch-assisted CPR using the CARES registry. Prehosp Emerg Care 2018;22(2):222-228. [doi: 10.1080/10903127.2017.1376133] [Medline: 29220603]
18. Xiang D, Jin Y, Fang W, Su X, Yu B, Wang Y, et al. The national chest pain centers program: monitoring and improving quality of care for patients with acute chest pain in China. Cardiol Plus 2021;6(3):187. [doi: 10.4103/2470-7511.327239]
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
20. Chen T, Luo R, Zhou Z. The clinical effects of different pre-hospital emergency modes on patients with acute coronary syndrome (ACS). Chin Foreign Med Treatment 2021;40(5):7-10. [doi: 10.16662/j.cnki.1674-0742.2021.05.007]
21. Teng F, Li M, Zhang Y, Wu D, Lu F. Application of Seamless Connection Mode in Pre-hospital and Intra-hospital in First Aid of Acute Coronary Syndrome. Journal of Frontiers of Medicine 2014 Mar;3(8):40-41. [doi: 10.3969/j.issn.2095-1752.2014.08.030]
22. Bohm K, Kurland L. The accuracy of medical dispatch - a systematic review. Scand J Trauma Resusc Emerg Med 2018 Nov 09;26(1):94 [FREE Full text] [doi: 10.1186/s13049-018-0528-8] [Medline: 30413213]
23. Du L. Medical emergency resource allocation model in large-scale emergencies based on artificial intelligence: algorithm development. JMIR Med Inform 2020 Jun 25;8(6):e19202 [FREE Full text] [doi: 10.2196/19202] [Medline: 32584262]
24. Wilmer I, Chalk G, Davies G, Weaver A, Lockey DJ. Air ambulance tasking: mechanism of injury, telephone interrogation or ambulance crew assessment? Emerg Med J 2015 Oct;32(10):813-816. [doi: 10.1136/emermed-2013-203204] [Medline: 25527473]
25. Atkins D, de Caen AR, Berger S, Samson R, Schexnayder S, Joyner BJ, et al. 2017 American Heart Association focused update on pediatric basic life support and cardiopulmonary resuscitation quality: an update to the American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2018 Jan 02;137(1):A. [doi: 10.1161/cir.0000000000000540]

XSL•FO
RenderX

26. El-Deeb MH. The chain of survival for ST-segment elevation myocardial infarction: insights into the Middle East. Crit Pathw Cardiol 2013 Sep;12(3):154-160. [doi: 10.1097/HPC.0b013e3182901f28] [Medline: 23892947]
27. El Khoury R, Jung R, Nanda A, Sila C, Abraham M, Castonguay A, et al. Overview of key factors in improving access to acute stroke care. Neurology 2012 Sep 25;79(13 Suppl 1):S26-S34. [doi: 10.1212/WNL.0b013e3182695a2a] [Medline: 23008407]

Abbreviations

ACS: acute coronary syndrome
EMS: emergency medical service
MPDS: medical priority dispatch system
NSTEMI: non–ST-elevation myocardial infarction
STEMI: ST-elevation myocardial infarction
UA: unstable angina pectoris

Edited by G Eysenbach; submitted 30.01.22; peer-reviewed by M Tomey, K Yamada, J Kennel; comments to author 12.08.22; revised version received 04.10.22; accepted 20.10.22; published 23.11.22.

Please cite as:
Dong X, Ding F, Zhou S, Ma J, Li N, Maimaitiming M, Xu Y, Guo Z, Jia S, Li C, Luo S, Bian H, Luobu G, Yuan Z, Shi H, Zheng ZJ, Jin Y, Huo Y
Optimizing an Emergency Medical Dispatch System to Improve Prehospital Diagnosis and Treatment of Acute Coronary Syndrome: Nationwide Retrospective Study in China
J Med Internet Res 2022;24(11):e36929
URL: https://www.jmir.org/2022/11/e36929
doi:10.2196/36929
PMID:36416876

©Xuejie Dong, Fang Ding, Shuduo Zhou, Junxiong Ma, Na Li, Mailikezhati Maimaitiming, Yawei Xu, Zhigang Guo, Shaobin Jia, Chunjie Li, Suxin Luo, Huiping Bian, Gesang Luobu, Zuyi Yuan, Hong Shi, Zhi-jie Zheng, Yinzi Jin, Yong Huo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Hospital Digitization on Clinical Outcomes and Patient Satisfaction: Nationwide Multiple Regression Analysis Across German Hospitals

Philip von Wedel[1], MSc; Christian Hagist[1], Prof Dr; Jan-David Liebe[2], Prof Dr; Moritz Esdar[3], MA; Ursula Hübner[3], Prof Dr; Christoph Pross[4], Dr

[1]Chair of Economic and Social Policy, WHU - Otto Beisheim School of Management, Vallendar, Germany

[2]Department of Medical Management, Faculty of Health Sciences, Medical School Hamburg, Hamburg, Germany

[3]Health Informatics Research Group, Faculty of Business Management and Social Sciences, University of Applied Sciences Osnabrueck, Osnabrueck, Germany

[4]Department of Health Care Management, Berlin University of Technology, Berlin, Germany

Corresponding Author:
Philip von Wedel, MSc
Chair of Economic and Social Policy
WHU - Otto Beisheim School of Management
Burgplatz 2
Vallendar, 56179
Germany
Phone: 49 02616509 ext 255
Email: philip.wedel@whu.edu

Abstract

Background: The adoption of health information technology (HIT) by health care providers is commonly believed to improve the quality of care. Policy makers in the United States and Germany follow this logic and deploy nationwide HIT adoption programs to fund hospital investments in digital technologies. However, scientific evidence for the beneficial effects of HIT on care quality at a national level remains mostly US based, is focused on electronic health records (EHRs), and rarely accounts for the quality of digitization from a hospital user perspective.

Objective: This study aimed to examine the effects of digitization on clinical outcomes and patient experience in German hospitals. Hence, this study adds to the small stream of literature in this field outside the United States. It goes beyond assessing the effects of mere HIT adoption and also considers user-perceived HIT value. In addition, the impact of a variety of technologies beyond EHRs was examined.

Methods: Multiple linear regression models were estimated using emergency care outcomes, elective care outcomes, and patient satisfaction as dependent variables. The adoption and user-perceived value of HIT represented key independent variables, and case volume, hospital size, ownership status, and teaching status were included as controls. Care outcomes were captured via risk-adjusted, observed-to-expected outcome ratios for patients who had stroke, myocardial infarction, or hip replacement. The German Patient Experience Questionnaire of Weisse Liste provided information on patient satisfaction. Information on the adoption and user-perceived value of 10 subdomains of HIT and EHRs was derived from the German 2020 Healthcare IT Report.

Results: Statistical analysis was based on an overall sample of 383 German hospitals. The analyzed data set suggested no significant effect of HIT or EHR adoption on clinical outcomes or patient satisfaction. However, a higher user-perceived value or quality of the installed tools did improve outcomes. Emergency care outcomes benefited from user-friendly overall digitization (β=−.032; *P*=.04), which was especially driven by the user-friendliness of admission HIT (β=−.023; *P*=.07). Elective care outcomes were positively impacted by user-friendly EHR installations (β=−.138; *P*=.008). Similarly, the results suggested user-friendly, overall digitization to have a moderate positive effect on patient satisfaction (β=−.009; *P*=.01).

Conclusions: The results of this study suggest that hospital digitization is not an end in itself. Policy makers and hospitals are well advised to not only focus on the mere adoption of digital technologies but also continuously work toward digitization that is perceived as valuable by physicians and nurses who rely on it every day. Furthermore, hospital digitization strategies should consider that the assumed benefits of single technologies are not realized across all care domains.

XSL•FO
RenderX

(J Med Internet Res 2022;24(11):e40124) doi:10.2196/40124

KEYWORDS

health care information technology; electronic health records; hospital digitization; quality of care; clinical outcomes; patient satisfaction; user-perceived value

Introduction

The Promise and Policy-Based Promotion of Digital Health

For decades, digitization has been discussed as a promising answer to the various issues health care systems face today. A growing stream of research has revealed the positive effects of health information technology (HIT), specifically on the quality of health care providers, such as hospitals [1-4]. However, the adoption of promising HIT, such as electronic health records (EHRs), computerized clinical decision support systems, or telemedical tools, continues to lag expectations, particularly in Western Europe [5-7]. In 2017, for example, almost 50% of German and Austrian hospitals lacked an EHR system entirely and relied on paper-based documentation [8,9]. Policy makers reacted and introduced comprehensive financial incentives for HIT adoption. In the United States, the 2008 Health Information Technology for Economic and Clinical Health Act provided >US $28 billion to health care providers for adopting EHR systems [10]. Through this program, >80% of the hospitals have installed an EHR system since 2015 [11]. In 2020, German policy makers announced the establishment of a Hospital Future Fund (Krankenhauszukunftsfonds), which provides up to €4.3 billion (US $4.24 billion) to hospitals for investments in digital infrastructure and emergency capacities [12]. Policy makers and hospital decision makers clearly state attaining clinical outcome improvements as a fundamental goal of these measures and investments [13]. However, what appears to be an intuitive relationship reveals uncertainties, especially for the quality effects of nationwide HIT programs.

Existing Research on the Relationship Between Hospital Digitization and Quality

Several studies have examined the relationship between health care provider digitization and quality of care. Most research represents single-case studies in which the introduction of a certain type of HIT in a single institution has been investigated [1,2,14-21]. In most cases, HIT is associated with benefits to quality in terms of clinical outcomes or patient satisfaction. For example, HIT deployment supported timely pneumococcal vaccinations, improved guideline adherence related to antibiotic prescriptions, drove medication adherence among patients with diabetes, and improved patient satisfaction scores [14,22-24]. These studies represent a valuable indication of the effects of HIT on quality; however, most research covers the application of customized solutions in single organizations, and the potential existence of publication bias is frequently stated [1,21]. Hence, policy makers would benefit from additional information on whether these effects at the micro level also translate into nationwide hospital quality.

A smaller stream in the literature represents studies that analyze nationwide data on HIT adoption and quality [4,25-30]. For example, Jones et al [26] analyzed data from 2021 US hospitals and identified process quality improvements for patients with heart failure following the installation of basic EHRs. However, most of these studies exclusively focused on process quality in contrast to actual clinical outcomes or patient satisfaction. Furthermore, studies to date almost exclusively covered EHRs and the United States, which is likely because of the comparably early introduction of the Health Information Technology for Economic and Clinical Health Act. Policy makers in other geographical locations such as Europe could have reason to doubt the applicability of the results from these studies to their respective health care systems.

Research Contribution of This Study

In summary, this study assessed the effect of the availability and user-perceived value of multiple digital HIT tools on clinical outcomes and patient satisfaction across German hospitals. Thereby, the authors attempted to comprehensively address several previously identified research gaps. First, the goal was to add to the small stream of nationwide HIT studies by analyzing a geographical location outside the United States, namely Germany. Other European countries are likely to follow Germany's approach of introducing a HIT adoption program. Hence, information on whether the previously described findings from US-based studies translate to European health systems is of interest. Second, we attempted to capture quality in terms of patient satisfaction and actual clinical outcomes in contrast to process quality. The challenge related to clinical outcomes, such as mortality or surgical revisions, is that absolute values are prone to several confounding factors and impede comparability across hospitals. Relative outcome measures involving patient-specific risk adjustment resolve these issues and can be considered the gold standard in terms of clinical outcome metrics [31,32]. In addition, previous research gives reason to assume that the influence of hospital digitization on outcomes potentially differs between unpredictable emergency care and planned elective treatments [16,26,28]. Hence, this study captured quality in terms of risk-adjusted outcome measures and differentiated between elective and emergency care. Third, we attempted to capture the digital maturity of hospitals comprehensively in contrast to focusing on the mere adoption of single technologies. Today, several technologies beyond EHRs assist physicians and nurses from admission to discharge. In addition, research has identified human factors as the most significant barriers to the proper use of HIT [1,6,33]. Hence, it seemed reasonable to also consider the value of hospital digitization as perceived by everyday clinical users in the analysis. This study relied on a sophisticated digital maturity score capturing both the availability and user-perceived value of several HIT categories. Ultimately, the goal was to provide policy-relevant insights for the meaningful design of nationwide HIT incentivization programs. In addition, the study can also generate insights for

XSL•FO
RenderX

hospitals pursuing the promise of clinical outcome improvements via digitization.

Methods

Measures for Hospital Digitization

This study relied on the 2020 Healthcare IT Report for data on digital maturity of German hospitals. The Healthcare IT Report (IT Report Gesundheitswesen) is a comprehensive survey of German, Swiss, and Austrian hospitals that was first executed in 2002 and is continuing with varying core themes until today. The 2020 version of the report that surveyed 492 German physicians and nurses in 2017 went far beyond the mere adoption of technologies [8]. The report structured hospital digitization around the workflow from admission to discharge and captured comprehensive information on the adoption and user-perceived value related to >50 subtechnologies. Furthermore, the survey methodology has been constantly scientifically validated [34,35]. The report captured the digital maturity of hospitals across 10 domains composed of several technologies. A subscore was determined for both the adoption and user-perceived values of the underlying technologies for each domain. Table 1 presents an overview of all domains and the maximum attainable subscores. The domain admission, for example, was composed of 3 underlying technologies, namely occupancy control, collection of medical history and patient information, and steering of emergency patients (triage). The underlying technologies of the other domains can be found in the report publication and Multimedia Appendix 1 [8]. Summing the adoption subscores of all domains ultimately resulted in an overall maturity in terms of *HIT availability*. Similarly, averaging the user-perceived value across all domains resulted in an overall maturity in terms of *HIT value*. The report also covered EHR installations' availability and user-perceived value in a separate question. Hence, this is not captured in the scores illustrated in Table 1. EHR adoption was captured via a simple yes-or-no question, and the user-perceived value of EHRs was captured on a scale from 1 to 10.

Table 1. Overview of the subdomains of hospitals' digital maturity captured by the 2020 Healthcare IT Report (including maximum attainable scores).

HIT[a] domain	Maximum attainable scores	
	Adoption *HIT availability*	User-perceived value *HIT value*
Admission	30	10
Surgery preparation	30	10
Discharge	35	10
Clinical documentation	77	10
Order entry and reporting	42	10
Decision support	42	10
Patient safety	49	10
Supply functions	40	10
Interface functions	50	10
Telemedicine and monitoring	20	10
Total digital maturity	415	10

[a]HIT: health information technology.

Measures for Clinical Outcomes

This study considered risk-adjusted clinical outcome measures provided in the Qualitätssicherung mit Routinedaten (QSR) data set for both elective and emergency care. At the hospital level, these measures took the form of observed-over-expected (O/E) ratios, with the value of expected incidents being risk adjusted. This implied a worse-than-expected performance at values >1. In Germany, a reliable source of risk-adjusted clinical outcomes is the hospital data of Germany's largest statutory sickness fund, the Allgemeine Ortskrankenkasse (AOK). The internal research institute of the AOK, Wissenschaftliches Institut der AOK, is responsible for the central calculation of the risk-adjusted outcome data consolidated in the QSR database [36,37]. A consolidated O/E ratio for elective care was determined by averaging 4 risk-adjusted indicators related to hip replacement surgery due to coxarthrosis. The consolidated O/E ratio for emergency care was determined by averaging the 30-day risk-adjusted standardized mortality rates for patients who had stroke and myocardial infarction (heart attack), which are the most important and representative emergency care cases (Multimedia Appendix 2). Details on patient-based risk adjustments can be found in the indicator handbook for QSR data [38].

Measures for Patient Satisfaction

This study relied on data from the Patient Experience Questionnaire (PEQ) to assess the relationship between HIT adoption or user-perceived value and patient satisfaction in German hospitals. In Germany, Weisse Liste collects data on patient satisfaction in cooperation with the statutory insurance funds AOK, Barmer, and Kaufmännische Krankenkasse (KKH) [39]. Weisse Liste relies on the PEQ, covering 15 questions sent to patients 2 to 8 weeks after hospital discharge [40]. The questions cover various factors from personnel friendliness to admission and discharge process satisfaction. The complete

XSL•FO
RenderX

questionnaire is publicly available for review [41]. Experience is rated on a scale of 1 to 6, from best to worst.

Model Development and Estimation

Since several outcome metrics were of interest, a multiequation framework, similar to previous research in this field, was used [32,42]. A 3-equation multiple linear regression model was formulated with emergency care outcomes, elective care outcomes, and patient satisfaction as dependent variables:

$$\text{Emergency care outcomes} = \alpha_0 + \alpha_1(HIT_{adoption(t-1)}) + \alpha_2(HIT_{user\_value(t-1)}) + \alpha_3(\#\text{emergency cases}) + \alpha_4(\#\text{emergency cases}^2) + \alpha_5(\#\text{total cases}) + \alpha_6(\text{Private ownership}) + \alpha_7(\text{Teaching status}) + \alpha_8(\#\text{beds}) + \varepsilon \quad \textbf{(1)}$$

$$\text{Elective care outcomes} = \alpha_0 + \alpha_1(HIT_{adoption(t-1)}) + \alpha_2(HIT_{user\_value(t-1)}) + \alpha_3(\#\text{elective cases}) + \alpha_4(\#\text{total cases}) + \alpha_5(\text{Private ownership}) + \alpha_6(\text{Teaching status}) + \alpha_7(\#\text{beds}) + \varepsilon \quad \textbf{(2)}$$

$$\text{Patient satisfaction} = \alpha_0 + \alpha_1(HIT_{adoption(t-1)}) + \alpha_2(HIT_{user\_value(t-1)}) + \alpha_3(\#\text{total cases}) + \alpha_4(\text{Geography}) + \alpha_5(\text{Private ownership}) + \alpha_6(\text{Teaching status}) + \alpha_7(\#\text{beds}) + \varepsilon \quad \textbf{(3)}$$

In equations 1 and 2, which capture care outcomes, the dependent variables were represented by the consolidated O/E ratios based on the QSR database for emergency or elective care. Since this study deliberately differentiates between technology adoption and user-perceived value, both $HIT_{adoption}$ and HIT_{user_value} measures were included as independent variables. Here, the total digital maturity score of the Healthcare IT Report (Table 1) was considered. Importantly, a time lag of 1 year (t − 1) was introduced. This implies that full technology operability and user education require time after the initial installment [33,43,44]. In addition, several other independent variables were included to control for hospital-level effects. First, the underlying number of cases related to clinical outcomes was included to capture the potential volume effects directly linked to emergency or elective case volumes. Since several studies proved a positive relationship between case volume and quality, this is a critical confounding factor to control for [45,46]. Second, the total number of inpatient cases, #total cases, was included to capture the overall busyness of the respective hospitals [47]. The hospital ownership status was considered via the independent dummy variable Private ownership, which controls for potential organizational effects [48]. Similarly, teaching effects were included by another independent dummy variable: teaching status [48,49]. Finally, the number of hospital beds, #beds, was considered to capture the possible size effects of hospitals [42,50]. Equations 1 and 2 differ only in the inclusion of *#related cases*^{2} in equation 1. The existing research has indicated an inverse U–shaped relationship between case volumes and outcomes for emergency care [42]. By including both the regular and squared emergency case volumes, this effect could be captured if present in the underlying data.

In equation 3, the dependent variable patient satisfaction was represented by the average PEQ score across all questions. Independent variables remained mostly the same as in equation 2, except that potential volume effects were captured by considering only all inpatient cases, #total cases, of the hospital. Interestingly, controlling for geographic effects seems important when considering patient satisfaction, as research has shown higher average satisfaction in Eastern Germany [51].

The linear-in-parameter models were separately estimated by relying on ordinary least squares (OLS) regression analysis performed by the SAS JMP software. The HIT-related independent variables HIT *adoption (t − 1)* and HIT *user_value (t − 1)* were captured as 2017 values based on the 2020 Healthcare IT Report. Considering the previously mentioned 1-year lag of full operability and user education, all other variables were based on 2018 values. All dependent and independent variables capturing case volumes were measured in natural logarithms to account for unequal variation and ensure that the assumptions of OLS regressions were met. The robustness of all the estimated OLS regressions was tested to ensure that the coefficients were unbiased and close to the actual population values. This also included estimations of the regression models without control variables (Multimedia Appendix 3).

Ethics Approval

First, because the present study does not involve human biological material, this research was exempt from ethics approval in accordance with the 1961 German Drug Law (BGBI. I S. 533) and the 1994 Medical Devices Act (BGBI. I S. 1963). Second, according to statutes of the ethics committees of the institutions of all authors, data analyses processing data which does not relate to identifiable living individuals, which is the case for this study, do not require ethical approval in line with national guidelines (Regulation [European Union] 2016/679 [General Data Protection Regulation]). This study was conducted in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All methods were carried out in accordance with relevant guidelines and regulations.

Results

Overall Study Sample

The initial 492 responses of the 2020 Healthcare IT Report were cleaned for significantly incomplete questionnaires and contradictory answers from 2 respondents of the same hospital. Moreover, only hospitals for which information on either clinical outcomes or patient satisfaction was available were included. The final study sample comprised a total of 383 hospitals. The study sample characteristics and sample frames corresponding to all 1925 German hospitals in 2018 are presented in Table 2.

XSL•FO
RenderX

Table 2. Sample characteristics (N=383).

Sample characteristics	Values	Sample frame[a]
Structural		
Ownership, n (%)		
Private	88 (22.9)	37%
Nonprivate	295 (77)	63%
Teaching status, n (%)		
Teaching	199 (51.9)	51%
Nonteaching	184 (48)	49%
Beds, mean (SD)	319.72 (318.35)	258.8
Total inpatient cases, mean (SD)	12,861.6 (14,194.0)	10,290.2
Geography, n (%)		
Eastern states	72 (18.7)	18%
Western states	311 (81.2)	82%
Digital maturity		
HIT[b] adoption (2017; maximum 415), mean (SD)	183.1 (98.4)	N/A[c]
HIT user-value (2017; maximum 10), mean (SD)	6.3 (1.7)	N/A
EHR[d]a doption (2017), n (%)		
Yes	126 (32.8)	N/A
No	118 (30.8)	N/A
Unsure	6 (1.5)	N/A
No information	133 (34.7)	N/A
Clinical outcomes		
O/E[e] ratio		
Emergency care[f], mean (SD)	1.07 (0.48)	1.07
Elective care[g], mean (SD)	1.05 (1.23)	1.07
Patient satisfaction		
Overall PEQ score[h], mean (SD)	1.99 (0.31)	2.02

[a]Official 2018 hospital report [52], nationwide Qualitätssicherung mit Routinedaten data, and nationwide Patient Experience Questionnaire data.

[b]HIT: health information technology.

[c]N/A: not applicable.

[d]EHR: electronic health record.

[e]O/E: observed-over-expected.

[f]Value of 1 indicates as-expected outcome performance (ie, higher values indicate worse outcomes), n=267.

[g]Value of 1 indicates as-expected outcome performance (ie, higher values indicate worse outcomes), n=249.

[h]On a 1-to-6 scale from best to worst (ie, higher values indicate lower satisfaction), n=354.

The study sample of 383 hospitals represented approximately 20% of the 2018 sample frame (1925 hospitals). Table 2 indicates that the key descriptive metrics of the sample are very much in line with the broader German hospital landscape, that is, the sample frame, hence supporting sample representativeness. With an average of 319.72 (SD 318.35) beds and 12,861.6 (SD 14,194.0) inpatient cases, the sample seemed to capture slightly larger hospitals than the 2018 German average. Overall, patients appeared to be relatively satisfied, with an average patient satisfaction score of 1.99 (SD 0.31) on a scale from 1 to 6 (best to worst). Considering digital maturity, hospitals still failed to adopt several technologies, resulting in a comparably mediocre average score of 183.1 (SD 89.4) out of the maximum attainable 415. Finally, 50% (126/250) of the hospitals that provided information on EHRs indicated adoption, which is largely in line with the existing research [8,9].

HIT and EHR Effects on Clinical Outcomes

Several regressions were estimated based on the linear-in-parameter model in equations 1 and 2. The clinical

XSL•FO
RenderX

outcome dependent variables were represented by the O/E ratios of emergency care and elective care. For both the dependent variables separately, the effects of the overall HIT adoption and user-perceived value, adoption of EHRs, and user-perceived EHR value were captured in 3 separate regressions. The adoption and user-perceived value of EHRs were captured in 2 separate regressions to allow for a higher subsample size for the regression-covering availability. Hospitals with <5 cases as the basis for the clinical outcomes O/E ratio calculations were excluded owing to high outcome variability.

For emergency care, the adoption of either HIT (*P*=.54) or EHRs (*P*=.11) did not significantly affect outcomes. However, the clinical user-perceived value of overall HIT did have a significant positive effect on the outcomes, which was mirrored by a decrease in the O/E ratio (β=–.032; *P*=.04). On the other hand, EHR adoption or user-perceived value did not influence the emergency outcomes. When looking at the control variables, model I indeed revealed an inverted U–shaped relationship between emergency case volumes and outcomes. This was represented by a positive influence of regular volumes (β=–.559; *P*<.001) and a negative influence of squared volumes (β=.058; *P*=.004). In addition, hospitals with higher overall inpatient case volumes generated worse emergency care outcomes (β=.102; *P*=.04; Table 3).

Table 3. Ordinary least squares (OLS) estimates for linear-in-parameter regressions capturing the impact of health information technology (HIT) and electronic health record (EHR) on emergency care clinical outcomes.

Dependent variable	ln[a] (O/E ratio emergency care)[b]					
	Model I (HIT)		Model II (EHR adoption)		Model III (EHR user value)	
	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
Intercept	.557 (0.506)	.27	.479 (0.65)	.46	−.563 (1.007)	.58
HIT adoption[c]	−.001 (0.001)	.54	N/A[d]	N/A	N/A	N/A
HIT user-value[e]	−.032 (0.016)	.04	N/A	N/A	N/A	N/A
EHR adoption[f]	N/A	N/A	.053 (0.034)	.11	N/A	N/A
EHR user-value[e]	N/A	N/A	N/A	N/A	−.018 (0.022)	.42
#beds						
<150	.037 (0.087)	.67	−.082 (0.108)	.45	−.087 (0.147)	.55
150-300	.027 (0.046)	.56	−.002 (0.061)	.97	.033 (0.077)	.66
301-600	−.035 (0.049)	.47	−.002 (0.064)	.97	.065 (0.082)	.43
>600	−.028 (0.086)	.74	.087 (0.114)	.45	−.01 (0.14)	.94
ln(#total cases)	.102 (0.051)	.04	.026 (0.068)	.70	.134 (0.086)	.12
ln(#emergency cases)	−.559 (0.167)	<.001	−.288 (0.214)	.18	−.203 (0.286)	.48
ln(#emergency cases)^2[g]	.058 (0.019)	.004	.027 (0.025)	.28	.018 (0.033)	.58
Teaching(yes)	−.036 (0.032)	.26	−.015 (0.214)	.72	−.022 (0.054)	.68
Private(yes)	.007 (0.037)	.83	.017 (0.046)	.72	.059 (0.060)	.33
Subsample size	261		174		82	
R^2	0.098		0.047		0.117	
F value	2.727	.003	0.905	.52	1.061	.40

[a]ln implies natural logarithm.

[b]O/E (observed-over-expected) ratio implies better performance with lower values.

[c]On a 0-to-415 scale from worst to best.

[d]N/A: not applicable.

[e]On a 1-to-10 scale from worst to best.

[f]Adoption of EHR.

[g]Tests for an inverse U–shaped relationship between case volumes and outcomes for emergency care.

Following an exploratory approach, additional regressions were estimated to better understand whether any subdomain (Table 1) had a prominent effect on the identified significant relationship at the overall HIT level. All HIT subdomains were separately regressed against emergency care outcomes. The admission subdomain generated statistically significant results (Multimedia Appendix 3). The user-perceived value of admission HIT had significant effects on emergency care outcomes (β=–.023; *P*=.07). An inverted U–shaped relationship between emergency case volumes and outcomes was also identified in this regression.

XSL•FO
RenderX

For elective care, the analysis of the effects of HIT and EHR on outcomes revealed not only similarities to but also differences from emergency care. In line with the emergency care results, the adoption of either HIT (*P*=.54) or EHRs (*P*=.84) did not significantly affect the elective care outcomes (Table 4). Although a higher user-perceived value of overall HIT did not affect outcomes either, clinical users' satisfaction with EHRs had a significant effect, as shown in model III (β=–.138; *P*=.008). Of the control variables included in the regression, the number of elective care cases showed significant effects in models I and III, indicating a positive relationship between case volume and outcome quality in the elective care field. Interestingly, hospital size was a significant factor, with smaller hospitals performing better than larger ones.

Table 4. Ordinary least squares (OLS) estimates for linear-in-parameter regressions capturing the impact of health information technology (HIT) and electronic health record (EHR) on elective care clinical outcomes.

Dependent variable	ln[a] (O/E ratio elective care)[b]					
	Model I (HIT)		Model II (EHR adoption)		Model III (EHR user value)	
	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
Intercept	3.005 (1.495)	.046	3.4 (2.07)	.10	2.973 (2.741)	.28
HIT adoption[c]	.001 (0.001)	.54	N/A[d]	N/A	N/A	N/A
HIT user-value[e]	.038 (0.034)	.26	N/A	N/A	N/A	N/A
EHR adoption[f]	N/A	N/A	−.013 (0.063)	.84	N/A	N/A
EHR user-value[e]	N/A	N/A	N/A	N/A	−.138 (0.05)	.008
#beds						
<150	−.529 (0.199)	.009	−.716 (0.229)	.002	−.42 (0.329)	.21
150-300	−.075 (0.099)	.45	−.071 (0.125)	.57	−.089 (0.189)	.64
301-600	.415 (0.108)	<.001	.5 (0.123)	<.001	.393 (0.181)	.03
>600	.189 (0.191)	.32	.287 (0.228)	.21	.116 (0.329)	.73
ln(#total cases)	−.239 (0.157)	.13	−.266 (0.211)	.21	−.132 (0.283)	.64
ln(#elective cases)	−.324 (0.064)	<.001	−.269 (0.079)	.001	−.278 (0.115)	.02
Teaching(yes)	−.059 (0.065)	.36	−.114 (0.078)	.15	−.03 (0.12)	.81
Private(yes)	−.054 (0.072)	.45	−.062 (0.082)	.45	−.173 (0.119)	.15
Subsample size	184		118		59	
R^2	0.237		0.265		0.375	
F value	5.997	<.001	4.915	<.001	3.753	.002

[a]ln implies natural logarithm.

[b]O/E (observed-over-expected) ratio implies better performance with lower values.

[c]On a 0-to-415 scale from worst to best.

[d]N/A: not applicable.

[e]On a 0-to-10 scale from worst to best.

[f]Adoption of EHR.

HIT and EHR Effects on Patient Satisfaction

Several regressions were estimated based on the linear-in-parameter model in equation 3, where the PEQ scores represented patient satisfaction. To ensure representativeness, the estimation excluded patient satisfaction PEQ scores based on <20 respondents per hospital. An overview of all the estimated models and their respective subsample sizes are presented in Table 5. In line with the effects on clinical outcomes, mere technology adoption did not significantly affect the overall patient satisfaction, that is, the overall PEQ score (*P*=.41). However, the user-perceived value of HIT did moderately affect patient satisfaction (β=–.009; *P*=.01), as indicated by model I of Table 5. Models II and III, which examined the impact of EHRs, did not identify any significant EHR-related effects. Across all models, higher bed numbers negatively affected satisfaction. In addition, in line with previous research, patients visiting hospitals in the Eastern German federal states were, on average, more satisfied.

An additional exploratory regression explicitly examining the relationship between admission HIT and admission satisfaction was estimated (Multimedia Appendix 3). Interestingly, the adoption of more technology in the admission implied a decrease in patients' admission satisfaction (β=.002; *P*=.03). However, by contrast, the user-perceived value of admission HIT and patients' admission satisfaction were significantly positively related (β=–.009; *P*=.02).

XSL•FO
RenderX

Table 5. Ordinary least squares (OLS) estimates for linear-in-parameter regressions capturing the impact of health information technology (HIT) and electronic health record (EHR) on patient satisfaction.

Dependent variable	ln[a] (overall PEQ score)[b]					
	Model I (HIT)		Model II (EHR adoption)		Model III (EHR user value)	
	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
Intercept	.694 (0.098)	<.001	.658 (0.124)	<.001	.822 (0.172)	<.001
HIT adoption[c]	.001 (0.001)	.41	N/A[d]	N/A	N/A	N/A
HIT user-value[e]	−.009 (0.004)	.01	N/A	N/A	N/A	N/A
EHR adoption[f]	N/A	N/A	.001 (0.007)	.93	N/A	N/A
EHR user-value[e]	N/A	N/A	N/A	N/A	−.006 (0.005)	.24
#beds						
<150	−.097 (0.017)	<.001	−.087 (0.02)	<.001	−.122 (0.034)	<.001
150-300	−.008 (0.009)	.42	−.019 (0.013)	.14	−.014 (0.018)	.44
301-600	.032 (0.011)	.005	.026 (0.014)	.07	.033 (0.021)	.12
>600	.072 (0.018)	<.001	.08 (0.022)	<.001	.104 (0.031)	.001
ln(#total cases)	.002 (0.01)	.82	.002 (0.013)	.90	−.013 (0.019)	.50
Geography (east)[g]	−.031 (0.007)	<.001	−.029 (0.01)	.005	−.029 (0.015)	.05
Teaching(yes)	.003 (0.008)	.67	.002 (0.009)	.83	.001 (0.014)	.97
Private(yes)	.003 (0.007)	.69	−.001 (0.009)	.95	−.016 (0.015)	.30
Subsample size	310		203		93	
R^2	0.322		0.259		0.35	
F value	15.824	<.001	8.516	<.001	5.669	<.001

[a]ln implies natural logarithm.

[b]PEQ (Patient Experience Questionnaire) on a 1-to-6 scale from best to worst.

[c]On a 0-to-415 scale from worst to best.

[d]N/A: not applicable.

[e]On a 1-to-10 scale from worst to best.

[f]Adoption of EHR.

[g]Effect of a hospital being located in an Eastern German state.

Discussion

Principal Findings

The analysis of our data set suggested that the adoption of HIT alone does not have a significant effect on either clinical outcomes or patient satisfaction. However, the degree of HIT value as perceived by hospital users did affect both clinical outcomes and patient satisfaction. Hospital digitization appears to be about quality instead of mere quantity.

In the elective care field, higher user-friendliness of EHRs for hospital staff appears to result in better outcomes, whereas digitization in other areas appears to matter less. By contrast, emergency care outcomes benefit from user-friendly overall digitization. Admission HIT, deemed valuable by physicians and nurses, is the strongest driver in this field. These identified differences appear to confirm practical experiences and logic. Elective care, represented by elective hip replacement surgery, relies on a comprehensive compilation of patients' medical histories to support patient-specific diagnosis and therapy decisions. A user-friendly EHR is at the center of this collection and preparation process, representing a single source of information for each patient. The results of this study suggested that physicians and nurses who can rely on easy-to-access EHRs also generated better elective surgery outcomes. When it comes to emergency care, outcomes were positively affected by higher user-perceived value of overall hospital digitization. This overall effect was especially driven by user-friendly admission technology. This implied that rapid decision-making shortly after the incident most likely determines the outcomes. Taking an exclusive look at the roughly 270,000 annual stroke cases in Germany and their average 6.8% 30-day mortality, our results indicated that clinical user-perceived HIT value has the potential to reduce deaths [53,54]. Specifically, a 1-point improvement in user-perceived HIT value may result in a 3.15% improvement in mortality, which would ultimately translate to approximately 580 avoided deaths within 30 days per year. At this point, a brief look at the control variables of equations 1 and 2 also revealed interesting insights. Clinical outcomes for elective care, that is, hip replacements, benefit from higher procedure

XSL•FO
RenderX

volumes. This finding was very much in line with the existing research [45,46,55-57]. The same applied to emergency care outcomes, at least to some extent. Here, an inverted U–shaped relationship between case volume and outcomes was identified, implying that emergency departments and their care teams are potentially overburdened at some point. Our analysis also suggested that smaller hospitals tend to perform better on elective care outcomes.

A similar picture emerged when considering patient satisfaction. Our analysis suggests that the mere adoption of overall HIT does not have significant effects on patient satisfaction. Digitization that appears as value adding from a hospital user perspective has a moderate positive impact on patients' experiences. There are several potential reasons for this. First, well-designed digital tools covering admission, clinical documentation, order entry, discharge, or even catering management help physicians and nurses effectively execute care. This could mean shorter waiting times, better informed treating doctors, and smooth patient discharge processes, all of which likely positively influence patient satisfaction. Second, however, one can also assume hospital staff that are content with their everyday work environment, which includes HIT, to approach patients in a less stressed and more personal manner. This can also ultimately influence patient satisfaction through interpersonal mechanisms. Finally, an interesting phenomenon related to patient satisfaction with admission was revealed. Results showed that higher levels of admission digitization, for example, via tools for patient education and collection of medical history, resulted in lower admission satisfaction. Research has repeatedly warned of the undesired effects that digitization has on the patient-physician relationship, considering the very personal elements of health care [58]. Fortunately, results also showed that a higher degree of hospital user-perceived value of admission digitization can work against this effect and improve patient satisfaction. Hence, in cases such as admission, mere digitization could even be detrimental to patient satisfaction if not executed in a user-friendly manner for the hospital staff. Looking at the control variables, patients appeared to be less satisfied in larger hospitals. It can only be hypothesized that this might be related to a more personal care approach in smaller institutions.

The results of this study also provided valuable insights for both policy makers and hospitals. The results of this study suggested that an exclusive focus on driving HIT adoption will likely not be sufficient to achieve improvements in care quality. For example, the 2020 German Hospital Future Act also covers the costs of training and user education. Furthermore, it introduced a mandatory yearly digital maturity survey among all hospitals that applied for funding. Considering the limited data on the digital maturity of hospitals in Germany, this represents a desirable approach. However, of the 209 final questions included in the questionnaire, only 6 questions capture aspects of user satisfaction [59]. The results of this study advise policy makers to focus on the value of digitization as perceived by everyday clinical users instead of incentivizing the mere adoption of advanced digital tools. From the hospital's perspective, much of what was previously discussed applies. First, hospital decision makers are advised to consider the views of the ultimate clinical users in their procurement decisions. Second, hospitals should not consider digitization projects completed after mere adoption but emphasize ongoing user training. Third, digitization strategies of health care providers benefit from differentiated approaches based on the respective area of care.

Limitations

The results of this study are subject to several potential limitations. First, this study had to rely on a somewhat limited data set on hospital digitization in Germany. Although the 2020 Healthcare IT Report captured initial answers for almost 500 hospitals, the exclusion of significantly incomplete questionnaires and contradictory answers of 2 respondents from the same hospital resulted in a final overall sample of 383 hospitals. Moreover, responses related to the adoption and user-perceived value of EHRs were even more limited. This ultimately resulted in very differently sized subsamples, potentially impairing the comparability of the estimated regressions. These limitations strongly reveal the need for a structured and periodic digital maturity assessment among hospitals. Fortunately, as previously described, the 2020 Hospital Future Act introduced this assessment in Germany. However, the collected data must also be available to the research community.

Second, the study relied on patient satisfaction data from the PEQ, which collects answers on a Likert-type scale. There is some controversy about whether Likert-type responses can be averaged when used in statistical analyses with no unanimous results. However, since the PEQ data used in this study were normally distributed, we followed the standard approach of using average values [60].

Third, clinical outcome data were provided by the QSR database of the statutory insurer AOK. Hence, only the clinical outcome data for patients insured by the AOK were captured. However, as the AOK is by far the largest statutory insurer in Germany, with approximately 35% market share, the representativeness of the data can be assumed. Moreover, this study focused on the outcomes for two emergency care cases, namely stroke and myocardial infarction, and one elective care indication, namely hip replacement. Although stroke and myocardial infarction indeed represent the most essential emergency care indications in Germany, the results of this study might not be fully applicable to other areas of elective care.

Besides these primary variables of interest, this study also relied on several control variables when estimating OLS regressions. Looking at the R-square values of regressions, including care outcomes as the dependent variable, a significant share of variance was seemingly not explained by the included independent variables. Hence, risk-adjusted care outcomes appear to be influenced by a much wider variety of factors that were not captured in this study. This seems reasonable since care represents a highly personalized process that is also related to activities outside the hospital. Nevertheless, several additional mediator or moderator variables that could potentially impact the dependent variables of interest were not included in the regressions. Hence, the results of this study can only suggest a causal effect that the availability and user-perceived value of HIT have on care quality. On a similar note, reverse causality

XSL•FO
RenderX

cannot be entirely ruled out. This would imply that hospitals which perform better on outcomes and patient satisfaction also invest more in HIT. However, ultimately, this study included the most prominent control variables used in research when assessing the effects on clinical outcomes and satisfaction to counter these potential limitations [2,20,21,47].

Conclusions

This study examined the effects of digitization on clinical outcomes and patient satisfaction across German hospitals. The analysis of our data set suggested that the adoption of HIT alone does not significantly influence either outcomes or patient satisfaction, whereas the value of these technologies as perceived by physicians and nurses does positively influence patient outcomes and satisfaction. However, the results also implied that it is essential to differentiate between care indications and HIT subtypes. Whereas emergency care outcomes significantly benefit from user-friendly admission HIT, elective care outcomes significantly benefit from user-friendly EHR installations. Besides clinical outcomes, the user-perceived value of HIT significantly influenced patient satisfaction. Hospital staff working with HIT that is value adding from their perspective can treat patients in a manner that enhances the overall patient satisfaction. Besides HIT-related effects, the results suggest a positive relationship between case volume and outcomes in the elective care field. For emergency care, an inverse U–shaped relationship between the volume and outcomes was identified. Furthermore, elective care outcomes and patient satisfaction benefit from smaller hospital sizes in terms of bed numbers. Policy makers attempting to improve care quality via HIT are advised to focus on the value of digitization as perceived by everyday clinical users instead of incentivizing, capturing, or even imposing the mere adoption of digital tools. Hospitals are well advised to consider the views of the ultimate clinical users in their procurement decisions and invest in continuous training. In conclusion, hospital digitization can improve both clinical outcomes and patient satisfaction but only if deemed to be value adding by the physicians and nurses who rely on it every day.

Acknowledgments

The authors thank Christian Günster and Andreas Klöss of the Scientific Institute of the Allgemeine Ortskrankenkasse (AOK; Wissenschaftliches Institut der AOK) and the Federal Association of the AOK for the provision of data and their conceptual feedback. Similarly, the authors thank Hannah Wehling and the Weisse Liste GmbH for providing Patient Experience Questionnaire data.

Conflicts of Interest

For transparency purposes, the authors declare that apart from his research positions, CP currently works at Stryker GmbH & Co. KG as Director Market Access, Health Economics, and Government Affairs.

Multimedia Appendix 1
Overview of the 2020 health care IT report scoring model.
[PDF File (Adobe PDF File), 103 KB - jmir_v24i11e40124_app1.pdf]

Multimedia Appendix 2
Overview of the single outcome indicators considered for the calculation of consolidated overexpected values for elective and emergency care.
[PDF File (Adobe PDF File), 69 KB - jmir_v24i11e40124_app2.pdf]

Multimedia Appendix 3
Robustness tests and statistical analysis of additional exploratory ordinary least squares regressions.
[PDF File (Adobe PDF File), 176 KB - jmir_v24i11e40124_app3.pdf]

References

1. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
2. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
3. Kruse CS, Beane A. Health information technology continues to show positive effect on medical outcomes: systematic review. J Med Internet Res 2018 Feb 05;20(2):e41 [FREE Full text] [doi: 10.2196/jmir.8793] [Medline: 29402759]
4. McCullough JS, Casey M, Moscovice I, Prasad S. The effect of health information technology on quality in U.S. hospitals. Health Aff (Millwood) 2010 Apr;29(4):647-654. [doi: 10.1377/hlthaff.2010.0155] [Medline: 20368594]

XSL•FO
RenderX

5. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
6. Sun TQ, Medaglia R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Gov Inf Q 2019 Apr;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
7. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020 Feb 6;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
8. Hübner U, Esdar M, Hüsers J, Liebe JD, Naumann L, Thye J, et al. IT-Report Gesundheitswesen: wie reif ist die Gesundheits-IT aus Anwenderperspektive. University of Applied Sciences. 2020. URL: https://www.hs-osnabrueck.de/fileadmin/HSOS/Homepages/IT-Report_Gesundheitswesen/IT-Report_Gesundheitswesen_2020_final.pdf [accessed 2022-01-19]
9. Stephani V, Busse R, Geissler A. Benchmarking der krankenhaus-IT: Deutschland im internationalen vergleich. In: Klauber J, Geraedts M, Friedrich J, Wasem J, editors. Krankenhaus-Report 2019: Das digitale Krankenhaus. Berlin, Germany: Springer; 2019:17-32.
10. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 04;362(5):382-385. [doi: 10.1056/NEJMp0912825] [Medline: 20042745]
11. Adler-Milstein J, Jha AK. HITECH act drove large gains in hospital electronic health record adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]
12. Krankenhauszukunftsgesetz für die Digitalisierung von Krankenhäusern. Bundesministerium für Gesundheit. 2020. URL: https://www.bundesgesundheitsministerium.de/krankenhauszukunftsgesetz.html [accessed 2021-09-14]
13. 98,1 Prozent aller Kliniken wollen KHZG-Antrag stellen. Online KMA. 2021. URL: https://www.kma-online.de/aktuelles/it-digital-health/detail/981-prozent-aller-kliniken-wollen-khzg-antrag-stellen-a-46415 [accessed 2022-01-19]
14. Ali SM, Giordano R, Lakhani S, Walker DM. A review of randomized controlled trials of medical record powered clinical decision support system to improve quality of diabetes care. Int J Med Inform 2016 Mar;87:91-100. [doi: 10.1016/j.ijmedinf.2015.12.017] [Medline: 26806716]
15. Atasoy H, Greenwood BN, McCullough JS. The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. Annu Rev Public Health 2019 Apr 01;40:487-500. [doi: 10.1146/annurev-publhealth-040218-044206] [Medline: 30566385]
16. Bennett P, Hardiker NR. The use of computerized clinical decision support systems in emergency care: a substantive review of the literature. J Am Med Inform Assoc 2017 May 01;24(3):655-668 [FREE Full text] [doi: 10.1093/jamia/ocw151] [Medline: 28031285]
17. Blum D, Raj SX, Oberholzer R, Riphagen II, Strasser F, Kaasa S, EURO IMPACT, European Intersectorial Multidisciplinary Palliative Care Research Training. Computer-based clinical decision support systems and patient-reported outcomes: a systematic review. Patient 2015 Oct;8(5):397-409. [doi: 10.1007/s40271-014-0100-1] [Medline: 25432150]
18. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 03;157(1):29-43 [FREE Full text] [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
19. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. Eur J Public Health 2016 Feb;26(1):60-64. [doi: 10.1093/eurpub/ckv122] [Medline: 26136462]
20. Jaspers MW, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011 May 01;18(3):327-334 [FREE Full text] [doi: 10.1136/amiajnl-2011-000094] [Medline: 21422100]
21. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 07;160(1):48-54 [FREE Full text] [doi: 10.7326/M13-1531] [Medline: 24573664]
22. Dexter PR, Perkins SM, Maharry KS, Jones K, McDonald CJ. Inpatient computer-based standing orders vs physician reminders to increase influenza and pneumococcal vaccination rates: a randomized trial. JAMA 2004 Nov 17;292(19):2366-2371. [doi: 10.1001/jama.292.19.2366] [Medline: 15547164]
23. Evans RS, Pestotnik SL, Classen DC, Burke JP. Evaluation of a computer-assisted antibiotic-dose monitor. Ann Pharmacother 1999 Oct;33(10):1026-1031. [doi: 10.1345/aph.18391] [Medline: 10534212]
24. Werder M. Health information technology: a key ingredient of the patient experience. Patient Exp J 2015 Apr 29;2(1):143-147. [doi: 10.35680/2372-0247.1071]
25. Appari A, Eric Johnson M, Anthony DL. Meaningful use of electronic health record systems and process quality of care: evidence from a panel data analysis of U.S. acute-care hospitals. Health Serv Res 2013 Apr;48(2 Pt 1):354-375 [FREE Full text] [doi: 10.1111/j.1475-6773.2012.01448.x] [Medline: 22816527]
26. Jones SS, Adams JL, Schneider EC, Ringel JS, McGlynn EA. Electronic health record adoption and quality improvement in US hospitals. Am J Manag Care 2010 Dec;16(12 Suppl HIT):SP64-SP71 [FREE Full text] [Medline: 21314225]
27. Linder JA, Ma J, Bates DW, Middleton B, Stafford RS. Electronic health record use and the quality of ambulatory care in the United States. Arch Intern Med 2007 Jul 09;167(13):1400-1405. [doi: 10.1001/archinte.167.13.1400] [Medline: 17620534]

XSL•FO
RenderX

28. McCullough JS, Parente ST, Town R. Health information technology and patient outcomes: the role of information and labor coordination. RAND J Econ 2016 Jan 08;47(1):207-236. [doi: 10.1111/1756-2171.12124]
29. Romano MJ, Stafford RS. Electronic health records and clinical decision support systems: impact on national ambulatory care quality. Arch Intern Med 2011 May 23;171(10):897-903 [FREE Full text] [doi: 10.1001/archinternmed.2010.527] [Medline: 21263077]
30. Zhou L, Soran CS, Jenter CA, Volk LA, Orav EJ, Bates DW, et al. The relationship between electronic health record use and quality of care over time. J Am Med Inform Assoc 2009;16(4):457-464 [FREE Full text] [doi: 10.1197/jamia.M3128] [Medline: 19390094]
31. Iezzoni LI. Using risk-adjusted outcomes to assess clinical practice: an overview of issues pertaining to risk adjustment. Ann Thorac Surg 1994 Dec;58(6):1822-1826. [doi: 10.1016/0003-4975(94)91721-3] [Medline: 7979776]
32. Restuccia JD, Cohen AB, Horwitt JN, Shwartz M. Hospital implementation of health information technology and quality of care: are they related? BMC Med Inform Decis Mak 2012 Sep 27;12:109 [FREE Full text] [doi: 10.1186/1472-6947-12-109] [Medline: 23016699]
33. Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to electronic health record adoption: a systematic literature review. J Med Syst 2016 Dec;40(12):252 [FREE Full text] [doi: 10.1007/s10916-016-0628-9] [Medline: 27714560]
34. Liebe JD, Hübner U. Developing and Trialling an independent, scalable and repeatable IT-benchmarking procedure for healthcare organisations. Methods Inf Med 2013;52(4):360-369. [doi: 10.3414/ME12-02-0016] [Medline: 23877646]
35. Liebe JD, Hübner UH, Buddrus U, Jahn F, Kümmel K. Reifegradmodelle: Von Frustration zu Innovation. F&W : Führen und Wirtschaften im Krankenhaus 2020;37(9):860-863.
36. Qualitätssicheurng der stationären Versorgung mit Routinedaten (QSR): Abschlussbericht. Wissenschaftliches Institut der AOK. 2007. URL: https://www.qualitaetssicherung-mit-routinedaten.de/imperia/md/qsr/publikationen/wido_kra_qsr-abschlussbericht_0407.pdf [accessed 2021-12-13]
37. Günster C, Jeschke E, Malzahn J, Schillinger G. Qualitätssicherung mit Routinedaten (QSR). In: Kray R, Koch C, Sawicki PT, editors. Qualität in der Medizin dynamisch denken: Versorgung - Forschung - Markt. Wiesbaden, Germany: Springer; 2013:111-129.
38. Indikatorenhandbuch Verfahrensjahr 2018. Wissenschaftliches Institut der AOK. 2018. URL: https://www.qualitaetssicherung-mit-routinedaten.de/imperia/md/qsr/methoden/indikatorenhandbuch_2018_final.pdf [accessed 2021-12-16]
39. Fischer S. Project “Weisse Liste”: a German best practice example for online provider ratings in health care. In: Gurtner S, Soyez K, editors. Challenges and Opportunities in Health Care Management. Cham, Switzerland: Springer; 2015:339-346.
40. Gehrlach C, Altenhöner T, Schwappach D. Der Patients' Experience Questionnaire: Patientenerfahrungen Vergleichbar Machen. Gütersloh, Germany: Bertelsmann Stiftung; 2010.
41. Liste W. Patients Experience Questionnaire (PEQ) V 1.1. www.weisse-liste.de. 2016. URL: https://www.weisse-liste.de/krankenhaus/static/files/Fragebogen-PEQ_Krankenhausaufenthalt-1.pdf [accessed 2021-12-01]
42. Choi JH, Park I, Jung I, Dey A. Complementary effect of patient volume and quality of care on hospital cost efficiency. Health Care Manag Sci 2017 Jun 4;20(2):221-231. [doi: 10.1007/s10729-015-9348-9] [Medline: 26637297]
43. Blumenthal D, Glaser JP. Information technology comes to medicine. N Engl J Med 2007 Jun 14;356(24):2527-2534. [doi: 10.1056/NEJMhpr066212] [Medline: 17568035]
44. Lin SC, Jha AK, Adler-Milstein J. Electronic health records associated with lower hospital mortality after systems have time to mature. Health Aff (Millwood) 2018 Jul;37(7):1128-1135. [doi: 10.1377/hlthaff.2017.1658] [Medline: 29985687]
45. Marcin JP, Romano PS. Impact of between-hospital volume and within-hospital volume on mortality and readmission rates for trauma patients in California. Crit Care Med 2004 Jul;32(7):1477-1483. [doi: 10.1097/01.ccm.0000127781.08985.03] [Medline: 15241091]
46. Wright JD, Chen L, Hou JY, Burke WM, Tergas AI, Ananth CV, et al. Association of hospital volume and quality of care with survival for ovarian cancer. Obstet Gynecol 2017 Sep;130(3):545-553 [FREE Full text] [doi: 10.1097/AOG.0000000000002164] [Medline: 28796677]
47. Jarvis B, Johnson T, Butler P, O'Shaughnessy K, Fullam F, Tran L, et al. Assessing the impact of electronic health records as an enabler of hospital quality and patient satisfaction. Acad Med 2013 Oct;88(10):1471-1477. [doi: 10.1097/ACM.0b013e3182a36cab] [Medline: 23969372]
48. Eggleston K, Shen YC, Lau J, Schmid CH, Chan J. Hospital ownership and quality of care: what explains the different results in the literature? Health Econ 2008 Dec;17(12):1345-1362. [doi: 10.1002/hec.1333] [Medline: 18186547]
49. Ayanian JZ, Weissman JS. Teaching hospitals and quality of care: a review of the literature. Milbank Q 2002;80(3):569-56v [FREE Full text] [doi: 10.1111/1468-0009.00023] [Medline: 12233250]
50. Menon NM, Lee B, Eldenburg L. Productivity of information systems in the healthcare industry. Inf Syst Res 2000 Mar;11(1):83-92. [doi: 10.1287/isre.11.1.83.11784]
51. Kraska RA, Weigand M, Geraedts M. Associations between hospital characteristics and patient satisfaction in Germany. Health Expect 2017 Aug;20(4):593-600 [FREE Full text] [doi: 10.1111/hex.12485] [Medline: 27447595]
52. Grunddaten der Krankenhäuser. Statistisches Bundesamt (Destatis). 2018. URL: https://tinyurl.com/2p88dh39 [accessed 2021-12-20]

XSL•FO
RenderX

53. Heuschmann PU, Busse O, Wagner M, Endres M, Villringer A, Röther J, et al. Schlaganfallhäufigkeit und Versorgung von Schlaganfallpatienten in Deutschland. Akt Neurol 2010 Oct 5;37(07):333-340. [doi: 10.1055/s-0030-1248611]
54. Stahmeyer JT, Stubenrauch S, Geyer S, Weissenborn K, Eberhard S. Häufigkeit und zeitpunkt von rezidiven nach inzidentem schlaganfall: eine analyse auf basis von GKV-routinedaten. Dtsch Arztebl Int 2019 Oct 18;116(42):711-717 [FREE Full text] [doi: 10.3238/arztebl.2019.0711] [Medline: 31711561]
55. Bozic KJ, Maselli J, Pekow PS, Lindenauer PK, Vail TP, Auerbach AD. The influence of procedure volumes and standardization of care on quality and efficiency in total joint replacement surgery. J Bone Joint Surg Am 2010 Nov 17;92(16):2643-2652. [doi: 10.2106/JBJS.I.01477] [Medline: 21084575]
56. Pross C, Berger E, Siegel M, Geissler A, Busse R. Stroke units, certification, and outcomes in German hospitals: a longitudinal study of patient-based 30-day mortality for 2006-2014. BMC Health Serv Res 2018 Nov 22;18(1):880 [FREE Full text] [doi: 10.1186/s12913-018-3664-y] [Medline: 30466414]
57. Vogel J, Polin K, Pross C, Geissler A. Implikationen von mindestmengen und zertifizierungsvorgaben: auswirkungen verschiedener vorgaben auf den deutschen krankenhaussektor. In: Dormann F, Klauber J, Kuhlen R, editors. Qualitätsmonitor 2019. Berlin, Germany: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2019.
58. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
59. Geissler A, Bleibtreu E, DigitalRadar Krankenhaus Konsortium. Instrument zur Evaluierung des Reifegrades der Krankenhäuser hinsichtlich der Digitalisierung. Schriftenreihe in Health Economics, Management and Policy. 2021 Jan. URL: https://tinyurl.com/3h7sycda [accessed 2022-02-15]
60. Sullivan GM, Artino Jr AR. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013 Dec;5(4):541-542 [FREE Full text] [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]

Abbreviations

AOK: Allgemeine Ortskrankenkasse
EHR: electronic health record
HIT: health information technology
KKH: Kaufmännische Krankenkasse
O/E: observed-over-expected
OLS: ordinary least squares
PEQ: Patient Experience Questionnaire
QSR: Qualitätssicherung mit Routinedaten

Edited by G Eysenbach; submitted 08.06.22; peer-reviewed by S Fetzer, E Loh, K Kroeker; comments to author 29.07.22; revised version received 30.08.22; accepted 23.09.22; published 10.11.22.

Please cite as:
von Wedel P, Hagist C, Liebe JD, Esdar M, Hübner U, Pross C
Effects of Hospital Digitization on Clinical Outcomes and Patient Satisfaction: Nationwide Multiple Regression Analysis Across German Hospitals
J Med Internet Res 2022;24(11):e40124
URL: https://www.jmir.org/2022/11/e40124
doi:10.2196/40124
PMID:36355423

©Philip von Wedel, Christian Hagist, Jan-David Liebe, Moritz Esdar, Ursula Hübner, Christoph Pross. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Relationship Between the Big Five Personality Traits and the Theory of Planned Behavior in Using Mindfulness Mobile Apps: Cross-sectional Survey

Sunghak Kim[1,2], PhD; Jin Young Park[1,2,3], MD, PhD; Kyungmi Chung[1,2,3], PhD

[1]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Yonsei University Health System, Seoul, Republic of Korea

[2]Department of Psychiatry, Yongin Severance Hospital, Yonsei University College of Medicine, Yonsei University Health System, Yongin, Republic of Korea

[3]Center for Digital Health, Yongin Severance Hospital, Yonsei University College of Medicine, Yonsei University Health System, Yongin, Republic of Korea

Corresponding Author:
Kyungmi Chung, PhD
Institute of Behavioral Science in Medicine
Yonsei University College of Medicine
Yonsei University Health System
50-1, Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 10 4234 3442
Email: chungkyungmi@yuhs.ac

Abstract

Background: Mindfulness has emerged as a promising approach toward improving mental health. Interest in mindfulness mobile app services has also increased in recent years. Understanding the determinants of mindfulness behavior is essential to predict people's utilization of mindfulness mobile apps and beneficial for developing and implementing relevant intervention strategies. Nevertheless, little has been done to determine the predictors of mindfulness behavior.

Objective: This study investigates the association between the Big Five personality traits and the Theory of Planned Behavior (TPB) variables in the context of using mindfulness mobile apps to explore the potential indirect effects of conscientiousness and neuroticism on people's behavioral intention for mindfulness, mediated by their attitude toward mindfulness, subjective norm about mindfulness, and perceived behavior control over mindfulness.

Methods: The authors conducted an online, cross-sectional survey in December 2021. Structural equation modeling was conducted to evaluate the overall model fit and test possible linkages among conscientiousness, neuroticism, attitude toward mindfulness, subjective norm about mindfulness, perceived behavior control over mindfulness, and behavioral intention for mindfulness. Bootstrapping mediation analyses were also conducted to test the potential mediating effect in the model.

Results: A total of 297 Korean participants' responses (153 males and 144 females) were analyzed. The proposed model had a good fit. Conscientiousness was correlated with attitude toward mindfulness (β=.384, *P*<.001), subjective norm about mindfulness (β=.249, *P*<.001), and perceived behavior control over mindfulness (β=.443, *P*<.001). Neuroticism was not correlated with attitude toward mindfulness (β=−.072, *P*=.28), but was correlated with subjective norm about mindfulness (β=.217, *P*=.003) and perceived behavior control over mindfulness (β=−.235, *P*<.001). Attitude toward mindfulness (β=.508, *P*<.001), subjective norm about mindfulness (β=.132, *P*=.01), and perceived behavior control over mindfulness (β=.540, *P*<.001) were separately correlated with behavioral intention for mindfulness. Conscientiousness was not directly correlated with behavioral intention for mindfulness (β=−.082, *P*=.27), whereas neuroticism was directly correlated with behavioral intention for mindfulness (β=.194, *P*=.001). Conscientiousness was indirectly linked with behavioral intention for mindfulness through attitude toward mindfulness (B=0.171, 95% CI 0.103-0.251) and perceived behavior control over mindfulness (B=0.198, 95% CI 0.132-0.273) but not through subjective norm about mindfulness (B=0.023, 95% CI −0.002 to 0.060). Neuroticism was indirectly linked with behavioral intention for mindfulness via perceived behavior control over mindfulness (B=−0.138, 95% CI −0.197 to −0.088) but not via subjective norm about mindfulness (B=0.021, 95% CI −0.002 to 0.059).

XSL•FO
RenderX

Conclusions: The results show that the integration of the Big Five personality traits and TPB constructs is useful in predicting the use of mindfulness mobile apps. Focusing on conscientiousness and neuroticism in developing information dissemination and implementation strategies for enhancing mindfulness behavior using mobile apps may lead to the successful promotion of mindfulness mobile apps and adherence to mindfulness techniques.

(J Med Internet Res 2022;24(11):e39501) doi:10.2196/39501

KEYWORDS

personality traits; Theory of Planned Behavior; mindfulness; mobile apps; mental health

Introduction

During the COVID-19 pandemic, the prevalence of mental health disorders, such as mood and anxiety disorders, has increased globally [1,2]. As interest in mental health promotion increases, mindfulness meditation can be introduced as an evidence-based intervention to reduce psychological distress and alleviate the psychological impact of the pandemic and long-term quarantine measures [3-5]. Reflecting the ability to pay full attention to the present moment without any judgment and not be overly reactive or overwhelmed by past or future events [6,7], mindfulness has been shown to have positive effects on mental health and psychological well-being [8-11]. As improved mental health is associated with better health behaviors, mindfulness, which reduces psychological distress, can play a key role in facilitating health behaviors [12-14]. The rapid development of information and communications technologies has led to the expansion of eHealth and mobile health (mHealth) and diverse mobile apps that contain mindfulness content. Fact.MR [15] has estimated that the market for mindfulness meditation apps is expected to reach US $180 million by 2032 with an annual compound growth rate of 8.4%. The amount of research examining the effectiveness of mindfulness mobile apps in mental health is also increasing [16-19]. Nevertheless, further research is required on the design of apps and factors that affect their use to connect this interest in mindfulness mobile apps to the development of successful mHealth intervention strategies [20-22]. Further, it is important to identify the determinants of mindfulness behavior that potentially drive the use of mindfulness mobile apps.

This paper focuses on behavioral intentions, one of the factors that enable the prediction of actual behavior. By exploring the determinants of behavioral intentions for mindfulness, this research aims to identify the factors that affect not only behavioral intentions for mindfulness but also adherence to mindfulness. To this end, the Theory of Planned Behavior (TPB), which deals with the role of behavioral intentions in actual behaviors, was adopted as the study's core theory. The TPB is a socio-psychological model that was developed to examine the psychological processes that influence behavior. This theory claims that attitude toward behavior, subjective norm about behavior, and perceived behavior control could impact the intention to perform the behavior so as to eventually change actual behavior [23]. It defines attitude as the overall evaluations of a person engaging in a single behavior or a set of behaviors. Subjective norm refers to a person's belief about whether others think that they should engage in a particular behavior. Perceived behavior control relates to a person's perception of how easy or difficult it would be to perform the behavior [24]. The TPB states that engaging in a behavior depends on the relationship between these fundamental concepts and their rational processes; strong behavioral intention incurred by attitude, subjective norm, and perceived behavior control is likely to make people perform the behavior [24,25]. The TPB has been widely used to explain and predict various health behaviors [26], including smoking [27], drinking [28], and exercise [29]. In terms of mindfulness, a study of individuals with no prior experience of mindfulness meditation found that subjective norm was a predictor of people's phone app usage time for practicing mindfulness, and that attitude and perceived behavior control were positively associated with an intention to practice mindfulness [30]. Although not focusing specifically on mindfulness behavior, another study showed that attitude and perceived behavior control predicted intention to seek mental health services [31]. As mindfulness and mental health are closely linked, it is possible to assume that the determinants of intention to seek mental health services would tend to be similar to the determinants of intention to practice mindfulness. Thus, the TPB is a useful theoretical framework for predicting mindfulness-related behavior.

There have been attempts to use the TPB to determine additional factors associated with mental health–related behaviors, thereby expanding the boundaries of the TPB and increasing the explanatory power of the theory. The Big Five personality traits (ie, openness, conscientiousness, neuroticism, extraversion, and agreeableness) [32] are representative variables. Openness describes a person's disposition toward doing new things and intellectual activities; people with a high level of openness tend to be creative, imaginative, and curious. Conscientiousness refers to a person's tendency to regulate themselves to perform goal-directed behaviors; people with a high level of conscientiousness are more inclined to be organized, self-disciplined, and competent. Neuroticism is related to a person's perception of the world and overall emotional stability; people with a high level of neuroticism tend to be emotionally vulnerable, anxious, and experience a lot of stress. Extraversion describes a person's willingness to interact with their environment; people with a high level of extraversion tend to be sociable, outgoing, and seek excitement. Agreeableness reflects the way people manage their relationships with others; people with a high level of agreeableness are more inclined to be sympathetic, cooperative, and trusting [32-34]. Previous studies have reported that these personality traits influence TPB variables and behavioral intention [35-38]. For example, affective attitudes and perceived behavioral control mediate the relationship between conscientiousness and intention for physical activity [38].

XSL•FO
RenderX

Previous mindfulness studies investigating the relationship between the Big Five personality traits and the TPB are relatively insufficient compared with studies on other health behaviors. However, considering the close correlation between the characteristics of the Big Five personality traits and mindfulness [39,40], there should be a similar close relationship between the Big Five personality traits and the TPB in the context of mindfulness behavior. For instance, conscientious people tend to be responsible and have good impulse control to achieve a goal. Mindful people are more likely to respond deliberately and carefully rather than impulsively and habitually. Thus, it is feasible to assume that conscientiousness and mindfulness have a positive relationship [39]. People with neuroticism tend to be more susceptible to stress and negative emotions and experience dramatic changes in their feelings, while mindful people are more likely to be calm and able to control their negative emotions. Therefore, it can be assumed that neuroticism and mindfulness have a negative relationship [39]. Despite the lack of evidence, we suppose that the Big Five personality traits and the TPB still have the potential to predict or change mindfulness behavior.

This paper investigates the mediation relationship in which the Big Five personality traits influence the factors of the TPB and ultimately affect behavioral intentions for mindfulness. This study reveals the mechanisms of the TPB with consideration to personality traits in the context of mindfulness behavior, thus enhancing the theoretical framework by embracing new potential variables that can predict behavioral intention. As conscientiousness and neuroticism display a stronger relationship with mindfulness than openness, extraversion, and agreeableness [39], this research focuses on the influence of conscientiousness and neuroticism on the TPB.

Methods

Study Design and Procedure

We conducted an online survey of adults to examine their experiences of mindfulness. Before participating in the survey, the participants were asked to read the research information and complete an electronic consent form. Only consenting participants were allowed to continue with the study. They provided sociodemographic, clinical, and individual difference information relevant to mindfulness. The questionnaire items were translated from English to Korean and modified to fit the context of the study's topic.

Recruitment

Participants were recruited in December 2021 using an online panel developed by a survey company, dataSpring Korea, Inc. Adults older than 19 years and who consented to participate in the survey were recruited. Based on their responses to the survey's screening question, they were categorized into 4 groups: (1) *mindfulness mobile app users*—those who used mindfulness mobile apps within a month of either downloading an app or subscribing to a paid premium service; (2) *mindfulness mobile app churners*—those who had experience of using mindfulness mobile apps in the past but either deleted the apps or did not use them within a month of downloading them; (3) *other mindfulness behavior performers*—those who did not have any experience of using mindfulness mobile apps but had experience of practicing mindfulness behaviors through other means; and (4) *no mindfulness behaviors*—those with no experience in performing any type of mindfulness behavior.

After the survey, participants were recategorized into 4 groups: (1) people using mindfulness mobile apps and practicing other types of mindfulness behaviors; (2) people using only mindfulness mobile apps; (3) people who only practice other types of mindfulness behavior; and (4) people not practicing any type of mindfulness behavior, with a focus on current (within 1 month) mindfulness activities. Participants only practicing other types of mindfulness behaviors were excluded from the data analysis because the study focuses on the use of mindfulness mobile apps. Consequently, the responses of 297 participants were used for the data analysis. At the time of the survey, 142/297 (47.8%) participants were using mindfulness mobile apps (either in conjunction with other types of mindfulness behavior or using mindfulness mobile apps only) and 155/297 (52.2%) were not practicing any type of mindfulness behavior. Of the 142 participants using mindfulness mobile apps, 46 (32.4%) were using Mabo, 46 (32.4%) were using Kokkiri, 17 (12%) were using Calm, 13 (9.2%) were using Harumeditation, and 20 (14.1%) were using other mindfulness mobile apps.

Of the total 297 participants, 47 (15.8%) were 20-29 years old, 68 (22.9%) were 30-39 years old, 108 (36.4%) were 40-49 years old, 72 (24.2%) were 50-59 years old, and 2 (0.7%) were 60 years old or older. In addition, 153 (51.5%) participants were male and 144 (48.5%) were female; 180 (60.6%) participants were married and 102 (34.3%) were single; and 170 (57.2%) participants had no religious affiliation. More than half of the participants were university graduates (180/297, 60.6%). Regarding the participants' employment status, 192 (64.6%) were permanently employed. More than half of the participants (188/297, 63.3%) responded that they were in the middle-income bracket. Table 1 presents more information about participants' demographic characteristics.

XSL•FO
RenderX

Table 1. Demographic information of participants (N=297).

Characteristics	Participants, n (%)[a]
Age (years)	
20-29	47 (15.8)
30-39	68 (22.9)
40-49	108 (36.4)
50-59	72 (24.2)
≥60	2 (0.7)
Gender	
Male	153 (51.5)
Female	144 (48.5)
Marital status	
Single (never married)	102 (34.3)
Domestic partnership/common law marriage	4 (1.3)
Married	180 (60.6)
Divorced	11 (3.7)
Religion	
None	170 (57.2)
Protestant	40 (13.5)
Catholic	36 (12.1)
Buddhist	49 (16.5)
Cheondoist	1 (0.3)
Won Buddhist	1 (0.3)
Highest level of education	
High school graduate	48 (16.2)
College graduate (2-3 years)	39 (13.1)
University graduate (4-6 years)	180 (60.6)
Master’s degree	26 (8.8)
Doctorate	4 (1.3)
Current employment status	
Permanently employed	192 (64.6)
Temporarily employed (eg, part-time workers, dispatched workers, daily workers, freelancers)	54 (18.2)
Not employed	49 (16.5)
Retired	2 (0.7)
Income	
Very low	12 (4)
Low	80 (26.9)
Middle	188 (63.3)
High	15 (5.1)
Very high	2 (0.7)
Current experience of mobile mindfulness apps (within 1 month)	
Yes	142 (47.8)
No	155 (52.2)

XSL•FO
RenderX

Characteristics	Participants, n (%)[a]
Experience of being diagnosed with a psychiatric disease	
Yes	45 (15.2)
No	252 (84.8)
Experience of being diagnosed with a medical or surgical condition	
Yes	143 (48.1)
No	154 (51.9)
Experience of taking medication (within the last 30 days)	
Yes	131 (44.1)
No	166 (55.9)

[a]Percentages may not add up to 100% due to rounding.

Measures

Overview

In this study, conscientiousness and neuroticism were chosen as exogenous variables and behavioral intention for mindfulness as an endogenous variable. Attitude, subjective norm, and perceived behavior control regarding mindfulness were intermediate variables.

Personality Factor

The Korean Big Five Inventory (BFI)-15 was used to assess respondents' Big Five personality traits. The Korean BFI-15 was translated, abbreviated, and verified in Korean by Kim and colleagues [41] from John and Srivastava's BFI items [42]. Along with the stem question "I see myself as someone who," the Korean BFI-15 includes 15 items as follows: 3 items for openness, 3 items for conscientiousness, 3 items for neuroticism, 3 items for extraversion, and 3 items for agreeableness. Each item is answered on a 5-point scale (1="strongly disagree" to 5="strongly agree"). The responses to the conscientiousness and neuroticism items were used in the data analysis.

- Conscientiousness: Respondents were asked to indicate their conscientiousness based on 3 items: "I see myself as someone who does a thorough job," "...does things efficiently," and "...is a reliable worker." These items were averaged to create a scale (mean 3.67, SD 0.77; Cronbach α=.840).
- Neuroticism: Neuroticism was measured using 3 items: "I see myself as someone who gets nervous easily," "...is depressed, blue," and "...worries a lot." These items were averaged to create a scale (mean 2.79, SD 1.04; Cronbach α=.876).

Behavioral Factor

We modified the TPB measurement developed by Kim [43] to predict drug users' intention to use treatment services for drug addiction to suit the context of mindfulness behavior in this study. Kim's TPB measurement was developed and verified based on methods and evidence related to the TPB as posited by Fishbein and Ajzen [25]. The measurement in this study assesses respondents' attitude, subjective norm, perceived behavior control, and behavioral intention regarding mindfulness.

- Attitude toward mindfulness: Five items, based on a 5-point scale (1="strongly disagree" to 5="strongly agree"), were modified in the context of mindfulness: "It is worthwhile to perform mindfulness," "It is wise to perform mindfulness," "It is practical to perform mindfulness," "It is desirable to perform mindfulness," and "I am positive about performing mindfulness." The items were used to measure respondents' attitude toward mindfulness and averaged to create a scale (mean 4.10, SD 0.63; Cronbach α=.878).
- Subjective norm about mindfulness: Four items, based on a 5-point scale (1="strongly disagree" to 5="strongly agree"), were modified in the context of mindfulness: "Most people who are important to me think that I need to perform mindfulness," "Most people who are important to me would endorse me performing mindfulness," "Most people who are important to me would support me performing mindfulness," and "I feel pressure to perform mindfulness from people around me." The items were used to measure respondents' subjective norm about mindfulness and averaged to create a scale (mean 3.19, SD 0.82; Cronbach α=.770).
- Perceived behavior control about mindfulness: Three items, based on a 5-point scale (1="strongly disagree" to 5="strongly agree"), were modified in the context of mindfulness: "I am confident in performing mindfulness," "It is entirely up to me to perform mindfulness," and "I can control the situation around me to perform mindfulness." The items were used to measure respondents' perceived behavior control about mindfulness and averaged to create a scale (mean 3.84, SD 0.66; Cronbach α=.744).
- Behavioral intention for mindfulness: Five items, based on a 5-point scale (1="strongly disagree" to 5="strongly agree"), were modified in the context of mindfulness: "I intend to perform mindfulness," "I will perform mindfulness," "I plan to perform mindfulness," "I want to perform mindfulness," and "I am willing to perform mindfulness." The items were used to measure respondents' behavioral intention for mindfulness and averaged to create a scale (mean 3.91, SD 0.72; Cronbach α=.901).

XSL•FO
RenderX

Statistical Analysis

Structural equation modeling (SEM) was conducted to verify the proposed research model established to predict mindfulness behavior by integrating the Big Five personality traits and the TPB. First, a confirmatory factor analysis was conducted to examine the measurement model and assess the model fit using several goodness-of-fit indices. Next, SEM was conducted to evaluate the structural model and test the study hypotheses. Lastly, bootstrapping mediation analyses were conducted to test the significance of the mediation pathways more precisely. SPSS 22.0 (IBM, Inc.), Amos 22.0 (IBM, Inc.), and PROCESS macro for SPSS 4.1 [44] software were used for the data analysis.

Ethics Approval

This study was approved by the Institutional Review Board of Yongin Severance Hospital in Yonsei University Health System (IRB No. 9-2021-0167). We made a data collection request to an online panel research company. Only the online panel members older than 19 years had access to this study's online survey. The members interested in the survey could thoroughly review the study explanation and voluntarily decide to participate. They could participate in the survey only if they provided informed voluntary consent. However, participants could withdraw their consent or stop participating in the study at any time according to their free will. This study was conducted through a survey; no special side effects or physical damage was expected. We received deidentified raw data from the online panel research company after the survey completion. The data were password protected, and only the research team had access.

Results

Scale Validation and Model Specification

A confirmatory factor analysis was conducted to verify the factor structure of the proposed model. First, after examining the factor loading of the observed variables constituting the latent variables, all the factor loading values were found to be statistically significant (P<.05 in all cases). The factor loading values of all the observed variables were more than 0.50, except for an observed variable in subjective norm about mindfulness (ie, the "'I feel pressure to perform mindfulness from people around me" item), which was below 0.50. Therefore, we concluded that the observed variables make up the latent variables based on an appropriate theoretical conceptualization [45,46]. In other words, content validity was established.

Next, convergent validity—the explanatory power and validity of the latent variables itself—was examined through composite/construct reliability (CR), representing the internal consistency of the observed variables, and the average variance extracted (AVE) value, representing the size of the variance that the observed variables can explain. The convergent validity was secured for all the variables because the CR exceeded 0.70 and the AVE exceeded the threshold of 0.50 [46,47].

The details of each observed and latent variable are summarized in Table 2.

After examining convergent validity, discriminant validity was examined to determine whether there was an overlap or similarity between each latent variable and whether there was differentiation. Each AVE value had to be greater than the square value of the correlation coefficient between certain variables to secure discriminant validity [47]. In this measurement model, discriminant validity was secured because the minimum value of the AVE (0.502) was greater than the largest square value of the correlation coefficient (maximum=0.696×0.696=0.484).

In terms of reliability, the Cronbach α value of each variable measurement was calculated to examine the internal consistency of the measurement tool. As the Cronbach α values of all the variables were above .70, all the measurements showed good reliability [45,48].

After confirming that the study's measurement model met the validity and reliability requirements, an evaluation to test overall model fit was conducted. As a result, the values χ^2_{215}=431.1 (P<.001), χ^2/df=2.005, incremental fit index (IFI)=0.942, comparative fit index (CFI)=0.941, Tucker-Lewis index (TLI)=0.930, and root mean square error of approximation (RMSEA)=0.058 met all the criteria: χ^2 (P<.001), χ^2/df≤3, IFI≥0.90, CFI≥0.90, TLI≥0.90, RMSEA≤0.08 [48]. Therefore, the overall model showed a good fit.

The details of the measurement model verified in this study are summarized in Table 3.

XSL•FO
RenderX

Table 2. Confirmatory factor analysis: items and loadings.

Construct and scale items	Factor loading[a]	SMC[b]	1 – SMC	CR[c]	AVE[d]
Personality factor (Korean BFI[e]-15)					
Conscientiousness				0.840	0.637
Does a thorough job.	0.780	0.608	0.392		
Does things efficiently.	0.840	0.706	0.294		
Is a reliable worker.	0.772	0.596	0.404		
Neuroticism				0.880	0.712
Gets nervous easily.	0.904	0.817	0.183		
Is depressed, blue.	0.881	0.776	0.224		
Worries a lot.	0.736	0.542	0.458		
Behavioral factor (TPB[f])					
Attitude toward mindfulness				0.879	0.593
It is worthwhile to perform mindfulness.	0.768	0.590	0.410		
It is wise to perform mindfulness.	0.803	0.645	0.355		
It is practical to perform mindfulness.	0.718	0.516	0.484		
It is desirable to perform mindfulness.	0.774	0.599	0.401		
I am positive about performing mindfulness.	0.783	0.613	0.387		
Subjective norm about mindfulness				0.801	0.515
Most people who are important to me think that I need to perform mindfulness.	0.814	0.663	0.337		
Most people who are important to me would endorse me performing mindfulness.	0.851	0.724	0.276		
Most people who are important to me would support me performing mindfulness.	0.705	0.497	0.503		
I feel pressure to perform mindfulness from people around me.	0.421	0.177	0.823		
Perceived behavior control about mindfulness				0.749	0.502
I am confident in performing mindfulness.	0.812	0.659	0.341		
It is entirely up to me to perform mindfulness.	0.648	0.420	0.580		
I can control the situation around me to perform mindfulness.	0.653	0.426	0.574		
Behavioral intention for mindfulness				0.903	0.651
I intend to perform mindfulness.	0.786	0.618	0.382		
I will perform mindfulness.	0.861	0.741	0.259		
I plan to perform mindfulness.	0.811	0.658	0.342		
I want to perform mindfulness.	0.790	0.624	0.376		
I am willing to perform mindfulness.	0.775	0.613	0.387		

[a]All factor loadings are significant at $P<.001$.

[b]SMC: squared multiple correlation.

[c]CR: composite/construct reliability.

[d]AVE: average variance extracted.

[e]BFI: Big Five Inventory.

[f]TPB: Theory of Planned Behavior.

XSL•FO
RenderX

Table 3. Descriptive statistics and associated measures of the measurement model.

Variables	Cronbach α	Mean (SD)	AVE[a]	1	2	3	4	5	6
1. Conscientiousness	.840	3.67 (0.77)	0.637	*0.840* [b]	−0.370	0.359	0.132	0.494	0.294
2. Neuroticism	.876	2.79 (1.04)	0.712	0.137	*0.880*	−0.211	0.130	−0.400	−0.074
3. Attitude toward mindfulness	.878	4.10 (0.63)	0.593	0.129	0.045	*0.879*	0.345	0.666	0.695
4. Subjective norm about mindfulness	.770	3.19 (0.82)	0.515	0.017	0.017	0.119	*0.801*	0.231	0.363
5. Perceived behavior control over mindfulness	.744	3.84 (0.66)	0.502	0.244	0.160	0.444	0.053	*0.749*	0.696
6. Behavioral intention for mindfulness	.901	3.91 (0.72)	0.651	0.086	0.005	0.483	0.132	0.484	*0.903*

[a]AVE: average variance extracted.

[b]Composite reliabilities are along the diagonal and represented in italic font. Correlations are above the diagonal. Squared correlations are below the diagonal.

Hypothesis Testing

SEM Analysis

Based on the verification results of the measurement model, an SEM analysis was conducted to investigate the relationship between the variables and the model fit of the structural model. The values χ^2_{218}=524.3 (*P*<.001), χ^2/df=2.405, IFI=0.917, CFI=0.916, TLI=0.903, and RMSEA=0.069 satisfied all the criteria: χ^2 (*P*<.001), χ^2/df≤3, IFI≥0.90, CFI≥0.90, TLI≥0.90, RMSEA≤0.08 [48] and therefore showed a good fit. We therefore concluded that the structural model had an appropriate explanatory power to test the hypotheses predicting a relationship between variables.

The Direct Relationships Between Personality Traits and the TPB

With respect to personality traits, the relationships between conscientiousness and attitude toward mindfulness (β=.384, *P*<.001), conscientiousness and subjective norm about mindfulness (β=.249, *P*<.001), and conscientiousness and perceived behavior control over mindfulness (β=.443, *P*<.001) were all significant and positive. The relationship between neuroticism and attitude toward mindfulness (β=−.072, *P*=.28) was not significant. However, the relationships between neuroticism and subjective norm about mindfulness (β=.217, *P*=.003) and between neuroticism and perceived behavior control over mindfulness (β=−.235, *P*<.001) were significant. The former was positive, whereas the latter was negative.

The relationships between attitude toward mindfulness and behavioral intention for mindfulness (β=.508, *P*<.001), subjective norm about mindfulness and behavioral intention for mindfulness (β=.132, *P*=.01), and perceived behavior control over mindfulness and behavioral intention for mindfulness (β=.540, *P*<.001) were significant and positive. The relationship between conscientiousness and behavioral intention for mindfulness (β=−.082, *P*=.27) was not significant; however, the relationship between neuroticism and behavioral intention for mindfulness (β=.194, *P*=.001) was significant and positive.

Table 4 summarizes all the predictable direct effect paths of this structural model.

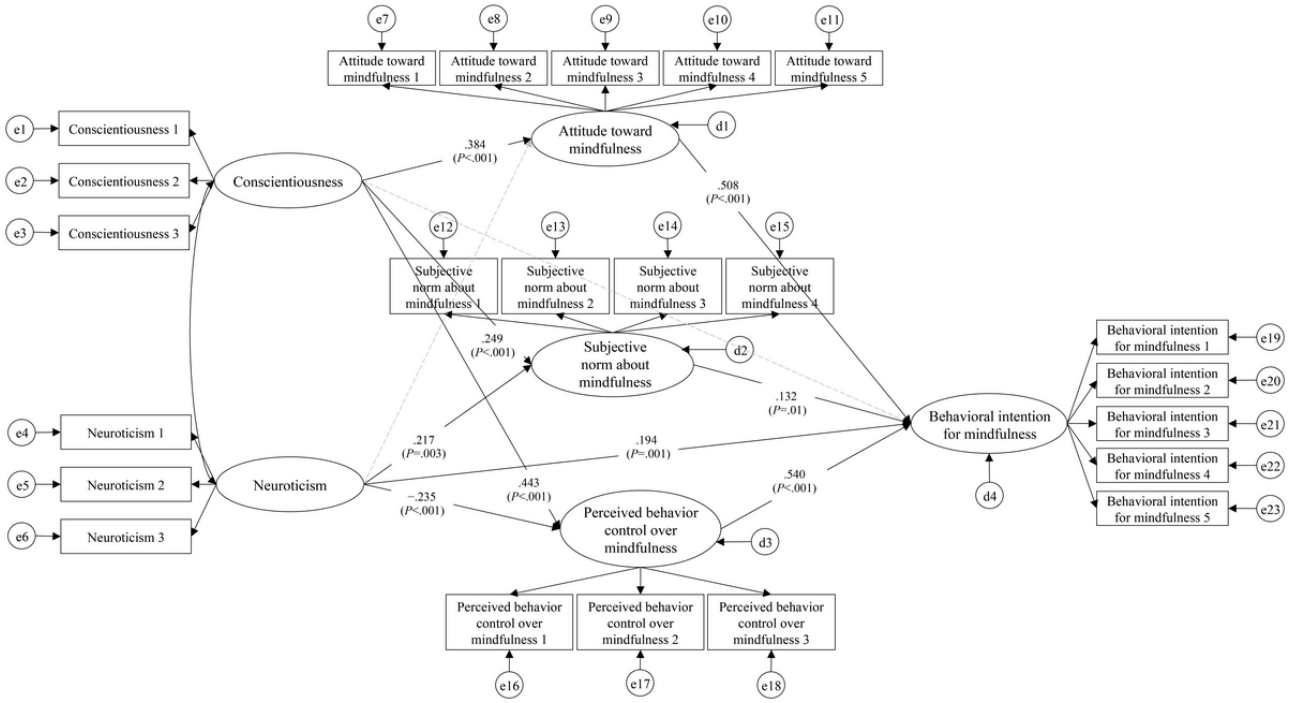
The results of verifying the research model are shown in Figure 1. It offers an understanding of possible mediating mechanisms of the model.

Table 4. The predictable direct effect paths of the structural model.

Direct path	Coefficients (β)	*P* value	Decision
Conscientiousness → Attitude toward mindfulness	.384	<.001	Supported
Conscientiousness → Subjective norm about mindfulness	.249	<.001	Supported
Conscientiousness → Perceived behavior control over mindfulness	.443	<.001	Supported
Neuroticism → Attitude toward mindfulness	−.072	.28	Not supported
Neuroticism → Subjective norm about mindfulness	.217	.003	Supported
Neuroticism → Perceived behavior control over mindfulness	−.235	<.001	Supported
Attitude toward mindfulness → Behavioral intention for mindfulness	.508	<.001	Supported
Subjective norm about mindfulness → Behavioral intention for mindfulness	.132	.01	Supported
Perceived behavior control over mindfulness → Behavioral intention for mindfulness	.540	<.001	Supported
Conscientiousness → Behavioral intention for mindfulness	−.082	.27	Not supported
Neuroticism → Behavioral intention for mindfulness	.194	.001	Supported

XSL•FO
RenderX

Figure 1. The path diagram and coefficients of the research model. e1-e23 are error terms for observed variables (ie, measurement errors); d1-d4 are disturbance terms for endogeneous variables (ie, residual errors).



The Indirect Relationships Between Personality Traits and the TPB

Next, the indirect effect path hypotheses were tested. Mediation analyses were conducted using the PROCESS macro model 4 (set to 5000 bootstrapped samples) [44]. In the PROCESS macro, the indirect effect is statistically significant when the CI does not include 0. Considering the paths of the model, the indirect effect path of "neuroticism → attitude toward mindfulness → behavioral intention for mindfulness" was not tested.

The results showed that conscientiousness had a potential indirect effect on behavioral intention for mindfulness mediated through attitude toward mindfulness (B=0.171, 95% CI 0.103-0.251) and perceived behavior control over mindfulness (B=0.198, 95% CI 0.132-0.273). Last, neuroticism had a potential indirect effect on behavioral intention for mindfulness mediated through perceived behavior control over mindfulness (B=−0.138, 95% CI −0.197 to −0.088). The relationships between conscientiousness and behavioral intention for mindfulness mediated by subjective norm about mindfulness (B=0.023, 95% CI −0.002 to 0.060) and between neuroticism and behavioral intention for mindfulness mediated by subjective norm about mindfulness (B=0.021, 95% CI −0.002 to 0.059) were not significant. As neuroticism was also directly associated with behavioral intention for mindfulness, neuroticism's potential indirect effect was considered a partial mediation path. Conscientiousness's potential indirect effects were considered full mediation paths.

Table 5 summarizes all the predictable indirect effect paths of this structural model.

Table 5. The predictable indirect effect paths of the structural model.

Indirect path	Coefficients (B)	Boot SE	Boot LLCI[a]	Boot ULCI[b]	Decision
Conscientiousness → Attitude toward mindfulness → Behavioral intention for mindfulness	0.171	0.037	0.103	0.251	Supported
Conscientiousness → Subjective norm about mindfulness → Behavioral intention for mindfulness	0.023	0.015	−0.002	0.060	Not supported
Conscientiousness → Perceived behavior control over mindfulness → Behavioral intention for mindfulness	0.198	0.036	0.132	0.273	Supported
Neuroticism → Subjective norm about mindfulness → Behavioral intention for mindfulness	0.021	0.015	−0.002	0.059	Not supported
Neuroticism → Perceived behavior control over mindfulness → Behavioral intention for mindfulness	−0.138	0.027	−0.197	−0.088	Supported

[a]LLCI: lower level of confidence interval.

[b]ULCI: upper level of confidence interval.

XSL•FO
RenderX

Discussion

Principal Findings

This research theorized and examined the indirect effect of the Big Five personality traits on behavioral intention for mindfulness based on the TPB. The study's main finding was that conscientiousness and neuroticism could potentially impact behavioral intention for mindfulness through attitude toward mindfulness and perceived behavior control over mindfulness. The results showed that among the Big Five personality traits, only conscientiousness and neuroticism were determinants influencing people's mindfulness behavior.

This study found no direct association between conscientiousness and behavioral intention for mindfulness; however, conscientiousness was positively associated with behavioral intention for mindfulness, via attitude toward mindfulness and perceived behavior control over mindfulness. This finding indicates that the higher the level of conscientiousness, the more positive the attitude toward mindfulness, and the stronger the perceived behavior control over mindfulness. In addition, attitude toward mindfulness and perceived behavior control over mindfulness were positively associated with behavioral intention for mindfulness, indicating that the more positive the attitude toward mindfulness and higher perceived behavior control over mindfulness, the stronger the behavioral intention for mindfulness. A strong sense of responsibility, good discipline, and effective self-regulation are therefore closely related to the characteristics of high conscientiousness [34]. People with a high level of conscientiousness are more likely to think positively about mindfulness and perceive that they are confident in performing mindfulness activities. Consequently, conscientiousness increases the likelihood of inducing behavioral intention for mindfulness.

Most importantly, neuroticism was negatively associated with behavioral intention for mindfulness via perceived behavior control over mindfulness. Inconsistent with previous studies [39,40], neuroticism was directly and positively associated with behavioral intention for mindfulness. Unlike conscientiousness, neuroticism showed a negative association with perceived behavior control over mindfulness. In other words, the higher the level of neuroticism, the weaker the perceived behavior control over mindfulness. In addition, perceived behavior control over mindfulness was positively associated with behavioral intention for mindfulness. Thus, the lower the level of perceived behavior control over mindfulness, the weaker the behavioral intention for mindfulness. As vulnerability to negative emotions is one of the characteristics of high levels of neuroticism [34], individuals with high neuroticism are less likely to perceive that they are confident in performing mindfulness. Hence, neuroticism may play a key role in reducing behavioral intention for mindfulness.

The results of this study suggest that people with high levels of neuroticism are more likely to have strong behavioral intentions for mindfulness regardless of perceived behavior control over mindfulness. One explanation is that people with high neuroticism are more likely to have poor mental health [49] but may seek treatment rather than avoidance under certain conditions. Those who perceived a high need for treatment for depression or social support are also more likely to seek professional help for depression [50]. Their strong motivation for mental health recovery may lead to stronger behavioral intention for mindfulness. In this regard, more studies are needed to verify this possible explanation and find a condition that may in turn produce results that are inconsistent with the findings of previous studies.

Implications

In the context of mindfulness, this study suggests an integrated model to explain the relationship between personality and behavior by combining the Big Five personality traits and the TPB. The results show that personality factors (ie, conscientiousness and neuroticism) and behavioral factors (ie, attitude toward mindfulness, subjective norm about mindfulness, perceived behavior control over mindfulness, and behavioral intention for mindfulness) are closely associated with changes in mindfulness behavior.

This study shows that the relationships or patterns between the Big Five personality traits and TPB variables in the context of mindfulness differ depending on personality traits. Attitude toward mindfulness and perceived behavior control over mindfulness mediate the association between conscientiousness and behavioral intention for mindfulness. However, only perceived behavior control over mindfulness mediates the association between neuroticism and behavioral intention for mindfulness. Furthermore, 3 other Big Five personality traits (ie, openness, extraversion, and agreeableness) are known to be less associated with mindfulness [39]. We posit that such inconsistencies might stem from the different characteristics of the Big Five personality traits. For example, the characteristics of high openness and liking new and diverse experiences are quite far removed from those attributes of mindfulness that require relaxation and peace. The conflict between personality traits and mindfulness characteristics may impact their associations. Examining the underlying mechanism of the effect of personality traits' different characteristics on mindfulness behavior could require a more advanced theoretical model with higher explanatory power on mindfulness behavior. Scrutinizing the multifaceted nature of mindfulness [51,52] may also provide a useful lens to interpret the relationship between personality traits and mindfulness. Future studies could examine the reasons why different relationships exist between personality traits and mindfulness behavior.

Our findings could also help to establish successful persuasion strategies that can encourage mindfulness behavior contingent on each individual's personality. If health professionals are able to identify a conscientious person, they could provide an intervention that may improve their attitude toward mindfulness and perceived behavior control over mindfulness to induce behavioral intentions for mindfulness and thereby promote mindfulness behavior. If health professionals are able to identify a person with high neuroticism, they could concentrate on improving the person's perceived behavior control over mindfulness and provide appropriate intervention effectively.

XSL•FO
RenderX

When considering mindfulness mobile apps as health intervention tools, applying immersive technology in developing mindfulness mobile apps to provide users with virtual experiences of mindfulness could be useful to improve users' perceived behavior control over mindfulness [53]. Embedding information on personality traits when designing the content of mindfulness mobile apps could be helpful in allowing users to perceive the content as personally relevant. As personalized health programs are known to have more persuasive effects than regular health messaging [54], tailoring a health intervention based on individual characteristics may improve mindfulness behavior. If a person with high neuroticism uses a mindfulness mobile app service to practice mindfulness behavior virtually and receives mindfulness tips for people with high neuroticism, the user may be more successful in overcoming the characteristics of neuroticism and improve their perceived behavior control over mindfulness. Taken together, scholars and health practitioners need to consider the indirect effects of personality traits to influence change or adherence to mindfulness behavior.

In this study, we only explored the potential of personality and behavioral factors to impact mindfulness behavior likely to drive the use of mindfulness mobile apps. We acknowledge that demonstrating other factors impacting mHealth adoption, not limited to personality traits, is important to build on the growing body of research examining mindfulness mobile apps usage. In their systematic review, Jacob and colleagues [55] introduced social and personal factors, technical and material factors, and health-related factors affecting mHealth adoption. Demographic factors (eg, age, gender, education, ethnicity, and socioeconomic factors), personal characteristics (eg, attitudes, motivation, and psychological factors), and cultural and social elements (eg, social influence, language, and culture) were representative social and personal factors. Usefulness (eg, perceived benefit, communication, and self-management), ease of use, technical factors (eg, access to technology, training, and tech support), monetary factors, data related (eg, privacy, credibility, and relevance), and user experience (eg, usability and personalization) composed technical and material factors. Disease or health condition, care team's role, health consciousness and literacy, relation to other therapies, health behavior, and insurance status were labeled as health-related factors [55]. Jacob and colleagues [55] emphasized taking a more holistic view of these factors and developing more patient-centered approaches (eg, fit into patient journey, inclusive design, and patient education) to facilitate mHealth adoption. They also highlighted that these factors are not mutually exclusive and showed mixed results on mHealth adoption depending on the context. It means that focusing on 1 or only a few factors alone on mHealth adoption and implementation is not likely to achieve success.

The complexity of the factors affecting mHealth adoption that Jacob and colleagues [55] stressed provides an implication to this study. The different potentials of personality and behavioral factors to impact mindfulness behavior and the use of mindfulness mobile apps can be explained by not only the characteristics of personality traits themselves but also other factors and contexts beyond personality traits. For instance, even people with high conscientiousness may only perform mindfulness behavior and use mindfulness mobile apps if they have strong motivation. They may not use mindfulness mobile apps if they feel such apps are uncomfortable to use or cannot protect personal data well. By contrast, people with low conscientiousness may use mindfulness mobile apps if they were educated about the importance of mindfulness and how to use mHealth services. This complexity may result in different and inconsistent outcomes of the associations between personality traits and mindfulness behavior. More comprehensive approaches to knowing the mediating and moderating roles of various factors are crucial to predict the use of mindfulness mobile apps. Future research is required to discover successful strategies for using the characteristics of the factors affecting mHealth adoption and their interconnectivity to facilitate the use of mindfulness mobile apps.

Limitations

The study has a few limitations. First, this research only looked at the Big Five personality traits and TPB variables and the mediating relationships among them. The study did not consider other possible variables that could impact mediations or other possible relationships among the Big Five personality traits and TPB variables. If this deficiency can be addressed and the proposed model can be expanded through follow-up studies, our understanding of the relationship between personality traits and behavior in the context of mindfulness will be improved. Second, although this study examined mediations supported by theoretical discussions and previous evidence, a careful interpretation of the causal assumptions is merited because this study used cross-sectional data. Future studies using experimental research methods are needed to verify causal relationships. Third, considering the development of this research's proposed model, future studies should be conducted with other types of health behaviors to refine the model or increase the model's degree of generalizability. Fourth, this study's participants recruited through an online panel were already likely to favor digital technology use itself. Therefore, it is encouraged to adopt a more randomized participant selection method to deal with this inherent bias and recruit participants regardless of technology preference or knowledge. Last, the data were self-reported and could be strengthened by physiological response measures.

Conclusions

The COVID-19 pandemic has highlighted the importance of mental health and mindfulness and various types of eHealth and mHealth mindfulness services have appeared [56]. The question is how to promote mindfulness behavior and encourage people to adhere to mindfulness behavior. Identifying the determinants of mindfulness behavior is the first to answer this question. This paper focused on personality traits and verified that the combination of the Big Five personality traits and the TPB provides a useful theoretical framework for predicting mindfulness behavior. The research results provide a foundation to develop an advanced model that is able to illustrate the relationships between personality and behavioral factors by adding other potential variables and moderation paths in the

XSL•FO
RenderX

context of mindfulness. The findings also underscore the potential of integrating other personality and behavior change theories to build a new model to explain mindfulness behavior. Furthermore, this research provides evidence to extend the research field and explore how diverse characteristics of personality traits affect mindfulness behavior. The role of personality may well explain mindfulness behavior more than people realize.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant number NRF-2021R1I1A1A01058680). This work was also supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety; Project Number 1711174427, RS-2020-KD000143).

Conflicts of Interest

None declared.

References

1. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
2. World Health Organization (WHO). COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. WHO. 2022. URL: https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide [accessed 2022-03-07]
3. Antonova E, Schlosser K, Pandey R, Kumari V. Coping With COVID-19: Mindfulness-Based Approaches for Mitigating Mental Health Crisis. Front Psychiatry 2021 Mar 23;12:563417 [FREE Full text] [doi: 10.3389/fpsyt.2021.563417] [Medline: 33833695]
4. Zhu JL, Schülke R, Vatansever D, Xi D, Yan J, Zhao H, et al. Mindfulness practice for protecting mental health during the COVID-19 pandemic. Transl Psychiatry 2021 May 28;11(1):329 [FREE Full text] [doi: 10.1038/s41398-021-01459-8] [Medline: 34050125]
5. Farris SR, Grazzi L, Holley M, Dorsett A, Xing K, Pierce CR, et al. Online Mindfulness May Target Psychological Distress and Mental Health during COVID-19. Glob Adv Health Med 2021 Mar 17;10:21649561211002461 [FREE Full text] [doi: 10.1177/21649561211002461] [Medline: 34497735]
6. Mindful staff. Jon Kabat-Zinn: Defining Mindfulness. Mindful. 2017 Jan 11. URL: https://www.mindful.org/jon-kabat-zinn-defining-mindfulness/ [accessed 2021-12-29]
7. Mindful staff. What is Mindfulness? Mindful. 2020 Jul 08. URL: https://www.mindful.org/what-is-mindfulness/ [accessed 2022-02-10]
8. Creswell JD. Mindfulness Interventions. Annu Rev Psychol 2017 Jan 03;68(1):491-516. [doi: 10.1146/annurev-psych-042716-051139] [Medline: 27687118]
9. Evans S, Ferrando S, Findler M, Stowell C, Smart C, Haglin D. Mindfulness-based cognitive therapy for generalized anxiety disorder. J Anxiety Disord 2008 May;22(4):716-721. [doi: 10.1016/j.janxdis.2007.07.005] [Medline: 17765453]
10. Hofmann SG, Gómez AF. Mindfulness-Based Interventions for Anxiety and Depression. Psychiatr Clin North Am 2017 Dec;40(4):739-749 [FREE Full text] [doi: 10.1016/j.psc.2017.08.008] [Medline: 29080597]
11. Kabat-Zinn J, Massion AO, Kristeller J, Peterson LG, Fletcher KE, Pbert L, et al. Effectiveness of a meditation-based stress reduction program in the treatment of anxiety disorders. Am J Psychiatry 1992 Jul;149(7):936-943. [doi: 10.1176/ajp.149.7.936] [Medline: 1609875]
12. Kushlev K, Drummond DM, Diener E. Subjective Well-Being and Health Behaviors in 2.5 Million Americans. Appl Psychol Health Well Being 2020 Mar 03;12(1):166-187. [doi: 10.1111/aphw.12178] [Medline: 31482675]
13. Walsh JL, Senn TE, Carey MP. Longitudinal associations between health behaviors and mental health in low-income adults. Transl Behav Med 2013 Mar 5;3(1):104-113 [FREE Full text] [doi: 10.1007/s13142-012-0189-5] [Medline: 23997836]
14. Cohen S, Schwartz JE, Bromet EJ, Parkinson DK. Mental health, stress, and poor health behaviors in two community samples. Preventive Medicine 1991 Mar;20(2):306-315. [doi: 10.1016/0091-7435(91)90029-4]
15. Mindfulness Meditation Apps Market. Fact.MR. 2022. URL: https://www.factmr.com/report/3075/mindfulness-meditation-apps-market [accessed 2022-03-25]
16. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]

XSL•FO
RenderX

17. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the Mindfulness Meditation Mobile App "Calm" to Reduce Stress Among College Students: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
18. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. The Mindfulness App Trial for Weight, Weight-Related Behaviors, and Stress in University Students: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Apr 10;7(4):e12210 [FREE Full text] [doi: 10.2196/12210] [Medline: 30969174]
19. Taylor G, Bylund CL, Kastrinos A, Alpert JM, Puig A, Krajewski JMT, et al. Practicing Mindfulness through mHealth Applications: Emerging Adults' Health-Enhancing and Inhibiting Experiences. Int J Environ Res Public Health 2022 Feb 24;19(5):2619 [FREE Full text] [doi: 10.3390/ijerph19052619] [Medline: 35270309]
20. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and Evaluation of Mindfulness-Based iPhone Apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
21. Huberty J, Green J, Puzia M, Stecher C. Evaluation of Mood Check-in Feature for Participation in Meditation Mobile App Users: Retrospective Longitudinal Analysis. JMIR Mhealth Uhealth 2021 Apr 23;9(4):e27106 [FREE Full text] [doi: 10.2196/27106] [Medline: 33890865]
22. Plaza I, Demarzo MMP, Herrera-Mercadal P, García-Campayo J. Mindfulness-based mobile applications: literature review and analysis of current features. JMIR Mhealth Uhealth 2013 Nov 01;1(2):e24 [FREE Full text] [doi: 10.2196/mhealth.2733] [Medline: 25099314]
23. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
24. Montaño DE, Kasprzyk D. Theory of Reasoned Action, Theory of Planned Behavior, and the Integrated Behavioral Model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; Jul 27, 2015:95-124.
25. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach. New York, NY: Psychology Press; Jul 24, 2009.
26. Godin G, Kok G. The theory of planned behavior: a review of its applications to health-related behaviors. Am J Health Promot 1996 Aug 25;11(2):87-98. [doi: 10.4278/0890-1171-11.2.87] [Medline: 10163601]
27. Godin G, Valois P, Lepage L, Desharnais R. Predictors of smoking behaviour: an application of Ajzen's theory of planned behaviour. Br J Addict 1992 Sep;87(9):1335-1343. [doi: 10.1111/j.1360-0443.1992.tb02742.x] [Medline: 1392555]
28. Collins SE, Witkiewitz K, Larimer ME. The theory of planned behavior as a predictor of growth in risky college drinking. J Stud Alcohol Drugs 2011 Mar;72(2):322-332 [FREE Full text] [doi: 10.15288/jsad.2011.72.322] [Medline: 21388605]
29. Shahar S, Ahmad MH, Mohd Fahmi Teng NI, Mohd Sakian NI, Omar B, Abd Manaf Z. Applying theory of planned behavior to predict exercise maintenance in sarcopenic elderly. CIA 2014 Sep:1551. [doi: 10.2147/cia.s60462]
30. Crandall A, Cheung A, Young A, Hooper AP. Theory-Based Predictors of Mindfulness Meditation Mobile App Usage: A Survey and Cohort Study. JMIR Mhealth Uhealth 2019 Mar 22;7(3):e10794 [FREE Full text] [doi: 10.2196/10794] [Medline: 30900992]
31. Bohon LM, Cotter KA, Kravitz RL, Cello PC, Fernandez Y Garcia E. The Theory of Planned Behavior as it predicts potential intention to seek mental health services for depression among college students. J Am Coll Health 2016 Jul 07;64(8):593-603 [FREE Full text] [doi: 10.1080/07448481.2016.1207646] [Medline: 27386898]
32. John OP, Naumann LP, Soto CJ. Paradigm shift to the integrative big-five trait taxonomy: history, measurement, and conceptual issues. In: John OP, Robins RW, Pervin LA, editors. Handbook of Personality: Theory and Research. New York, NY: The Guilford Press; Nov 24, 2010:114-158.
33. Cherry K. What Are the Big 5 Personality Traits?: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Verywell Mind. 2022 Oct 19. URL: https://www.verywellmind.com/the-big-five-personality-dimensions-2795422 [accessed 2022-11-14]
34. Lim A. The big five personality traits. Simply Psychology. 2020 Jun 15. URL: https://www.simplypsychology.org/big-five-personality.html [accessed 2022-11-14]
35. Davies CA, Mummery WK, Steele RM. The relationship between personality, theory of planned behaviour and physical activity in individuals with type II diabetes. Br J Sports Med 2010 Oct 27;44(13):979-984. [doi: 10.1136/bjsm.2008.050930] [Medline: 18753158]
36. Conner M, Abraham C. Conscientiousness and the Theory of Planned Behavior: Toward a more Complete Model of the Antecedents of Intentions and Behavior. Pers Soc Psychol Bull 2016 Jul 02;27(11):1547-1561. [doi: 10.1177/01461672012711014]
37. Courneya KS, Bobick TM, Schinke RJ. Does the Theory of Planned Behavior Mediate the Relation Between Personality and Exercise Behavior? Basic and Applied Social Psychology 2010 Jun 07;21(4):317-324. [doi: 10.1207/s15324834basp2104_5]
38. McEachan RR, Sutton S, Myers L. Mediation of personality influences on physical activity within the theory of planned behaviour. J Health Psychol 2010 Nov 03;15(8):1170-1180. [doi: 10.1177/1359105310364172] [Medline: 20522504]
39. Giluk TL. Mindfulness, Big Five personality, and affect: A meta-analysis. Personality and Individual Differences 2009 Dec;47(8):805-811. [doi: 10.1016/j.paid.2009.06.026]

XSL•FO
RenderX

40. Hanley AW, Garland EL. The mindful personality: A meta-analysis from a cybernetic perspective. Mindfulness (N Y) 2017 Dec 17;8(6):1456-1470 [FREE Full text] [doi: 10.1007/s12671-017-0736-8] [Medline: 29479377]
41. Kim JH, Kim BH, Ha MS. Validation of a Korean version of the Big Five inventory. Journal of Human Understanding and Counseling 2011 Jun;32(1):47-65 [FREE Full text]
42. John OP, Srivastava S. The Big Five trait taxonomy: history, measurement and theoretical perspectives. In: Pervin LA, John OP, editors. Handbook of Personality: Theory and Research. New York, NY: The Guilford Press; Oct 18, 1999:102-138.
43. Kim N. Development of a Measurement to Predict Drug Users' Intent to Use Treatment Services : Applying the Theory of Planned Behavior. SR 2018 Nov 30;19(4):1-34. [doi: 10.20997/sr.19.4.1]
44. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: The Guilford Press; Jan 24, 2022.
45. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. Upper Saddle River, NJ: Prentice Hall; Jan 16, 2009.
46. Arifin WN, Yusoff MSB. Confirmatory Factor Analysis of the Universiti Sains Malaysia Emotional Quotient Inventory Among Medical Students in Malaysia. SAGE Open 2016 May 17;6(2):215824401665024. [doi: 10.1177/2158244016650240]
47. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
48. Nguyen DM, Chiu YH, Le HD. Determinants of Continuance Intention towards Banks' Chatbot Services in Vietnam: A Necessity for Sustainable Development. Sustainability 2021 Jul 08;13(14):7625. [doi: 10.3390/su13147625]
49. Drake MM, Morris DM, Davis TJ. Neuroticism's susceptibility to distress: Moderated with mindfulness. Personality and Individual Differences 2017 Feb;106:248-252. [doi: 10.1016/j.paid.2016.10.060]
50. Shumet S, Azale T, Ayano G, Abebaw D, Amare T, Getnet W. Intention to seek help for depression and associated factors among residents of Aykel town, Northwest Ethiopia: cross-sectional study. Int J Ment Health Syst 2019 Mar 26;13(1):18 [FREE Full text] [doi: 10.1186/s13033-019-0274-y] [Medline: 30962818]
51. Bishop SR. Mindfulness: A Proposed Operational Definition. Clinical Psychology: Science and Practice 2004 Aug 01;11(3):230-241. [doi: 10.1093/clipsy/bph077]
52. Spinhoven P, Huijbers MJ, Zheng Y, Ormel J, Speckens AE. Mindfulness facets and Big Five personality facets in persons with recurrent depression in remission. Personality and Individual Differences 2017 May;110:109-114. [doi: 10.1016/j.paid.2017.01.045]
53. Wienrich C, Döllinger N, Hein R. Behavioral Framework of Immersive Technologies (BehaveFIT): How and Why Virtual Reality can Support Behavioral Change Processes. Front. Virtual Real 2021 Jun 24;2:627194. [doi: 10.3389/frvir.2021.627194]
54. Rimer BK, Kreuter MW. Advancing Tailored Health Communication: A Persuasion and Message Effects Perspective. J Communication 2006 Aug;56(s1):S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]
55. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical Factors Affecting Patients' Adoption of Mobile Health Tools: Systematic Literature Review and Narrative Synthesis. JMIR Mhealth Uhealth 2022 May 05;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]
56. Misitzis A. Increased Interest for Mindfulness Online. Int J Yoga 2020;13(3):247-249 [FREE Full text] [doi: 10.4103/ijoy.IJOY_15_20] [Medline: 33343156]

Abbreviations

AVE: average variance extracted
BFI: Big Five Inventory
CFI: comparative fit index
CR: composite/construct reliability
IFI: incremental fit index
LLCI: lower level of confidence interval
mHealth: mobile health
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SMC: squared multiple correlation
TLI: Tucker-Lewis index
TPB: Theory of Planned Behavior
ULCI: upper level of confidence interval

XSL•FO
RenderX

Edited by G Eysenbach; submitted 16.05.22; peer-reviewed by Z Dai, C Jacob; comments to author 22.06.22; revised version received 16.10.22; accepted 02.11.22; published 30.11.22.

Please cite as:
Kim S, Park JY, Chung K
The Relationship Between the Big Five Personality Traits and the Theory of Planned Behavior in Using Mindfulness Mobile Apps: Cross-sectional Survey
J Med Internet Res 2022;24(11):e39501
URL: https://www.jmir.org/2022/11/e39501
doi:10.2196/39501
PMID:36449344

©Sunghak Kim, Jin Young Park, Kyungmi Chung. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Artificial Intelligence in Intensive Care Medicine: Bibliometric Analysis

Ri Tang[1*], MD; Shuyi Zhang[1*], MD; Chenling Ding[1], MD; Mingli Zhu[1], MD; Yuan Gao[1], MD

Department of Intensive Care Medicine, Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

*these authors contributed equally

Corresponding Author:
Shuyi Zhang, MD
Department of Intensive Care Medicine
Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine
No 1630, Dongfang Road
Pudong New District
Shanghai, 200127
China
Phone: 86 13917816250
Email: shuishui286@qq.com

Abstract

Background: Interest in critical care–related artificial intelligence (AI) research is growing rapidly. However, the literature is still lacking in comprehensive bibliometric studies that measure and analyze scientific publications globally.

Objective: The objective of this study was to assess the global research trends in AI in intensive care medicine based on publication outputs, citations, coauthorships between nations, and co-occurrences of author keywords.

Methods: A total of 3619 documents published until March 2022 were retrieved from the Scopus database. After selecting the document type as articles, the titles and abstracts were checked for eligibility. In the final bibliometric study using VOSviewer, 1198 papers were included. The growth rate of publications, preferred journals, leading research countries, international collaborations, and top institutions were computed.

Results: The number of publications increased steeply between 2018 and 2022, accounting for 72.53% (869/1198) of all the included papers. The United States and China contributed to approximately 55.17% (661/1198) of the total publications. Of the 15 most productive institutions, 9 were among the top 100 universities worldwide. Detecting clinical deterioration, monitoring, predicting disease progression, mortality, prognosis, and classifying disease phenotypes or subtypes were some of the research hot spots for AI in patients who are critically ill. Neural networks, decision support systems, machine learning, and deep learning were all commonly used AI technologies.

Conclusions: This study highlights popular areas in AI research aimed at improving health care in intensive care units, offers a comprehensive look at the research trend in AI application in the intensive care unit, and provides an insight into potential collaboration and prospects for future research. The 30 articles that received the most citations were listed in detail. For AI-based clinical research to be sufficiently convincing for routine critical care practice, collaborative research efforts are needed to increase the maturity and robustness of AI-driven models.

(J Med Internet Res 2022;24(11):e42185) doi:10.2196/42185

KEYWORDS

intensive care medicine; artificial intelligence; bibliometric analysis; machine learning; sepsis

Introduction

Background

Artificial intelligence (AI) refers to a system that mimics human intelligence as characterized by the ability to perceive, reason, discover meaning, generalize, draw lessons from past experience and solve problems, or make decisions [1]. Machine learning (ML), natural language processing, and capability to visualize and recognize objects (computer vision) are all commonly used AI technologies [2].

ML is the dominant technique for implementing Al systems. ML refers to the science of programming computers that use

XSL•FO
RenderX

statistical analysis techniques to create algorithms to learn from data [1]. Because it uses statistical models and algorithms to evaluate enormous training data sets, it is also referred to as "programming with data." Supervised and unsupervised frameworks are two different types of ML (Figure 1).

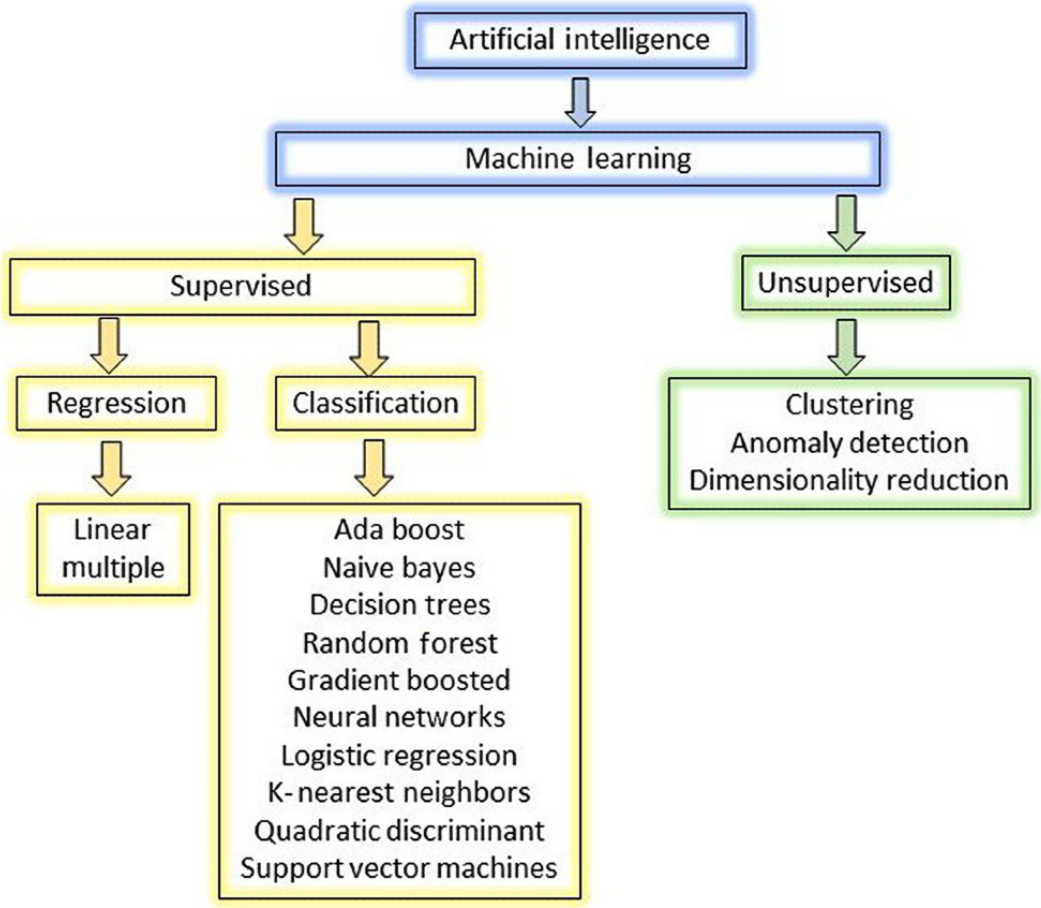
AI has been used in the medical field in molecular biology, bioinformatics, and medical imaging and to support population health management, provide tailored diagnosis and treatment, monitor patients, guide surgical care, and predict health trajectories [1,3,4].

The intensive care unit (ICU) is the most suitable ward among all the hospital wards to begin the transition to big data and the application of AI in research and even clinical practice in the near future. In ICUs, patients are closely monitored to detect physiological changes associated with deterioration that might require an appropriate reevaluation of the treatment plan. Nursing staff closely monitor patients in the ICU by charting neurological status, input, and output (including medication administration), etc. Bedside monitors facilitate this and continuously stream large amounts of data [5]. With advances in computer science, it has become possible to integrate and archive data in clinical documentation from various information systems and build a comprehensive system that is later transformed into a research database. Large public ICU databases include eICU and MIMIC databases. Open access to these databases encourages the use of AI technology in clinical research in intensive care medicine and the development of decision support tools.

The application of AI in critical care mainly involves disease diagnosis, prediction of disease progression (clinical deterioration), and characterization of specific disease phenotypes or endotypes in sepsis, septic shock, and acute respiratory distress syndrome (ARDS), etc.

Bibliometric study is a quantifiable informatics technique that analyzes the academic literature [6,7]. A general, quantitative, and qualitative overview of a certain topic can be provided via bibliometric analysis. It specifically identifies the most active authors, organizations, publications, influential studies, and international collaborations [7].

Figure 1. Machine learning is a branch of artificial intelligence encompassing two major approaches: supervised and unsupervised learning. Shown under each branch are algorithm types used in model development.



Goal of This Study

Although there has been a growing interest in critical care–related AI research, the literature is still lacking in comprehensive bibliometric studies that measure and analyze scientific publications globally. In this study, we aimed to (1) provide a holistic view of the research trends in AI application in ICUs; (2) highlight trending research topics in AI-related research focused on health care in ICUs; (3) highlight the contributions of prolific authors, leading countries, and the most productive academic institutions; and (4) provide an insight into potential collaboration and research directions in the future [8].

XSL•FO
RenderX

Methods

A bibliometric analysis study uses a mechanistic method to comprehend the global research trends in a certain field based on the outputs of the academic literature database. This approach distinguishes bibliometric analysis from reviews that are primarily designed to discuss the most recent advancements, challenges, and future directions of a particular topic [8].

Ethical Considerations

Ethics committee permission was not required, as this study was a retrospective bibliometric analysis of the existing published studies.

Data Source and Search Strategy

Data mining was conducted on March 18, 2022, using the Scopus database. Scopus is recognized as the largest abstract and citation database of peer-reviewed literature covering a wide range of subjects [8]. This study conducted a search for articles mentioning artificial intelligence or AI-related terms (neural network*, machine learning, deep learning, or natural language processing) and intensive care or ICU-related terms (critical care, critically ill, high dependency, or ICU) in the title, abstract, and keywords. The oldest publication dates to 1986, and the more recent ones are from 2022. The reproducible query string used for the search was: TITLE-ABS-KEY (“artificial intelligence” OR “neural network*” OR “machine learning” OR “deep learning” OR “natural language processing”) AND (“intensive care” OR “critical care” OR “critically ill” OR “high dependency” OR “ICU”) SEARCHED ON 18, March 2022.

Screening Strategy

The query string yielded 3619 documents. From those, only articles (2050/3619, 56.64%) were included (Table 1). A total of 4 duplicate articles were removed by using Stata (version 17; StataCorp) for data cleaning. The papers analyzed were restricted to those that (1) focused on intensive care medicine and (2) involved AI technologies. Two coauthors (SZ and RT) reviewed the titles of all studies as a pilot screening and removed irrelevant articles. Papers from the preliminary searches were categorized into include, exclude, or unsure. The abstracts and keywords of papers marked as unsure were further screened by 3 authors (SZ, RT, and MZ) and discussed until a consensus was reached in team meetings.

After screening the titles of all articles and abstracts, when necessary, 848 articles were excluded either because they did not focus on intensive care medicine or because they did not involve AI technologies. Finally, 1198 papers were included in the bibliometric analysis.

Table 1. The distribution of the bibliographic records by document type (N=3619).

Type of document	Frequency, n (%)
Article	2050 (56.64)
Conference paper	972 (26.85)
Review	241 (6.65)
Editorial	126 (3.48)
Note	66 (1.82)
Conference review	59 (1.6)
Letter	57 (1.57)
Book chapter	31 (0.85)
Short survey	9 (0.21)
Data paper	4 (0.1)
Erratum	4 (0.1)

Statistical Analysis

Overview

We used VOSviewer (version 1.6.18; Centre for Science and Technology Studies, Leiden University), a software tool for constructing and visualizing bibliometric maps, for bibliometric network visualization. The citation, bibliographical, and author keyword information of 1198 articles were exported to VOSviewer. The countries, authors, institutions, or keywords were included as objects of interest when creating maps using VOSviewer. We computed the growth rate of publications, research keywords, and publication patterns (countries, institutions, and journals). Bibliometric analyses were performed according to the instructions provided in the VOSviewer user manual [9].

Publication Output and Growth of Research Interest

The publication years were sorted, and the number of publications each year was counted using Stata 17.

The growth rate of publications over time was computed using the following compound annual growth rate formula:

Growth rate = ([number of publications in the last year or number of publications in the first year]1/(last year − first year) − 1) × 100 [10-12].

Preferred Journals

We used the citation analysis function of VOSviewer and set the unit of analysis as “sources.” Of the 443 sources (journals), 44 (9.9%) had >5 publications in total on AI in intensive care medicine. Journals were sorted according to the number of

XSL•FO
RenderX

publications. We listed the number of citations, an important index of the degree of attention and influence of the published papers [13,14]. CiteScore 2020 was obtained from the Scopus Preview website [15].

Leading Countries, International Collaboration, and Top Institutions

The citation trends of the top 10 most productive countries, top 15 most productive journals, and top 15 most productive research institutions were analyzed. The frequency and percentage of publications or citations in each country, journal, and institution were computed. This information was provided by Scopus and analyzed using the citation and coauthorship functions in VOSviewer. In the coauthorship analysis, the country-to-country link strength showed the number of publications coauthored by 2 linked countries. We created a thesaurus file to merge same institutions with different name variants.

Google Mymaps [16] was used for world map drawing. Using information from the International Monetary Fund's World Economic Outlook, the gross domestic product of the countries was estimated to ascertain whether the economic power of the countries had an impact on the productivity of publications [17].

Author Keywords

A total of 2267 keywords from 892 (74.5%) articles were analyzed for author co-occurrence. Owing to the lack of author keyword information, the remaining 306 (25.5%) articles were excluded. A thesaurus file was created to merge synonymic single words and congeneric phrases. For example, *coronavirus*, *coronavirus disease 2019*, and *sars-cov-2* were merged into 1 keyword and relabeled as *covid-19*. We identified high-frequency keywords and classified them into 3 categories: diseases, technology, and function.

Results

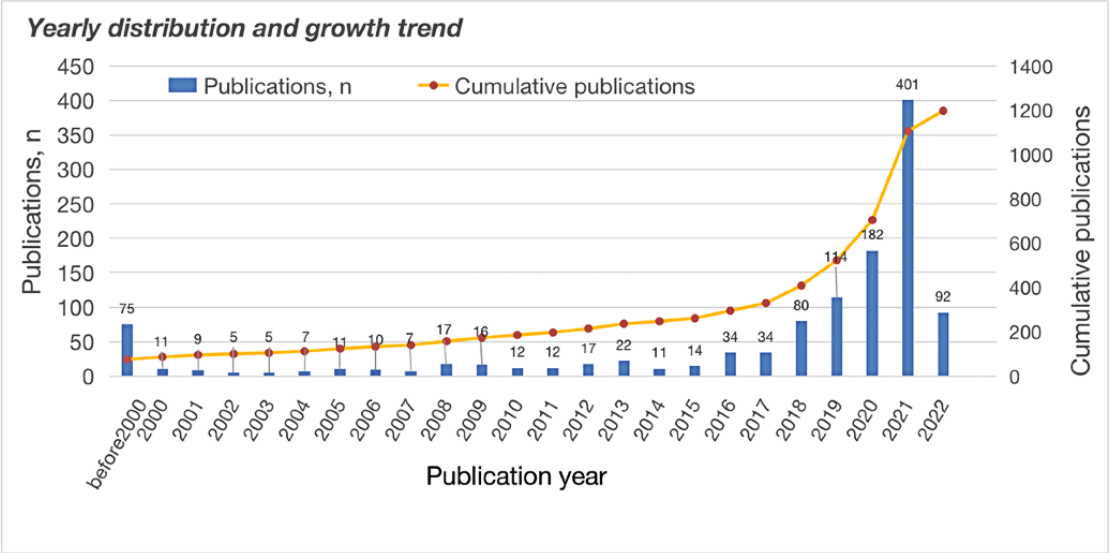
Publication Output and Growth of Research Interest

In all, 1198 research articles were published in 36 years (1986-2022; Figure 2). The oldest publication dates to 1986 [18].

It is suggested that a strong interest in AI in intensive care medicine started from 2018 when the annual growth rate increased by 135.3%. Since then, there has been a steep increase in annual publications, which has caused the cumulative number of publications to increase rapidly. The growth rate was 3.93% from 2011 to 2015 52.1% from 2016 to 2020, and the growth rate doubled last year to 120.3%. The number of publications increased steeply between 2018 and 2022, accounting for 72.53% (869/1198) of all the included papers.

Most articles (61.1%, 732/1198) were open-access articles that could be accessed for free. An article is likely to receive more citations if it is published in an open-access journal.

Figure 2. The annual and cumulative numbers of research articles on artificial intelligence in intensive care in Scopus from 1986 to 2022.



Preferred Journals

Our results revealed that the top 15 most productive journals were owned by 9 different publishers (Table 2). The most productive journal was *Artificial Intelligence in Medicine* with 42 articles, covering 3.5% of the total publications, followed by *Critical Care Medicine* (40, 3.3%), *PLoS ONE* (35, 2.9%), and *Scientific Reports* (33, 2.8%). *Critical Care Medicine*, a Lippincott Williams and Wilkins journal, received the highest number of citations—1166. One of their articles, "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU," published in 2018, was the most cited article, with 247 citations.

A total of 9 journals had a CiteScore of ≥5 according to the CiteScore 2020 report. *Critical Care Medicine* (CiteScore 12.7) and *JMIR Medical Informatics* (CiteScore 1.59) had the highest and lowest CiteScores, respectively. Although ranked 7th with 26 articles in Scopus, the total number of citations and number of citations per document of *Frontiers in Medicine* were significantly lower than those of other journals.

XSL•FO
RenderX

Table 2. The top 15 most productive journals on artificial intelligence in critical care research with their most cited article.

	TP[a] (N=1198), n (%)	TC[b], n	CiteScore 2020	Ranking based on citation	Citation per document	Most cited article	Times cited, n	Publisher
Artificial Intelligence in Medicine	42 (3.51)	983	8	2	23.40	A Working System for the Automated Control of Assisted Ventilation in ICUs	84	Elsevier
Critical Care Medicine	40 (3.34)	1166	12.7	1	29.15	An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU	249	Lippincott Williams and Wilkins Ltd
PLoS ONE	35 (2.92)	412	5.3	7	11.77	Machine Learning Models for Early Sepsis Recognition in the Neonatal Intensive Care Unit Using Readily Available Electronic Health Record Data	48	Public Library of Science
Scientific Reports	33 (2.75)	315	7.1	8	9.55	Prediction of Ventricular Tachycardia One Hour Before Occurrence Using Artificial Neural Networks	56	Springer Nature
Computer Methods and Programs in Biomedicine	26 (2.17)	241	7.7	11	9.27	An Empirical Evaluation of Deep Learning for ICD-9 Code Assignment Using MIMIC-III Clinical Notes	36	Elsevier
Critical Care	26 (2.17)	470	10.1	4	18.08	Development and Validation of a Novel Molecular Biomarker Diagnostic Test for the Early Detection of Sepsis	96	BioMed Central Ltd
Frontiers in Medicine	26 (2.17)	57	4.1	15	2.19	A Machine Learning-Based Prediction of Hospital Mortality in Patients With Postoperative Sepsis	14	Frontiers Media SA
BMC Medical Informatics and Decision Making	25 (2.09)	233	3.9	12	9.32	A Comparative Analysis of Multi-Level Computer-Assisted Decision Making Systems for Traumatic Injuries	33	Springer Nature
Computers in Biology and Medicine	25 (2.09)	464	7.3	6	18.56	A Computational Approach to Early Sepsis Detection	116	Elsevier
Journal of Clinical Monitoring and Computing	25 (2.09)	300	3.7	10	12.00	Using Physiological Models and Decision Theory for Selecting Appropriate Ventilator Settings	72	Springer Nature
IEEE Journal of Biomedical and Health Informatics	21 (1.75)	116	10.2	13	5.52	Multi-Sensor Fusion Approach for Cuff-Less Blood Pressure Measurement	32	IEEE
International Journal of Medical Informatics	21 (1.75)	468	7.1	5	22.29	User-Centered Design Techniques for a Computerised Antibiotic Decision Support System in an Intensive Care Unit	61	Elsevier
JMIR Medical Informatics	19 (1.59)	312	2.9	9	16.42	Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: a Machine Learning Approach	203	JMIR Publications Inc
Journal of Biomedical Informatics	19 (1.59)	522	8.1	3	27.47	Reducing False Alarm Rates for Critical Arrhythmias Using the Arterial Blood Pressure Waveform	151	Academic Press Inc
IEEE Access	17 (1.42)	112	4.8	14	6.59	Predicting Complications in Critical Care Using Heterogeneous Clinical Data	25	IEEE

[a]TP: total publications.

[b]TC: total citations.

XSL•FO
RenderX

Leading Countries, International Collaboration, and Top Institutions

The top 15 most productive countries contributing to the growth of AI in critical care research activities globally are listed in Table 3 and Figure 3. The United States was the leading country with 488 publications, accounting for 40.73% of all publications (1198) worldwide. With one-third of the total publications in the United States, China was the second most productive country (173/1198, 14.44%).

Table 3. The distribution of the bibliographic records by top 10 (by quantity) countries.

Country	Rank based on total output	Output (N=1198), n (%)	Citations (N=18,876),n (%)	Rank based on citations	Citation per document	GDP[a] rank
United States	1	488 (40.7)	8678 (46)	1	17.78	1
China	2	173 (14.4)	1294 (6.9)	3	7.48	2
United Kingdom	3	116 (9.7)	2765 (14.7)	2	23.84	6
Germany	4	62 (5.2)	873 (4.6)	5	14.08	4
Canada	5	60 (5)	740 (3.9)	6	12.33	8
Italy	6	58 (4.8)	386 (2)	12	6.66	9
France	7	52 (4.3)	905 (4.8)	4	17.40	7
Spain	8	47 (3.9)	351 (1.9)	13	7.47	15
South Korea	9	46 (3.8)	442 (2.3)	10	9.61	12
Australia	10	41 (3.4)	520 (2.8)	8	12.68	13
India	11	39 (3.3)	495 (2.6)	9	12.69	5
Netherlands	12	39 (3.3)	534 (2.8)	7	13.69	19
Taiwan	13	35 (2.9)	319 (1.7)	14	9.11	21
Iran	14	29 (2.4)	171 (0.9)	15	5.90	14
Belgium	15	24 (2)	403 (2.1)	11	16.79	25

[a]GDP: gross domestic product.

Figure 3. World map of the top 15 most productive countries based on publications on artificial intelligence in intensive care units.

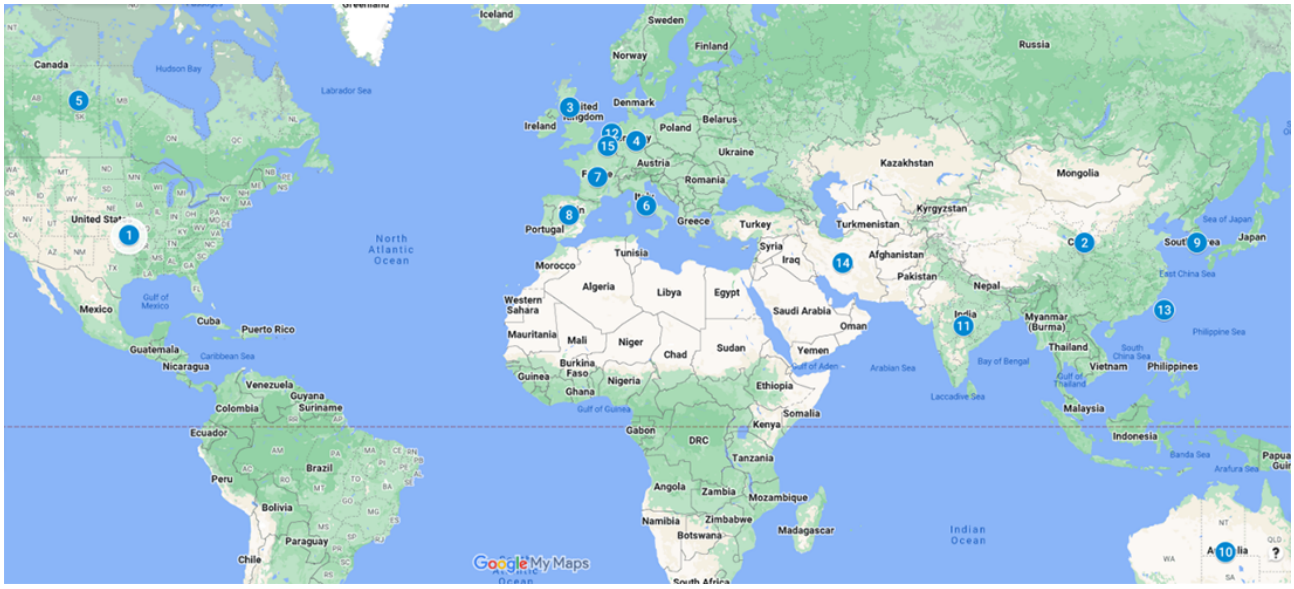


Figure 4 illustrates the distribution of the countries. This figure demonstrates how countries are establishing research networks and collaborating on the study of AI in intensive care. In VOSviewer, a country's proximity to another indicates how strong their relatedness is. The stronger the link between 2 countries, the thicker the line. The results of coauthorship showed that the United States was the most affiliated country, being linked to 34 countries or territories with 247 times of coauthorship. The list was followed by the United Kingdom (28 links, 122 coauthorships), Italy (27 links, 93 coauthorships), China (26 links, 94 coauthorships), and others.

We listed the 15 most productive institutions based on the number of articles these institutions have published on AI in intensive care medicine (Table 4). Of the 15 institutions, 13 were in the United States, which further suggested the dominant

XSL•FO
RenderX

role of the country in this research field. Of the 15 universities listed, 9 were in the top 100 best universities based on the World University Rankings 2022 [19].

Figure 4. Bibliometric map created based on coauthorship analysis between countries with network visualization mode. The QR code can be used to open this figure in VOSviewer.

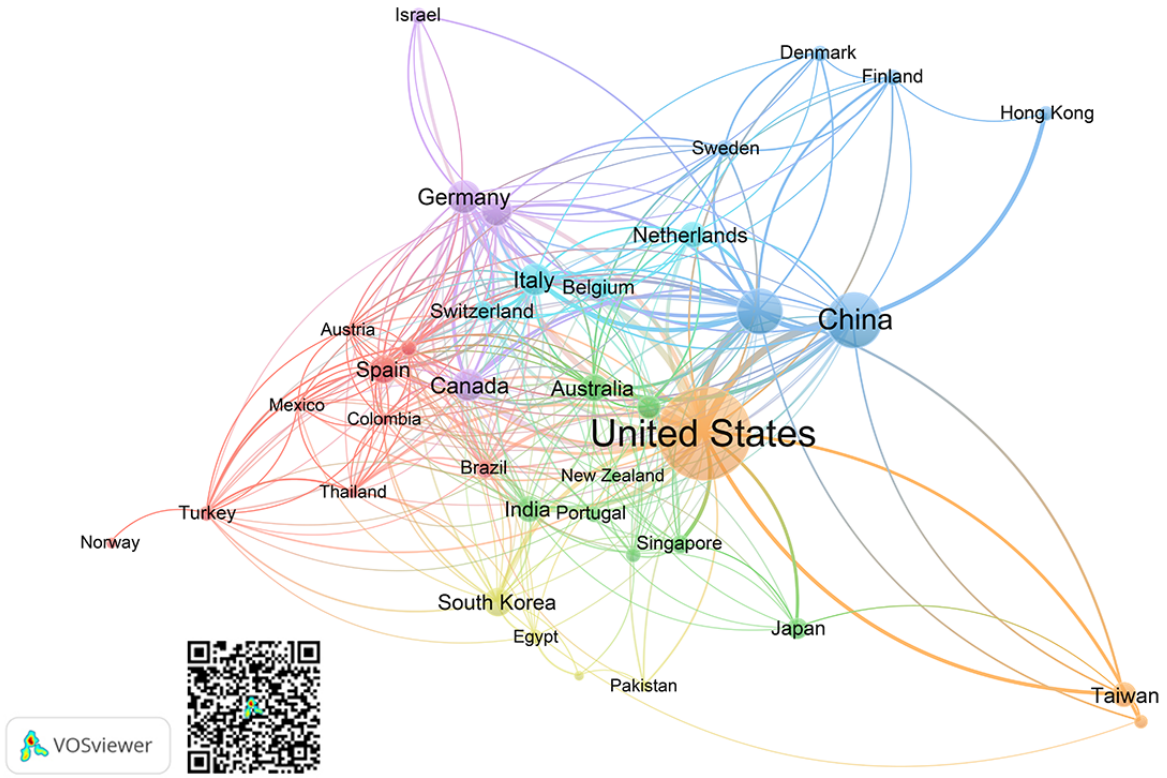


Table 4. The top 15 most productive institutions based on the number of articles published on artificial intelligence in intensive care medicine.

Rank	Organization	TPi[a]	Citations, n	Citation per document	University ranking
1	Massachusetts Institute of Technology, Cambridge, United States	12	411	34.25	5
2	University of Michigan, Ann arbor, United States	11	77	7.00	24
3	Beth Israel Deaconess Medical Center, Boston, United States	9	596	66.22	—[b]
4	Harvard Medical School, Boston, United States	9	122	13.56	2
5	Mayo Clinic, Rochester, United States	9	41	4.56	—
6	University of Pennsylvania, Philadelphia, United States	8	130	16.25	13
7	Johns Hopkins University, Baltimore, United States	7	23	3.29	13
8	Georgia Institute of Technology, Atlanta, United States	7	347	49.57	45
9	Icahn School of Medicine at Mount Sinai, New York, United States	7	85	12.14	—
10	University of California San Francisco, San Francisco, United States	7	670	95.71	—
11	Emory University School of Medicine, Atlanta, United States	6	324	54.00	82
12	Peking University, Beijing, China	6	115	19.17	16
13	University of Pittsburgh, Pittsburgh, United States	6	139	23.17	—
14	Kuopio University Hospital, Kuopio, Finland	5	87	17.40	—
15	University of Chicago, Chicago, United States	5	139	27.80	10

[a]TPi: total publications of a given academic institution.

[b]Not available.

XSL•FO
RenderX

Author Keywords

Among the 2267 author keywords recorded, 1785 (78.73%) occurred only once, 230 (10.14%) occurred twice, and 252 (11.11%) occurred thrice. After relabeling synonymic words and congeneric phrases using the thesaurus file, 102 (4.49%) keywords met the threshold of a minimum of 5 occurrences for mapping in VOSviewer (Figure 5 and Figure 6; Table 5). *Machine learning* was the most frequently encountered keyword, with 347 occurrences and 809 links to other keywords. In the same cluster of *machine learning*, general terms included *big data* (9 occurrences, 15 links), *data science* (6 occurrences, 13 links), *prediction models* (5 occurrences, 10 links), and *clustering* (5 occurrences, 14 links). *Machine learning* co-occurred with ICU-related professional keywords, including *critical care medicine*, *acute respiratory distress syndrome*, *acute respiratory failure*, *endotracheal intubation*, *mechanical ventilation*, and *personalized medicine*.

Figure 5. Bibliometric map created based on author keywords co-occurrence with network visualization mode. The QR code can be used to open the figure in VOSviewer. Colors show clustering. Keywords in the same cluster are of the same color. The circle size increases with the number of times a keyword is used. ARDS: acute respiratory distress syndrome; ECG: electrocardiogram; EEG: electroencephalogram.

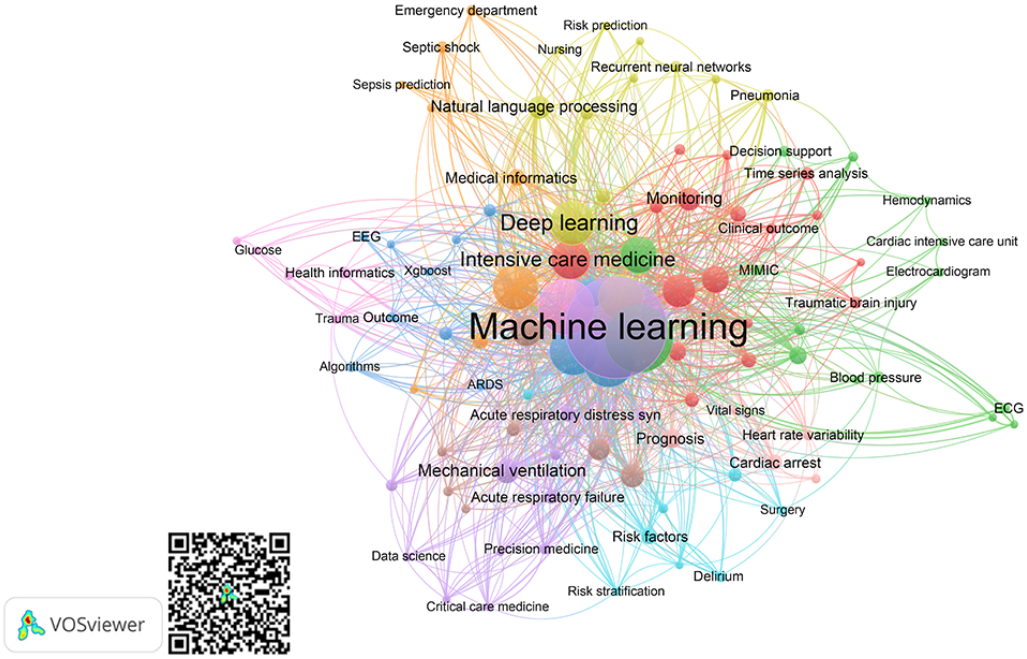
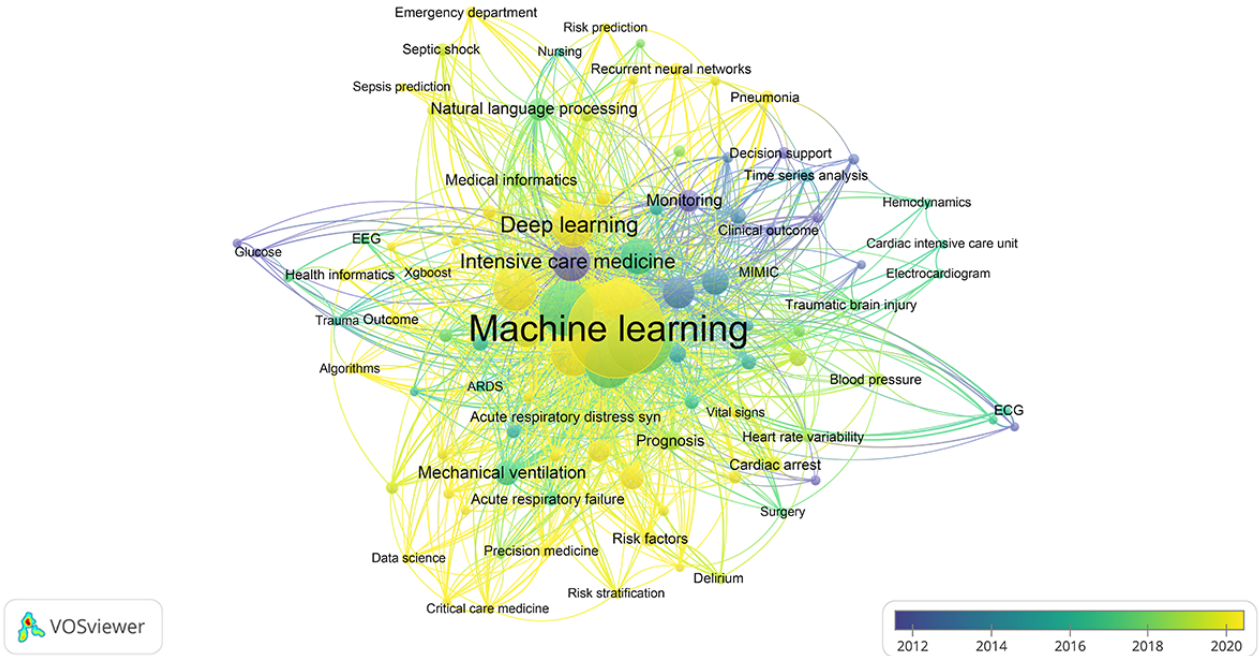


Figure 6. Bibliometric map created based on the co-occurrence of author keywords with overlay visualization mode. The color indicates the average publication year of the documents in which a keyword occurs. ARDS: acute respiratory distress syndrome; ECG: electrocardiogram; EEG: electroencephalogram.



XSL•FO
RenderX

Table 5. The 30 most frequently used keywords in articles on artificial intelligence in intensive care.

Keyword	Occurrences, n
Machine learning	347
Intensive care unit	172
Artificial intelligence	124
Critical care	102
Predictive analytics	97
Sepsis	86
COVID-19	85
Deep learning	84
Intensive care medicine	67
Decision support systems	60
Neural networks	50
Electronic health records	46
Mortality	42
Artificial neural networks	36
Mechanical ventilation	31
Predictive model	29
Natural language processing	27
Monitoring	26
Acute kidney injury	25
Prognosis	21
Data mining	17
Random forest	17
Medical informatics	16
Classification	15
Diagnosis	14
Risk factors	14
Acute respiratory distress syndrome	13
Cardiac arrest	13
Deterioration	13
Support vector machines	13

We categorized the keywords into 3 different categories based on the diseases in ICU, ML technologies used, and the function of ML in the research (Table 6). The top 5 diagnoses were sepsis, COVID-19, acute kidney injury, ARDS, and cardiac arrest. The top 5 AI technologies were ML, AI, deep learning, decision support systems, and neural networks. The top 5 functions were prediction (of clinical deterioration or mortality), mortality, monitoring, disease prognosis, and (disease phenotype) classification.

XSL•FO
RenderX

Table 6. Top author keywords in publications on artificial intelligence in critical care medicine.

Category	Frequency, n
Disease	
Sepsis	86
COVID-19	85
Acute kidney injury	25
Acute respiratory distress syndrome	13
Cardiac arrest	13
Technology	
Machine learning	347
Artificial intelligence	124
Deep learning	84
Decision support systems	60
Neural networks	50
Function	
Prediction	97
Mortality	42
Monitoring	26
Prognosis	21
Classification	15

Citation Analysis

The 30 most cited articles among the 1198 articles on AI in intensive care medicine between 1986 and 2022 are presented in Table 7. The total number of citations and average number of citations per year are given. Of the 30 most cited articles, 10 (33%) were related to sepsis, one of the hottest topics in intensive care, and 11 (37%) articles were related to the use of AI technologies for the prediction of acute kidney injury, sepsis, hypotension diagnosis, clinical outcomes (mortality, survival, and ICU length of stay), and complications.

XSL•FO
RenderX

Table 7. The 30 most cited articles on artificial intelligence in intensive care unit ranked in the descending order of the number of citations.

Rank	Article	Citations, n	CPY[a]	Rank by CPY
1	Matthieu Komorowski, Leo A. Celi et al "The Artificial Intelligence Clinician Learns Optimal Treatment Strategies for Sepsis in Intensive Care" *Nature Medicine* (2018)	290	73	1
2	Shamim Nemati, Andre Holder et al "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU" *Critical Care Medicine*,46,4 (2018)	249	62	2
3	Tom J. Pollard, Alistair E. W. Johnson et al "The eICU Collaborative Research Database, a Freely Available Multi-Center Database for Critical Care Research" *Scientific Data*,5,1 (2018)	211	53	3
4	Thomas Desautels, Jacob Calvert et al "Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: a Machine Learning Approach" *JMIR Med Inform*,4,3 (2016)	203	34	7
5	Brendon P. Scicluna, Lonneke A. van Vught et al "Classification of Patients With Sepsis According to Blood Genomic Endotype: a Prospective Cohort Study" *The Lancet Respiratory Medicine*,5,10, (2017)	164	33	9
6	Alistair E. W. Johnson, Mohammad M. Ghassemi et al "Machine Learning and Decision Support in Critical Care" *Proceedings of the IEEE*,104,2, (2016)	162	27	11
7	Richard Dybowski, Peter Weller et al "Prediction of Outcome in Critically Ill Patients Using Artificial Neural Network Synthesised by Genetic Algorithm" *Lancet*,347,9009,4 (1996)	160	6	28
8	Romain Pirracchio, Maya L. Petersen et al "Mortality Prediction in Intensive Care Units With the Super ICU Learner Algorithm (SICULA): a Population-Based Study" *The Lancet Respiratory Medicine*,3,1 (2015)	160	23	15
9	Q. Li, G. D. Clifford "Dynamic Time Warping and Machine Learning for Signal Quality Assessment of Pulsatile Signals" *Physiological Measurement*,33,9 (2012a)	156	16	20
10	Anton Aboukhalil, Larry Nielsen et al "Reducing False Alarm Rates for Critical Arrhythmias Using the Arterial Blood Pressure Waveform" *Journal of Biomedical Informatics*,41,3,6 (2008)	151	11	23
11	Feras Hatib, Zhongping Jian et al "Machine-learning Algorithm to Predict Hypotension Based on High-Fidelity Arterial Pressure Waveform Analysis" *Anesthesiology*,129,4, 10 (2018)	147	37	5
12	U. Rajendra Acharya, Hamido Fujita et al "Automated Identification of Shockable and Non-Shockable Life-Threatening Ventricular Arrhythmias Using Convolutional Neural Network" *Future Generation Computer Systems*,79,2 (2018)	144	36	6
13	Qingqing Mao, Melissa Jay et al "Multicentre Validation of a Sepsis Prediction Algorithm Using Only Vital Sign Data in the Emergency Department, General Ward and ICU" *BMJ Open*,8,1 (2018)	134	34	8
14	Zhengping Che, Sanjay Purushotham et al "Interpretable Deep Models for ICU Outcome Prediction" *AMIA. Annual Symposium Proceedings* (2016)	132	22	17
15	Abbas K. Abbas, Konrad Heimann et al "Neonatal NonContact Respiratory Monitoring Based on Real-Time Infrared Thermography" *BioMedical Engineering Online,10,1,10* (2011)	126	11	21
16	Jacob S. Calvert, Daniel A. Price et al "A Computational Approach to Early Sepsis Detection" *Computers in Biology and Medicine,74,*7 (2016)	116	19	18
17	Gilles Clermont, Derek C. Angus et al "Predicting Hospital Mortality for Patients in the Intensive Care Unit: a Comparison of Artificial Neural Networks with Logistic Regression Models" *Critical Care Medicine*,29,2 (2001)	112	10	24
18	David W. Shimabukuro, Christopher W. Barton et al "Effect of a Machine Learning-Based Severe Sepsis Prediction Algorithm on Patient Survival and Hospital Length of Stay: a Randomised Clinical Trial" BMJ *Open Respiratory Research*,4,1,11 (2017)	112	22	16
19	Jan Claassen, Kevin Doyle et al "Detection of Brain Activation in Unresponsive Patients with Acute Brain Injury" *New England Journal of Medicine*,380,26,6 (2019)	111	37	4
20	Sanjay Purushotham, Chuizheng Meng et al "Benchmarking Deep Learning Models on Large Healthcare Datasets" *Journal of Biomedical Informatics*,83,7 (2018)	105	26	12
21	Jay L. Kovner. Kyle A. Carey et al "The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model" *Critical Care Medicine*,46,7 (2018)	105	26	13
22	Alexander Meyer, Dina Zverinski et al "Machine Learning for Real-Time Prediction of Complications in Critical Care: a Retrospective Study" *The Lancet Respiratory Medicine*,6,12 (2018)	101	25	14
23	Michel Dojat, Laurent Brochard et al "A Knowledge-Based System for Assisted Ventilation of Patients in Intensive Care Units" *International journal of clinical monitoring and computing 9*,4 (1992)	99	3	30

XSL•FO
RenderX

Rank	Article	Citations, n	CPY[a]	Rank by CPY
24	Nicos Maglaveras, Telemachos Stamkopoulos et al “An Adaptive Backpropagation Neural Network for Real-Time Ischemia Episodes Detection: Development and Performance Analysis Using the European ST-T Database” *IEEE Transactions on Biomedical Engineering*,45,7 (1998)	96	4	29
25	Allison Sutherland, Mervyn Thomas et al “Development and Validation of a Novel Molecular Biomarker Diagnostic Test for the Early Detection of Sepsis” *Critical Care*,15,3,6 (2011)	96	9	25
26	Hye Jin Kam, Ha Young Kim “Learning Representations for the Early Detection of Sepsis With Deep Neural Networks” *Computers in Biology and Medicine*,89,10 (2017)	96	19	19
27	Michelle M. Clark, Amber Hildreth et al “Diagnosis of Genetic Diseases in Seriously Ill Children by Rapid Whole-Genome Sequencing and Automated Phenotyping and Interpretation” *Science Translational Medicine*,11,489,4 (2019)	95	32	10
28	Subramani Mani, Asli Ozdas et al “Medical Decision Support Using Machine Learning for Early Detection of Late-Onset Neonatal Sepsis” *Journal of the American Medical Informatics Association*,21,2,3 (2014)	89	11	22
29	K. Ashwin Kumar, Yashwardhan Singh et al “Hybrid Approach Using Case-Based Reasoning and Rule-Based Reasoning for Domain Independent Clinical Decision Support in ICU” *Expert Systems with Applications*,36,1 (2009)	86	7	27
30	Qiao Li, Gari D. Clifford “Signal Quality and Data Fusion For False Alarm Reduction in the Intensive Care Unit” *Journal of Electrocardiology*,45,6,11 (2012b)	85	9	26

[a]CPY: citations per year.

Discussion

Our study used a bibliometric method to analyze AI in intensive care medicine research by examining publication output, the growth of research interest, preferred journals, leading countries, international collaboration, top institutions, author keywords, and citation analysis.

Publication Output and Growth of Research Interest

Since the oldest publication in 1986, publications on AI in intensive care medicine had a slow growth for 30 years. The turning point appeared in 2018 when there was a significant growth in the interest in AI in intensive care medicine. This lags 6 years behind the rapid growth in the interest in AI in general medicine that started in 2012 [20]. This is likely owing to concerns regarding the safety and accountability of the AI model in critically ill patients. The AI technologies that emerged from 2014 to 2018 such as autonomous robots, voice recognition, neural networks, and ML provided unprecedented opportunities to predict, diagnose, and manage diseases. Large public critical care databases such as MIMIC and eICU became readily available to researchers in 2016 and 2018 [5,21]. Advancements in AI technology and large databases have contributed to the steep increase in annual publications since 2018. Based on the publication trend, it is anticipated that the annual publications will continue to increase.

Preferred Journals

Of the top 15 productive journals, 9 (60%) had a CiteScore of ≥5, which suggested that critical care medicine–related AI research is favored by the top journals in critical care and medical informatics. These include *Critical Care Medicine*, *Critical Care and IEEE Journal of Biomedical and Health Informatics*. Authors who want to publish critical care medicine–related AI research could first consider the top productive journals listed. CiteScore, an Elsevier-Scopus alternative to the Clarivate Analytics Impact Factor, is a metric for assessing journal impact based on citation data from the Scopus database. However, CiteScore is not the only factor considered when deciding which journal to publish in. Authors should consider the ability of the journal to disseminate the research work to the right audience and contribute to the progression of the field [8].

Leading Countries, International Collaboration, and Top Institutions

The fact that 9 of the top 15 productive institutions are among the top 100 best universities demonstrated that AI in critical care medicine has received attention at the top universities worldwide. Authors could consider joint research with those institutions or apply for their visiting scholar or educational programs.

When the distribution of publications by countries was examined, high-income countries were the leading force in critical care medicine–related AI research. The top 10 most productive countries are among the top 25 in terms of world gross domestic product, which suggested that the economic power of the countries affects the productivity of their publications. This result is the same as that of the bibliometric research on many other medical subjects [13,22,23]. About 60% of the global publications were contributed by the United States, China, and the United Kingdom, indicating that these 3 countries contributed the most to AI in critical care research. These countries also had the highest citations, although China had relatively low citations per document compared with the United States and the United Kingdom.

The number of citations was lower than that of other research hot spots in critical care medicine [24]. This is likely because of the limitations of AI-related studies in critical care, which include low maturity of AI in real-world application [25] and a lack of external validation process, prospective evaluation,

XSL•FO
RenderX

and clear protocols to examine the reproducibility of AI solutions [26].

Coauthorship analysis revealed that the United States, the United Kingdom, Italy, and China were the most affiliated countries with >90 coauthorships. The diversity of research partners, a high proportion of foreign postgraduates or visiting scholars, and adequate research funding were all factors that contributed to improved international collaboration. To ensure the sustainability of international collaboration, a flexible and stable research policy is also crucial [8].

Author Keywords and Citation Analysis

According to the author keywords in the identified categories, the top domains of disease covered in critical care medicine–related AI research were sepsis, COVID-19, acute kidney injury, ARDS, and cardiac arrest. These most prevalent ICU conditions have become a popular target for AI algorithms.

The top functions of AI include the prediction of clinical deterioration or disease evolution or mortality, monitoring, disease prognosis, and disease phenotype or subtype classification. The literature reported other important functions such as disease identification and guiding decision-making (reinforcement learning) [26]. The keywords of high occurrence in the titles of the 30 most cited articles include *sepsis*, *prediction*, *early detection*, and *clinical decision support*. Keyword analysis showed that *machine learning* was frequently related to respiratory diseases, especially COVID-19. The most widely used AI technologies in critical care include ML, deep learning, decision support systems, and neural networks.

Limitations

Our research provided a general review of the research trends and hot spots in critical care medicine–related AI research, highlighted the most productive countries and academic institutions to facilitate potential collaboration, and provided directions for future research. However, this bibliometric study has several limitations.

First, even though we included the most widely used AI technologies and made an effort to be specific about AI-related terms in the search keywords (eg, *neural network*, *machine learning*, *deep learning*, and *natural language processing*), they were still quite general and did not include all AI technologies. Because of restricting the search to only those keywords in the title, abstract, and keywords, the search result may not have covered all AI in critical care medicine–related studies available on Scopus. Furthermore, because of missing author keyword information, the co-occurrence analysis of author keywords included only 74.4% (891) of the 1198 articles.

Second, our study used only the Scopus database, which is the largest abstract and citation database that we think should be sufficient for our analysis. Bibliometric analysis using multiple data sources such as the Web of Sciences, PubMed, and Google Scholar will be more comprehensive. For instance, Web of Science has a feature called “hot paper” that is not available in Scopus, which automatically displays the most popular articles in the field [27]. The hot paper feature displays important papers that were identified immediately after publication, as indicated by a sharp rise in the number of citations [8].

Finally, we did not include papers published in the form of conference papers, reviews, editorials, notes, letters, book chapters, short surveys, or data papers owing to concerns about the low clinical readiness of those publications. As a result, we may have missed relevant studies published in forms other than articles. Because AI technology is a cutting-edge and rapidly evolving research area, papers published in conference proceedings and letters may have reviewed the latest updates in the field.

Future bibliometric analyses could use more specific AI technology terms in the research keywords; use other databases such as Web of Sciences, PubMed, and Google Scholar; and include conference papers in the type of articles to explore more potential papers.

Conclusions

Our study has provided an overview of the research trends of AI in critical care medicine based on 1198 publications retrieved from the Scopus database. Publication growth was rapid in the last 5 years and is expected to further increase. We have reviewed countries and academic institutions (eg, the United States, China, and the United Kingdom) that have a substantial number of publications and solid international collaborations. This provides potential collaboration opportunities to other countries, especially to low- and middle-income countries that lack AI technologies but have an increasing demand for health care resources.

We have discussed several conditions in critical care that are currently actively explored using AI technology, such as sepsis, COVID-19, acute kidney injury, ARDS, and cardiac arrest. AI research hot spots in critically ill patients involve detecting clinical deterioration, monitoring, predicting disease evolution, mortality, disease prognosis, and classifying disease phenotypes or subtypes. The most widely used AI technologies in critical care research are ML, deep learning, decision support systems, and neural networks. The 30 articles that received the most citations between 1986 and 2022 have been listed in detail.

AI research on critical care is a rising hot spot in both critical care and AI research, with potential applications being demonstrated across various domains of critical care medicine. However, the development and implementation of AI solutions still face many challenges. AI research application in clinical settings has been constrained by a lack of external validation processes, prospective evaluation, and clear protocols to examine the reproducibility of AI solutions. For AI-based clinical research to be sufficiently convincing for routine critical care practice, collaborative research efforts are needed to increase the maturity and robustness of AI-driven models [26].

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Gutierrez G. Artificial intelligence in the intensive care unit. Crit Care 2020 Mar 24;24(1):101 [FREE Full text] [doi: 10.1186/s13054-020-2785-y] [Medline: 32204716]
2. Greco M, Caruso PF, Cecconi M. Artificial intelligence in the intensive care unit. Semin Respir Crit Care Med 2021 Feb;42(1):2-9. [doi: 10.1055/s-0040-1719037] [Medline: 33152770]
3. Aggarwal N, Ahmed M, Basu S, Curtin JJ, Evans BJ, Matheny ME, et al. Advancing artificial intelligence in health settings outside the hospital and clinic. NAM Perspect 2020 Nov 30;2020:10.31478/202011f [FREE Full text] [doi: 10.31478/202011f] [Medline: 35291747]
4. Artificial Intelligence in Health Care: Benefits and Challenges of Technologies to Augment Patient Care. U.S. Government Accountability Office. 2020 Nov 30. URL: https://www.gao.gov/products/gao-21-7sp [accessed 2022-08-11]
5. Pollard TJ, Johnson AE, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 2018 Sep 11;5:180178 [FREE Full text] [doi: 10.1038/sdata.2018.178] [Medline: 30204154]
6. De Bellis N. Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. Lanham, MD, USA: Scarecrow Press; 2009.
7. Danış F, Kudu E. The evolution of cardiopulmonary resuscitation: global productivity and publication trends. Am J Emerg Med 2022 Apr;54:151-164. [doi: 10.1016/j.ajem.2022.01.071] [Medline: 35152126]
8. Md Khudzari J, Kurian J, Tartakovsky B, Raghavan G. Bibliometric analysis of global research trends on microbial fuel cells using Scopus database. Biochem Eng J 2018 Aug;136:51-60. [doi: 10.1016/j.bej.2018.05.002]
9. Van Eck NJ, Waltman L. VOSviewer Manual. Universiteit Leiden. 2018 Apr 27. URL: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.8.pdf [accessed 2022-08-07]
10. CAGR Calculator (Compound Annual Growth Rate). Omni Calculator. URL: https://www.omnicalculator.com/finance/cagr#the-simple-growth-rate-formula [accessed 2022-08-07]
11. Garrett SJ. An Introduction to the Mathematics of Finance: A Deterministic Approach. 2nd edition. Oxford, UK: Butterworth-Heinemann; 2013.
12. Cipra T. Financial and Insurance Formulas. Heidelberg, Germany: Springer; 2010.
13. Wan R, Li L, Xing C, Peng R, Gao L. Worldwide scientific productions with immunotherapy of sepsis: a bibliometric analysis. PeerJ 2019 Jun 17;7:e7116 [FREE Full text] [doi: 10.7717/peerj.7116] [Medline: 31245180]
14. Fardi A, Kodonas K, Gogos C, Economides N. Top-cited articles in endodontic journals. J Endod 2011 Sep;37(9):1183-1190. [doi: 10.1016/j.joen.2011.05.037] [Medline: 21846531]
15. Scopus - Sources. Scopus preview. URL: https://www.scopus.com/sources.uri?RN_AG_Sourced_400000654 [accessed 2022-08-12]
16. MyMaps. Google Maps. URL: https://www.google.com/maps/d/ [accessed 2022-08-07]
17. List of countries by GDP (nominal). Wikipedia - The Free Encyclopedia. URL: https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal) [accessed 2022-08-07]
18. Lutz H, Bender HJ. Möglichkeiten der computeranwendung in anästhesie und intensivmedizin. Anasthesiol Intensivmed Notfallmed Schmerzther 1986;21(02):68-71. [doi: 10.1055/s-2007-1002449]
19. World University Rankings. Times Higher Education (THE). URL: https://www.timeshighereducation.com/world-university-ránkings [accessed 2022-08-10]
20. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res 2020 Jul 29;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]
21. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
22. Vuillemin N, Pape HC, Rommens PM, Lippuner K, Siebenrock KA, Keel MJ, et al. A bibliometric analysis of fragility fractures: top 50. Medicina (Kaunas) 2021 Jun 21;57(6):639 [FREE Full text] [doi: 10.3390/medicina57060639] [Medline: 34205638]
23. Klingelhöfer D, Braun M, Schöffel N, Brüggmann D, Groneberg DA. Gastric cancer: bibliometric analysis of epidemiological, geographical and socio-economic parameters of the global research landscape. Int J Health Policy Manag 2021 Mar 01;10(3):118-128 [FREE Full text] [doi: 10.34172/ijhpm.2020.29] [Medline: 32610747]
24. Zhang Z, Van Poucke S, Goyal H, Rowley DD, Zhong M, Liu N. The top 2,000 cited articles in critical care medicine: a bibliometric analysis. J Thorac Dis 2018 Apr;10(4):2437-2447 [FREE Full text] [doi: 10.21037/jtd.2018.03.178] [Medline: 29850150]
25. Fleuren LM, Thoral P, Shillan D, Ercole A, Elbers PW, Right Data Right Now Collaborators. Machine learning in intensive care medicine: ready for take-off? Intensive Care Med 2020 Jul;46(7):1486-1488. [doi: 10.1007/s00134-020-06045-y] [Medline: 32399747]

XSL•FO
RenderX

26. Yoon JH, Pinsky MR, Clermont G. Artificial intelligence in critical care medicine. Crit Care 2022 Mar 22;26(1):75 [FREE Full text] [doi: 10.1186/s13054-022-03915-3] [Medline: 35337366]
27. Birkle C, Pendlebury DA, Schnell J, Adams J. Web of Science as a data source for research on scientific and scholarly activity. Quant Sci Stud 2020 Feb 1;1(1):363-376. [doi: 10.1162/qss_a_00018]

Abbreviations

AI: artificial intelligence
ARDS: acute respiratory distress syndrome
ICU: intensive care unit
ML: machine learning

Edited by T Leung; submitted 27.08.22; peer-reviewed by M Kapsetaki, Y Wang; comments to author 08.10.22; revised version received 23.10.22; accepted 31.10.22; published 30.11.22.

Please cite as:
Tang R, Zhang S, Ding C, Zhu M, Gao Y
Artificial Intelligence in Intensive Care Medicine: Bibliometric Analysis
J Med Internet Res 2022;24(11):e42185
URL: https://www.jmir.org/2022/11/e42185
doi:10.2196/42185
PMID:36449345

©Ri Tang, Shuyi Zhang, Chenling Ding, Mingli Zhu, Yuan Gao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Uncovering the Reasons Behind COVID-19 Vaccine Hesitancy in Serbia: Sentiment-Based Topic Modeling

Adela Ljajić[1*], PhD; Nikola Prodanović[1*], PhD; Darija Medvecki[1*], MSc; Bojana Bašaragin[1*], PhD; Jelena Mitrović[1,2], PhD

[1]The Institute for Artificial Intelligence Research and Development of Serbia, Novi Sad, Serbia

[2]Faculty of Computer Science and Mathematics, University of Passau, Passau, Germany

[*]these authors contributed equally

Corresponding Author:
Adela Ljajić, PhD
The Institute for Artificial Intelligence Research and Development of Serbia
Fruškogorska 1
Novi Sad, 21000
Phone: 381 652626347
Email: adela.ljajic@ivi.ac.rs

Abstract

Background: Since the first COVID-19 vaccine appeared, there has been a growing tendency to automatically determine public attitudes toward it. In particular, it was important to find the reasons for vaccine hesitancy, since it was directly correlated with pandemic protraction. Natural language processing (NLP) and public health researchers have turned to social media (eg, Twitter, Reddit, and Facebook) for user-created content from which they can gauge public opinion on vaccination. To automatically process such content, they use a number of NLP techniques, most notably topic modeling. Topic modeling enables the automatic uncovering and grouping of hidden topics in the text. When applied to content that expresses a negative sentiment toward vaccination, it can give direct insight into the reasons for vaccine hesitancy.

Objective: This study applies NLP methods to classify vaccination-related tweets by sentiment polarity and uncover the reasons for vaccine hesitancy among the negative tweets in the Serbian language.

Methods: To study the attitudes and beliefs behind vaccine hesitancy, we collected 2 batches of tweets that mention some aspects of COVID-19 vaccination. The first batch of 8817 tweets was manually annotated as either relevant or irrelevant regarding the COVID-19 vaccination sentiment, and then the relevant tweets were annotated as positive, negative, or neutral. We used the annotated tweets to train a sequential bidirectional encoder representations from transformers (BERT)-based classifier for 2 tweet classification tasks to augment this initial data set. The first classifier distinguished between relevant and irrelevant tweets. The second classifier used the relevant tweets and classified them as negative, positive, or neutral. This sequential classifier was used to annotate the second batch of tweets. The combined data sets resulted in 3286 tweets with a negative sentiment: 1770 (53.9%) from the manually annotated data set and 1516 (46.1%) as a result of automatic classification. Topic modeling methods (latent Dirichlet allocation [LDA] and nonnegative matrix factorization [NMF]) were applied using the 3286 preprocessed tweets to detect the reasons for vaccine hesitancy.

Results: The relevance classifier achieved an *F*-score of 0.91 and 0.96 for relevant and irrelevant tweets, respectively. The sentiment polarity classifier achieved an *F*-score of 0.87, 0.85, and 0.85 for negative, neutral, and positive sentiments, respectively. By summarizing the topics obtained in both models, we extracted 5 main groups of reasons for vaccine hesitancy: concern over vaccine side effects, concern over vaccine effectiveness, concern over insufficiently tested vaccines, mistrust of authorities, and conspiracy theories.

Conclusions: This paper presents a combination of NLP methods applied to find the reasons for vaccine hesitancy in Serbia. Given these reasons, it is now possible to better understand the concerns of people regarding the vaccination process.

(J Med Internet Res 2022;24(11):e42261) doi:10.2196/42261

KEYWORDS

topic modeling; sentiment analysis; LDA; NMF; BERT; vaccine hesitancy; COVID-19; Twitter; Serbian language processing; vaccine; public health; NLP; vaccination; Serbia

XSL•FO
RenderX

Introduction

Background

The COVID-19 pandemic has significantly disrupted the daily lives of individuals and the way in which organizations operate worldwide. One of the most effective strategies to tackle the COVID-19 pandemic is to achieve collective immunity through mass vaccination [1,2]. However, people have shown significant resistance and hesitancy to the global immunization process [3]. The World Health Organization (WHO) identified vaccine hesitancy as 1 of the top 10 threats to global health care in 2019 [4]. Therefore, the study of the public attitude toward the vaccination process is of utmost importance. In particular, it is useful to identify the prevailing beliefs and attitudes that may lead to a negative sentiment toward vaccination. According to WHO, many events have the potential to erode confidence in vaccines. Some of them are related to vaccine safety and adverse events following immunization, but some are related to social media stories or rumors [5], making it significant to analyze the beliefs, attitudes, and fears reflected in the user-generated content on social media.

This kind of research needs to be conducted regionally worldwide since attitudes of people from different world regions vary significantly [3,6]. This presents a fair challenge as numerous languages of small populations worldwide, Serbian being among them, lack electronic resources. Due to the rapid advancement of artificial intelligence and machine natural language processing (NLP), we believe it is now possible to tackle this challenge, and thus demonstrate a possible solution for the case of Serbian as an example. The main contribution of our work is in the application of a combination of NLP methods to a low-resourced language to discover hidden topics related to vaccine hesitancy with minimum data annotation.

The research community predominantly used Twitter to collect data on COVID-19 vaccination [7-22]. We also opted for this social media since this is the platform where users produce large amounts of data that can be used for analysis of perceptions and narratives [23], collective experiences, behaviors, and attitudes related to particular social events [24]. Additionally, Twitter provides an application programming interface (API) that enables easier extraction of data compared to other platforms [25]. The use of this API allowed us to collect 14,452 tweets related to vaccination in the Republic of Serbia. The collected data span from January 2021, right after the first COVID-19 vaccines got released, to June 2022. The goal of our research is to look for topics in the tweets that express negative attitudes toward vaccination, which we believe would be most revealing with regard to the reasons for vaccine hesitancy.

A part of the data set was manually annotated using 4 class labels: irrelevant, positive, negative, and neutral. This data set was used to train a sequential bidirectional encoder representations from transformers (BERT)-based classifier, which then served to automatically annotate the rest of the data. After gathering the set of tweets with a relevant and clear negative sentiment toward vaccination through both manual and automatic annotation, we conducted topic analysis in order to pinpoint the main reasons for vaccine hesitancy.

The aim of this study is to detect the main topics within tweets in Serbian that express a negative sentiment regarding COVID-19 vaccination under an assumption that these topics point to the main reasons for vaccine hesitancy in Serbia. This information can help local domain experts influence the public in a more informed way with regard to vaccination. Knowing why people, especially young people, are hesitant equips the key decision makers with the right tools for planning vaccination-oriented campaigns.

Related Work: Tweet Classification

The length and impact of the COVID-19 pandemic led to a surge in user-generated pandemic-related content on Twitter. The ability to automatically classify that content using machine learning and deep learning methods became especially important when information about COVID-19 vaccines started appearing. Previous work on sentiment analysis and human papillomavirus vaccination [26-28], and vaccination in general [11,12], served as a base for research into automatic classification of the sentiments of COVID-19–related tweets.

In recent years, there has been a significant shift in the design of machine learning architecture for the purpose of short text classification. With regard to public opinions about vaccination, the most traditionally exploited idea is that of static text embeddings combined with classical machine learning methods [11,12]. Relatively recently, systems based on recurrent neural networks (RNNs) started being used for such purposes [27,28]. A new family of methods based on attention neural networks was introduced in 2017. Their self-attention mechanism efficiently captures long-range dependencies through the pretraining process by maximally using parallel computation algorithms and hardware [29]. This gives this method a significant advantage over its predecessors based on RNNs to produce context and morphosyntactic aware embeddings. Historically, the sequence-to-sequence transduction model was the original model with the attention mechanism [29], but soon after, the first encoder-only architecture capable of providing only embeddings was published under the acronym BERT [30].

With the rise in computational power, many researchers were able to apply BERT to COVID-19 and vaccination content in English and test its results against older methods, such as bidirectional long-short term memory, support vector machines, and naïve Bayes. BERT-based architecture proved to be superior both for binary sentiment, relevance, or misinformation classification [9,13,19,28] and for tertiary stance or sentiment classification [14,17,19], which prompted us to choose such architecture for our research.

The pretraining strategy for BERT is usually defined as a masked language modeling task, which resembles the autoencoders, and a next sentence prediction task [30]. The most recent proposal for a pretraining strategy is the Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA) approach, where the BERT model is trained as a discriminator rather than a generator. This method was used to train BERTić [31], the first BERT-based model for South Slavic languages and the model we used to develop our classifiers.

XSL•FO
RenderX

BERTić has already been tested on tasks of short text classification for Serbian. Batanović [32] compared the results of BERT and BERTić to several linear classifiers on different classification tasks for movie reviews and showed that BERTić was the most optimal model for the tasks of binary and 4-class polarity classification. Mochtak et al. [33] worked on the tasks of ternary (negative-positive-neutral) and binary (negative and other) classification of sentences from parliamentary proceedings for Croatian, Serbian, and Bosnian. They tested several models: fastText with pretrained CLARIN.SI word embeddings, Cross Lingual Model – Roberta (XLM-Roberta), cseBERT, and BERTić. The best results were obtained with BERTić for all 3 languages. To the best of our knowledge, our work is the first attempt to apply BERTić to the classification of tweets in Serbian.

Related Work: Tweet Topic Modeling

Since the beginning of the COVID-19 pandemic, researchers have attempted to use topic modeling to determine public attitudes toward various aspects of the pandemic [7,10,34], particularly vaccination [8,15,16,20-22,34-36]. Topic modeling is a method that allows grouping of documents into a predetermined number of topics. As a method that does not require any supervision or prior data labeling, it is popular for detecting hidden attitudes in a large variety of documents. Historically devised for longer texts, topic modeling has been confronted in recent years with the challenge of unveiling topics in short, unstructured, and informal social media comments [37]. Despite proposing methods to specifically tackle short text [38,39], and aggregating shorter texts into pseudodocuments before applying topic modeling [40-42], classical topic modeling methods, such as latent Dirichlet allocation (LDA) [43] and nonnegative matrix factorization (NMF) [44], remain the preferred methods when tackling tweets and social media comments in general.

LDA is a generative probabilistic model for collection of discrete data and is therefore used for discovering latent semantic structures from text corpora by capturing the pattern of co-occurrence of words at the document level. It has been especially widely used during the COVID-19 pandemic to determine the most discussed topics [7,10], correlate the vaccination stance and events in the media [8,17] or other spatiotemporal factors [16,36] and determine vaccine hesitancy topics [21,35], the general sentiment toward COVID-19 vaccines [20], and its changes over time [15].

NMF is a nonprobabilistic method based on matrix decomposition actively used for topic modeling [44,45]. It has also been applied to the theme of COVID-19 to determine the main pandemic health effects [34] and the public sentiment toward vaccination [22]. Compared to LDA, which gives more general descriptions of broader topics [46], the architecture of NMF enables it to find more detailed, clear-cut, and coherent topics [37,46,47]. Chen et al [18] even claim that NMF can learn from data similarly to the way humans do, which makes its results more easily interpretable than in the case of LDA.

Given that the 2 models approach the data and the topics differently, we decided to use a combination of their results in order to determine the final list of topics in our research.

Even though substantial work has been conducted on sentiment analysis for Serbian [48-52], to the best of our knowledge, this is the first attempt to apply topic modeling to Serbian.

Methods

Study Design

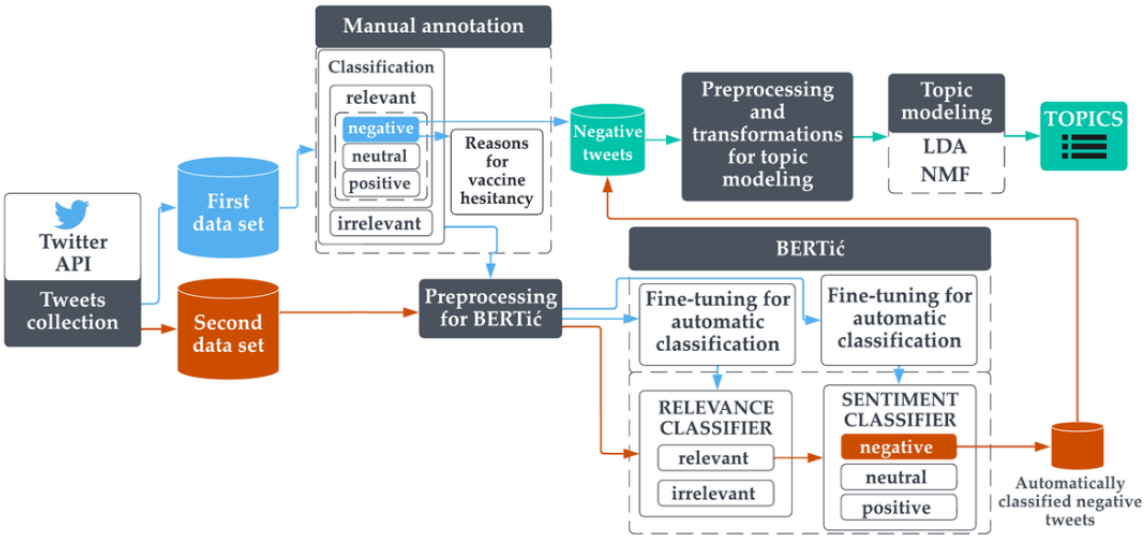
To study the attitudes and beliefs behind vaccine hesitancy, we first collected 2 batches of tweets that mention some aspect of COVID-19 vaccination. We manually annotated the first set of tweets as either relevant or irrelevant with regard to the COVID-19 vaccination sentiment and then annotated the relevant ones as positive, negative, or neutral. In addition, we manually searched for topics related to vaccine hesitancy in the negative tweets.

To augment this initial data set, we used the annotated tweets to train a sequential BERT-based classifier for 2 tweet classification tasks. In the first task, the classifier distinguished between relevant and irrelevant tweets. In the second task, the classifier took the relevant tweets as input and classified them as negative, positive, or neutral. We used this sequential classifier to annotate the second batch of tweets. We then combined the 2 data sets and applied 2 topic modeling methods (LDA and NMF) to them in order to detect the reasons for vaccine hesitancy.

This entire pipeline is presented in Figure 1. Each of the individual steps is described in detail in the following subsections.

XSL•FO
RenderX

Figure 1. Tweet classification and topic modeling pipeline. API: application programming interface; BERT: bidirectional encoder representations from transformers; LDA: latent Dirichlet allocation; NMF: nonnegative matrix factorization.



Data Collection and Annotation

We used the *Twarc* Python library [53] to extract the data in Serbian (in Cyrillic and Latin scripts) from the Twitter streaming API. The collection of tweets was divided into 2 phases, resulting in 2 subsets of data.

The first data set consisted of 8817 tweets collected between January 1 and November 23, 2021. Since the purpose of this data set was to reflect the opinions and topics of the citizens of Serbia, the query included the condition that the tweets either contain the location of the Republic of Serbia or be written in Serbian. We first tested the search using relevant hashtags (#COVID-19, #vakcina, etc), which did not yield enough tweets, because hashtags with Serbian words on this topic are not frequently used. For that reason, we based our search on keywords relevant to the topic of vaccination. The query consisted of all the writing and morphological variations for COVID-19 mutually connected with an OR operator (eg, “COVID-19” OR “corona” OR “kovid”) and all the writing and morphological variations for the words “vaccine” and “vaccination,” including vaccine types (“vakcina” OR “moderna” OR “fajzer”) in Latin and Cyrillic scripts. This enabled a search of all the tweets that were related to both COVID-19 and vaccines. Retweets were excluded from the search.

This entire data set needed to be annotated in order to train the classifiers. We compiled a detailed set of rules according to which the annotators conducted the labeling. The labels for the positive and negative sentiments were assigned to tweets with the respective type of attitude toward vaccination. A neutral sentiment was used for neutral attitudes about the topic but also for tweets that did not convey an explicit attitude of the user but contained some information about the topic. This included facts about COVID-19 vaccination, available doses or vaccination dates, objective questions about vaccination backed by the user’s obvious intention to seek other people’s opinion and information, jokes without attitude, and posting of neutral media headlines without additional personal comments. Furthermore, the annotators used a special class for irrelevant tweets, such as those containing an unclear or vague attitude. This class also included tweets that consisted of an external link and some user comments related to the content of the link, which was not sufficient to capture their attitude toward vaccination, because the links were not the subject of this analysis. The subjects of the annotation were text content and hashtags.

For the first 500 (5.7%) tweets, all the authors of this paper conducted the labeling and amended the initial set of rules through mutual discussion on the confused examples. The rest of the data set was individually and separately annotated by 2 annotators using the defined guidelines. After the whole data set was labeled, the Cohen κ score was 0.57 for all 4 classes, 0.67 for the 3 sentiment classes, and 0.73 for the positive and negative classes. The main point of disagreement between the annotators was in assigning the “neutral” versus the other 2 sentiment labels and the “irrelevant” versus the “relevant” label (positive, negative, and neutral), which was resolved by an author of this paper who was most involved in the COVID-19 vaccine discussion. The result was a data set of 5791 (65.7%) relevant tweets (irrelevant tweets=3026, 34.3%), divided into 3 sentiment classes. The statistics of the first subset can be seen in Table 1.

In addition to defining the sentiments of relevant tweets, the annotators separately indicated the topics that were prevalent in the negative tweets. The number of these topics was later used to set the upper limit for testing the optimal number of topics for the topic modeling methods.

The second subset of data was collected for the period from November 23, 2021, to June 6, 2022. After the first phase of tweet collection, we concluded that filtering the tweets by specifying the location and the Serbian language severely limited the number of tweets available for collection, so we decided to take a different approach.

XSL•FO
RenderX

Table 1. Vaccine hesitancy data set statistics for the relevant tweets in batch 1 (N=5791).

Sentiment class	Tweets, n (%)
Negative	1770 (30.6)
Positive	1965 (33.9)
Neutral	2056 (35.5)

Since the search condition regarding location can only be satisfied if the user shares the location at the time the tweet is published, which does not often seem to be the case for people from Serbia, this operator significantly limits the collection of tweets and excludes many potential results. Several problems occur when using the language operator. When Serbian is specified as the language, Cyrillic is the default script, so the collection of tweets written in Latin is omitted, as noticed in Ref. [49]. In fact, the Twitter API sorts out most of these tweets as an undefined language. In addition, some of the tweets collected in Cyrillic are in Northern Macedonian instead of Serbian. Therefore, we decided to exclude these 2 operators this time. As a result, our initial data set contained tweets in languages close to Serbian (Russian, Czech, Northern Macedonian, etc), which we filtered out using the language recognition library for Python *langID* [54].

This clean data subset consisted of 5635 tweets in Serbian. As this subset was meant to be used to test the performance of our classification model, it was not labeled by human annotators. The total number of tweets in both batches was 14,452.

Automatic Tweet Classification

Deciding which tweet contains a negative sentiment is not a straightforward task. In our data set described in the previous subsection, about two-thirds of the total number of gathered tweets have an attitude toward vaccination, and only a subset of these tweets has a negative sentiment. We assumed that our data set was representative enough and therefore concluded that any further pipeline must contain automatic filtration of tweets into negative-sentiment tweets with sufficient relevancy in order to be able to automatically detect a large number of negative tweets for further analysis. With this in mind, we decided to develop a deep learning classifier that could detect relevant tweets with a sufficiently clear negative attitude toward the vaccination process. To build both classifiers, we used BERTić, a BERT-based model for South Slavic languages [31]. Instead of pretraining BERT from scratch on a much larger corpus of tweets [55], we used the annotated data to fine-tune and test BERTić on a downstream task of short text classification.

The classifier consists of 2 sequential parts. The first part filters tweets based on their relevance to the topic, and the second part filters tweets based on their sentiment. The second classifier takes as input the tweets that have passed the first filter for relevancy. We considered unifying these 2 classifiers into a single BERT architecture with an increased number of classes but abandoned this idea due to prominent class imbalance. The most interesting discussion arose for the boundary between irrelevant tweets and neutral-sentiment tweets. This boundary had to be introduced clearly through the annotation process. It was intuitively clear that class separation efficiency between the neutral class and the positive and negative classes would be sharper if we forced training only on the tweets that indeed had vaccines as the main topic but had no clear sentiment. This was our main reasoning behind the serialization of the classifiers.

The minimum preprocessing steps that we took before the training consisted of switching to the Latin script for all the tweets (using the *srtools* Python library [56]); restoring the diacritics (using the *classla* Python library [57]); removing the mentions, links, emojis, and noninformative hashtags; and transforming the remaining hashtags into words using regular expressions. We trained our algorithm on only 1 iteration of the annotation process because we also wanted to analyze possible human annotation errors and the robustness of the algorithm to the quality of annotation.

For the relevance classifier, the annotated data set was split into training, validation, and test sets according to the 80%:10%:10% ratio. The total number of examples in this data set was 8817. The validation set was used to choose the most optimal network solution among the maximum number of 6 training epochs.

For the sentiment polarity classifier, we developed a set of 5791 relevant tweets, which we split according to the 80%:10%:10% training:validation:test ratio.

The number of epochs and batch size were chosen to be optimal for a fixed validation set, which may result in a slight but acceptable bias. This is justified by the recommended values of these hyperparameters given in the original paper describing the BERT model [30], namely 4 epochs and a training batch size of 16 tweets.

Topic Modeling

To uncover the reasons for vaccine hesitancy, we used 2 topic modeling methods on the data set of negative tweets: LDA and NMF. We decided to use these 2 models to compare the topics generated by completely different approaches.

For LDA, we used the implementation of Hoffman et al [58] and an open source *Gensim* Python library [59]. For NMF, we used the *sklearn* NMF decomposition the way it was implemented by Cichocki and Phan [60].

Before applying the topic modeling methods, we needed to go through several preprocessing steps to remove noise and reduce the space for topic modeling. The preprocessing pipeline consisted of switching from Cyrillic to Latin script; removal of URLs, mentions, numbers, new lines, emojis, images, special characters, etc; tokenization; lemmatization; and removal of stop words. We converted the tweets to Latin script using the *srtools* Python library, while tokenization and lemmatization were conducted using the *classla* pipeline for nonstandard Serbian. We removed the URLs, mentions, etc, using regular expressions. We used the list of stop words described by Marovac et al [61], which we extended with all the alternative

XSL•FO
RenderX

names for COVID-19 and derivatives of the word "vaccine." These terms naturally appear in most tweets since we applied them as our Twitter search keywords.

Building the Models

Both LDA and NMF require certain data set transformations. The transformations required to create the LDA model first include the creation of a vocabulary in the form of a list of unique words represented as integers. The next step is the pruning process: removing low- and high-frequency words. The final step is creating a corpus of all tweets as bag-of-words features. After these initial steps, we applied filters that excluded all the words that appeared in less than 3 tweets and more than 85% of tweets and limited the dictionary to 1000 terms. We chose to limit the dictionary since using more than 1000 terms resulted in less coherent topics. Additionally, a large dictionary allowed for less significant words to become more significant inside topic keywords due to the inability to quantify the importance of words.

For the NMF model, we used the term frequency–inverse document frequency (TF-IDF) transformation of the normalized text and applied the same filters as for LDA: we excluded words that appeared in less than 3 tweets and more than 85% of tweets and limited the dictionary to 1000 terms. We experimented with using several different combinations of filters for both models, which did not lead to significant changes in topics for the NMF model, but it did in the case of LDA. In general, NMF showed greater topic stability with the change in the dictionary size.

Each of the topic modeling methods requires a predefined number of topics. We calculated that number by tuning the model parameters and choosing the number of topics and parameters that yielded the highest coherence score value (c_v). The c_v score ranges from 0 to 1 and measures the co-occurrence of words in a topic inside the corpus. We opted for c_v as a metric since it increases monotonously with an increase in the number of topics, unlike another customarily used topic similarity metric, u_mass, which reaches the peak for a smaller number of topics and then decreases with an increase in the number of topics. When testing the models for the number of topics, we set the parameter α to "auto," which made the model learn an asymmetric prior from the corpus.

In addition to c_v, we used another similarity metric, namely the Jaccard similarity coefficient. The Jaccard similarity coefficient ranges from 0 to 1 and measures the topic overlap. The lower the Jaccard similarity coefficient and the higher the c_v value, the more optimal the number of topics. Since c_v increases with an increase in the number of topics, which was not proven adequate for our data set, we applied the Jaccard similarity coefficient to normalize the number of topics. We set the limit for the optimal number of topics for both models to 15, as that was also the number of topics initially identified by human annotators.

After applying both c_v and Jaccard similarity coefficient metrics, the resulting optimal number of topics for LDA proved to be 14 (see Figure 2).

To obtain cluster assignments, LDA uses 2 probability values: *P*(word|topics) and *P*(topics|documents). In the *Gensim* model, parameters α and β affect these 2 probabilities. The α parameter is an a priori belief on document-topic distribution, while β is an a priori belief on topic-word distribution. After determining the optimal number of topics, we tuned these 2 parameters to obtain the best distribution of keywords per topic (see Figure 3). We made the model for the first 5 best-ranking combinations of α and β, and by manually comparing the topics, we chose the second one as best, which was α="asymmetric" and β=0.91. A high value of β means that the topic can be assigned to more words. This was justified, given the nature of the data set focused on a narrow field where the same words often appear in different contexts, which makes the topics more similar based on the words they contain.

After applying c_v and Jaccard similarity coefficient metrics, the resulting optimal number of topics for NMF proved to be 13 (see Figure 4).

For the NMF model, we used an input document-term matrix normalized with TF-IDF. The matrices into which the starting document-term matrix is decomposed are document-topic and topic-term matrices. We obtained the starting values of these 2 matrices by using singular value decomposition initialization presented in Belford et al [62], which is suitable for sparse data. For the fast convergence rate, we used coordinate descent solver-cd in *sklearn*. We tested the κ parameter, which determines the model convergence speed, and concluded it did not significantly affect coherence (see Figure 5). We chose a κ learning rate of 0.1, limited the number of iterations to 500, and set the random state to 42. We used the default value of 1e-4 for the tolerance of the stopping condition, and we did not use regularization parameters.

XSL•FO
RenderX

Figure 2. Optimal number of topics according to the coherence score value (c_v) and the Jaccard similarity coefficient for LDA. LDA: latent Dirichlet allocation.

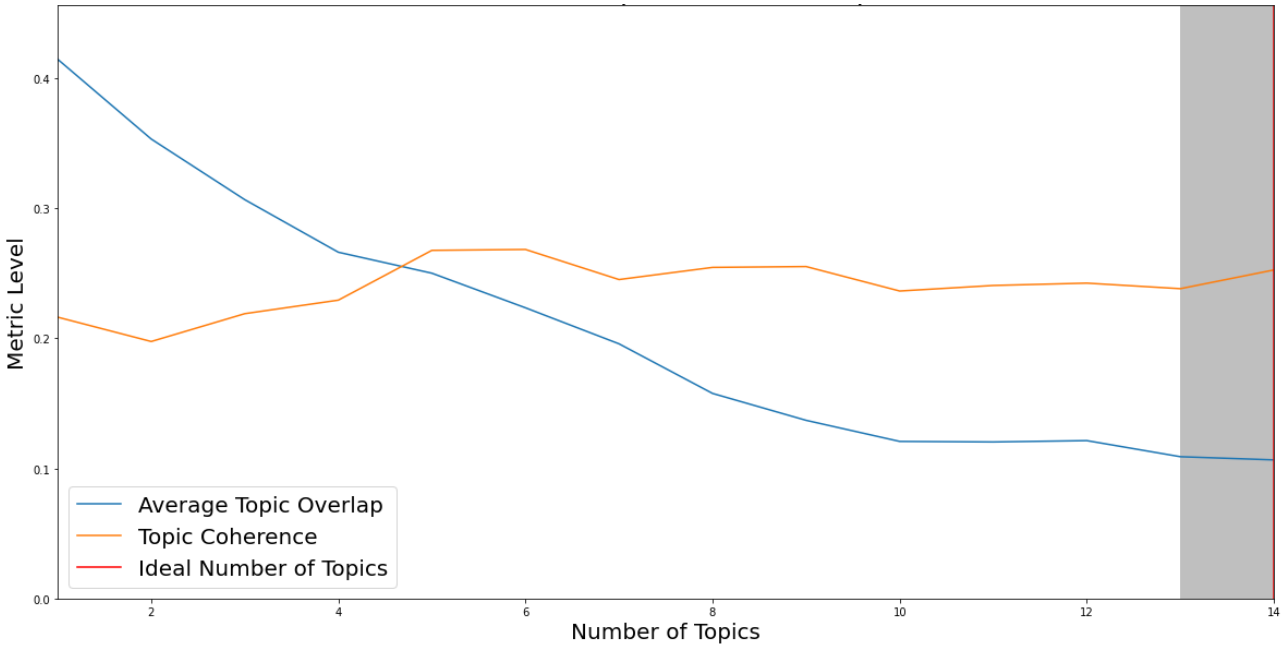
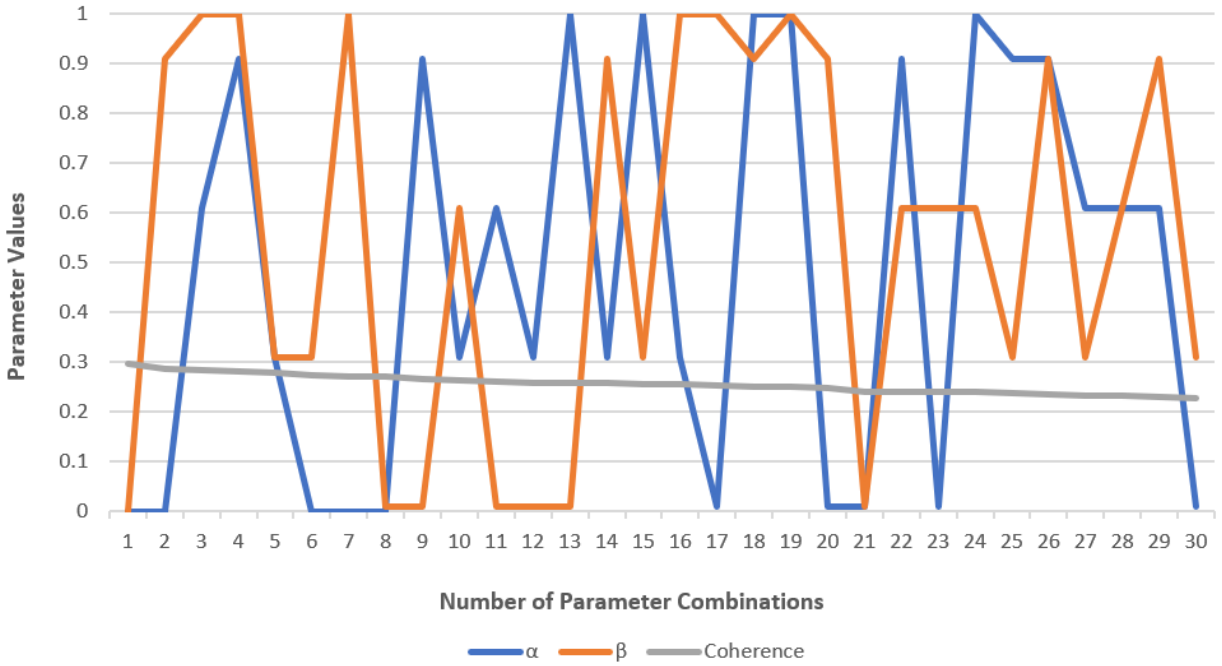


Figure 3. The c_v score for different values of α and β for 14 LDA topics. The "asymmetric" value is represented as 0 and the "symmetric" value as 1. LDA: latent Dirichlet allocation.



XSL•FO
RenderX

Figure 4. Optimal number of topics according to c_v and the Jaccard similarity coefficient for NMF. NMF: nonnegative matrix factorization.

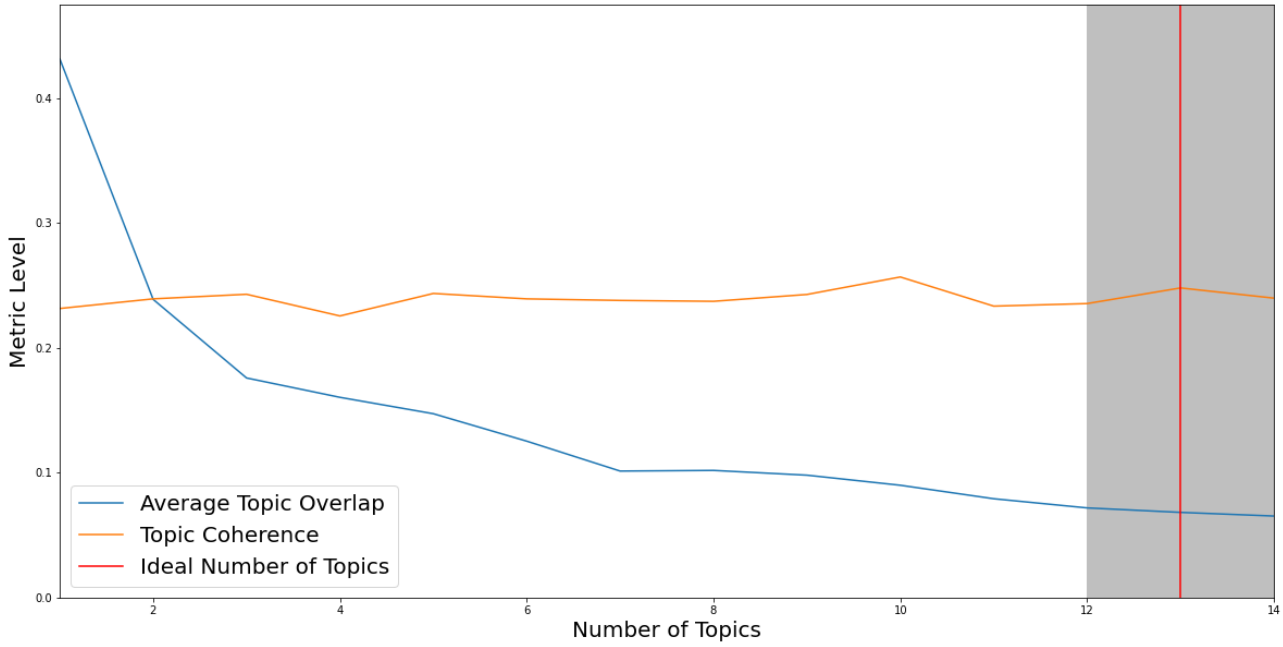
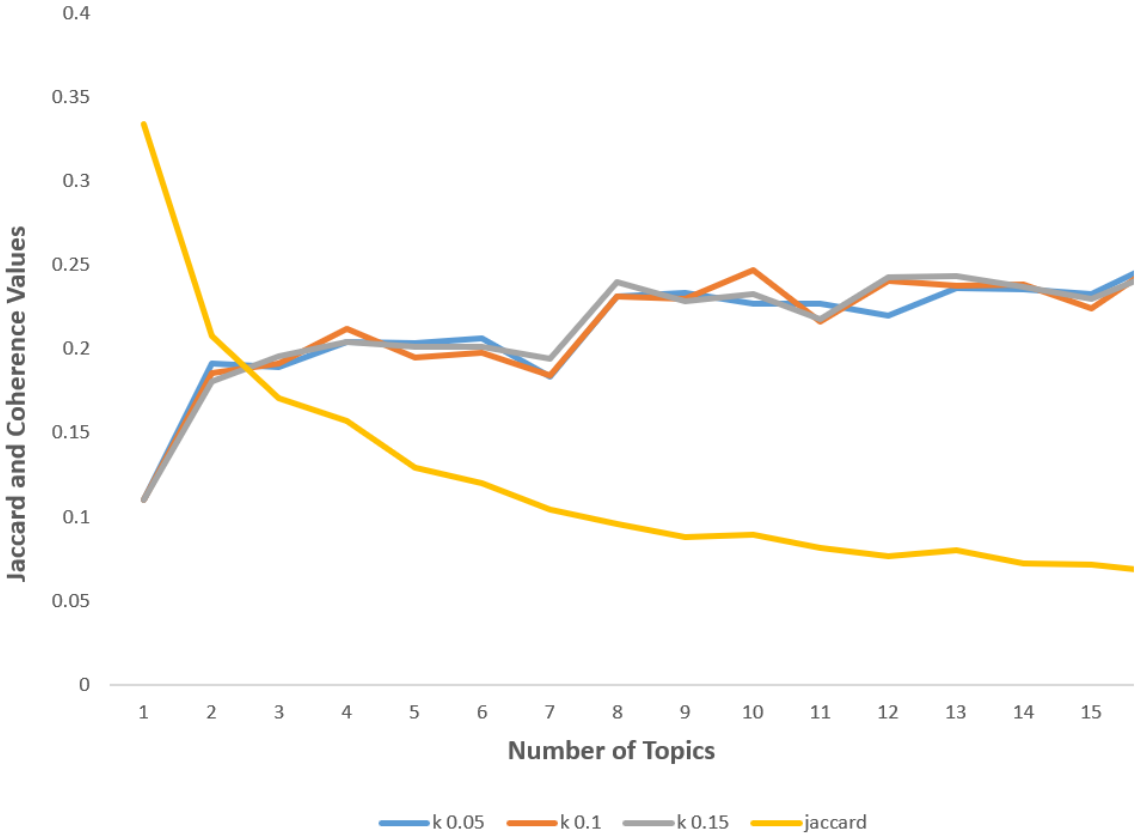


Figure 5. The c_v score and the Jaccard similarity coefficient for different learning rates (NMF). NMF: nonnegative matrix factorization.



Results

We grouped the individual results of automatic classification and topic modeling into 2 separate subsections, a*utomatic tweet classifier* and *topic modeling*.

Automatic Tweet Classifier

We designed a sequential tweet classifier consisting of 2 BERTić classifiers. The first classifier was binary, and it decided whether a tweet was relevant for further analysis, while the second classifier performed the task of ternary classification and decided the type of sentiment associated with the tweet.

XSL•FO
RenderX

Relevance Classifier

The relevance classifier detected whether a tweet was relevant enough to be considered as an opinion about vaccines. Usually, irrelevant tweets are strongly related to epidemics and politics but without a clear attitude toward vaccination. We found that the political attitudes of Twitter users often mask attitudes toward vaccination. We decided to label extremely complex examples with completely masked attitudes as irrelevant, because it was obvious that users were frustrated by some other issues rather than by vaccination itself.

The algorithm was tested on 10% of the total number of tweets, which in this case was 882 tweets. The outright accuracy was 94.7%. The irrelevant class was imbalanced according to the 35%:65% ratio. However, after test set reannotation, many of the tweets were labeled as relevant, which shifted this imbalance below 30% for the irrelevant class; thus, we obtained lower *F*- and recall scores for the irrelevant class, valued at 0.91 and 0.86, respectively. The *F*-score for the relevant class was above 0.96. All the scores can be seen in Table 2.

The biggest issue was to come to a conclusion about the exact semantic boundary between the irrelevant tweets and the relevant tweets with a neutral sentiment. A neutral sentiment may also be understood as no sentiment, and thus irrelevant.

Table 2. Confusion matrix and *F*-scores for the relevance classifier.

Class	Irrelevant (predicted)	Relevant (predicted)
Irrelevant	225	35
Relevant	12	610
F-score	0.91	0.96

Sentiment Polarity Classifier

The sentiment polarity classifier took as input only relevant tweets and output their sentiment toward vaccination.

The accuracy of the model on the test set was about 85.7% (see Table 3).

Most of the confused examples fell between the neutral and the other 2 classes. Recall was the lowest for the positive class, with a value of 0.82. By careful inspection, we found no systematic error tendency for the algorithm or the annotators to confuse the positive class. Thus, the lower recall for the positive class is a consequence of a slightly imbalanced data set against the number of positive examples, as can be seen from Table 3.

Table 3. Confusion matrix and *F*-scores for the sentiment classifier.

Class	Negative (predicted)	Neutral (predicted)	Positive (predicted)
Negative	166	17	6
Neutral	18	197	12
Positive	10	20	134
F-score	0.87	0.85	0.85

Topic Modeling

We performed topic modeling using a total of 3286 preprocessed tweets with a negative sentiment: 1770 (53.9%) tweets came from the manually annotated data set, and another 1516 (46.1%) tweets came as a result of automatic classification. We made this data set available on our GitHub repository [63].

The average word count in the data set was 22, with an SD of 8 words. The word count distribution in negative tweets can be seen in Figure 6. The distribution was slightly negatively skewed, but overall, it was a normal distribution, with the 25th percentile at 16 words and the 75th percentile at 28 words.

The text length distribution can be seen in Figure 7. It was also negatively skewed but more significantly than the word count distribution, with an average length of 152 characters and an SD of 53 characters. The length of tweets was often connected with the nature of the negative sentiment, which affected the grouping of such tweets into a certain topic.

Figure 8 displays the 20 most frequent words in the preprocessed data set. The top 20 words included the terms “virus,” “fraud,” and “experiment,” proving that the most frequent words reflect the nature of the data set consisting of tweets with a negative sentiment regarding vaccination.

XSL•FO
RenderX

Figure 6. Tweet word count distribution.

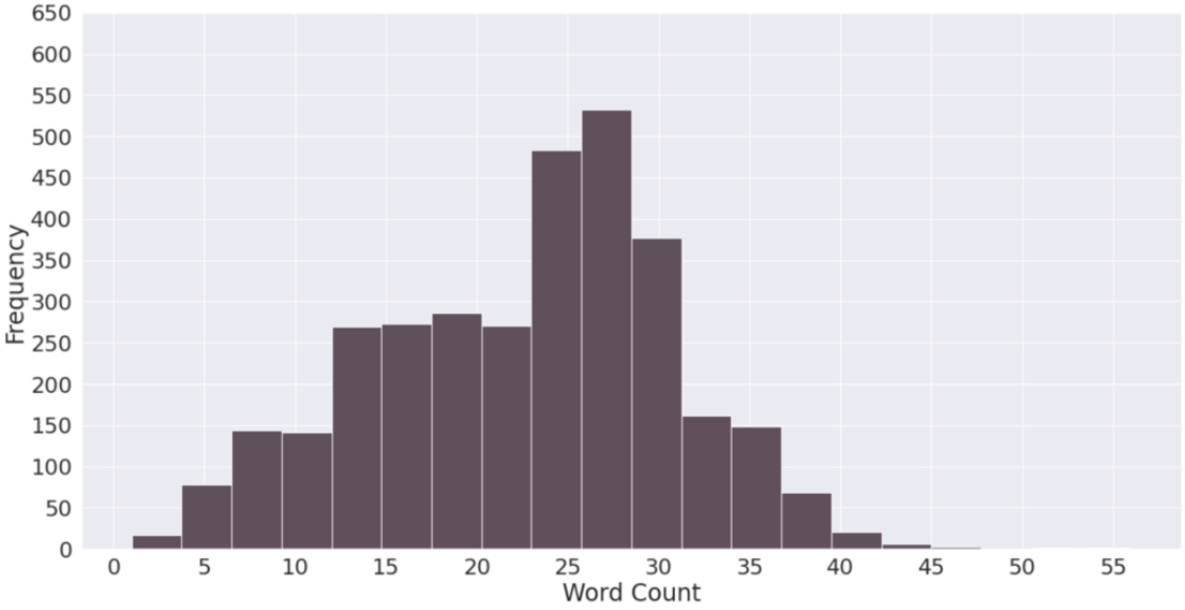


Figure 7. Tweet length distribution.

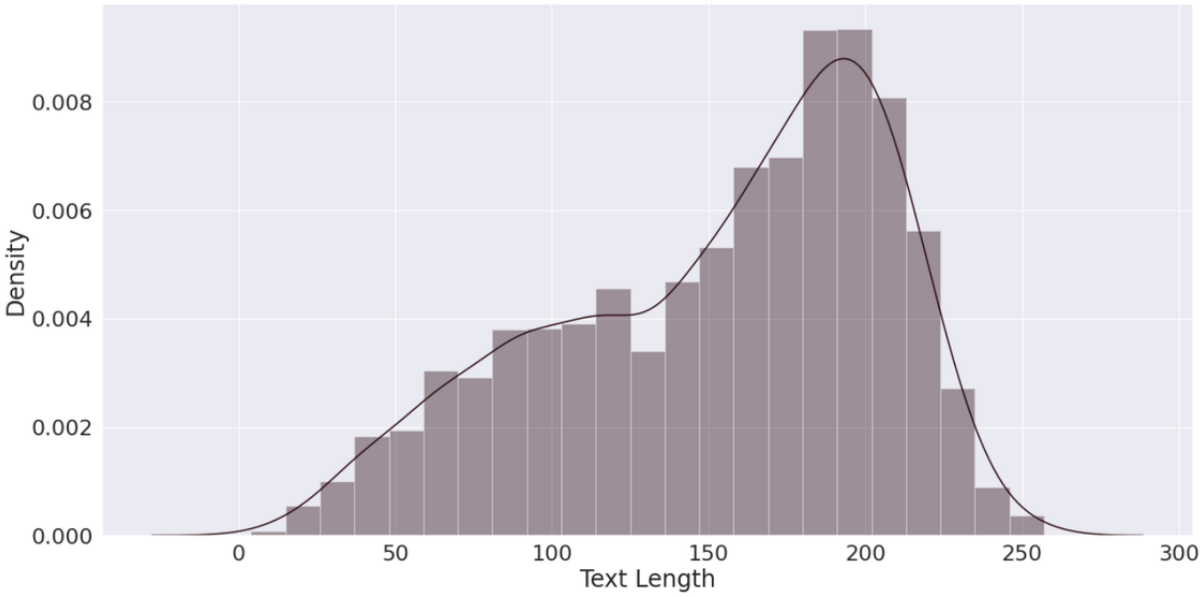
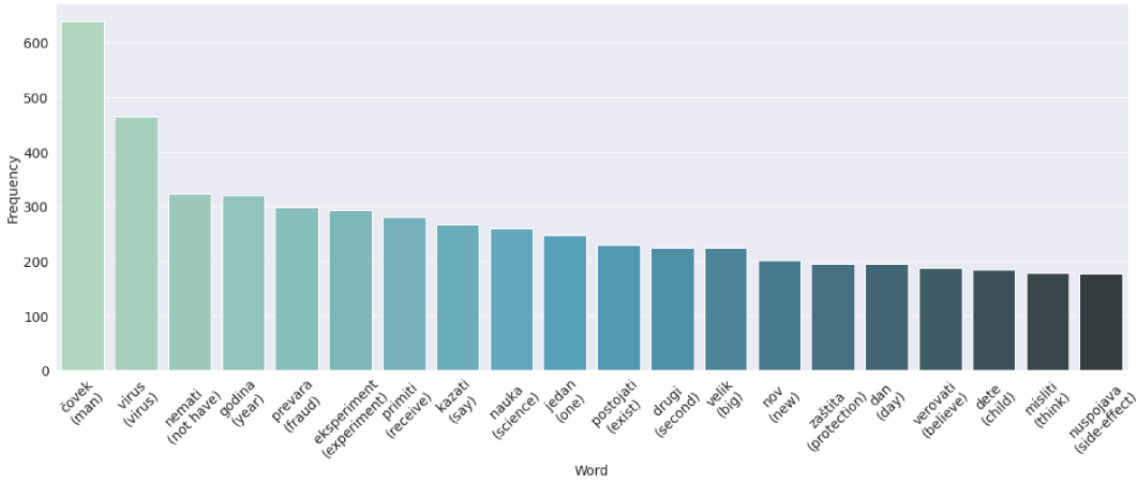


Figure 8. Frequency of top 20 words.



XSL•FO
RenderX

Topic Analysis

The optimal number of topics that we obtained for LDA and NMF was 14 and 13, respectively.

The direct output of both models were the most prominent keywords for each of the topics. We defined the topic names by first looking at the top 20 keywords per topic and then by checking the name against the 30 most prominent tweets assigned to that topic. The defined names and top 10 words per topic for both models can be seen in the table in Multimedia Appendix 1. To display the topics in the LDA method, we used the *Gensim* method "show topics," which returns an arbitrary order of topics. For NMF, like in the case of LDA, there is no natural ordering of topics. The topics are inferred from the highest frequency of words per topic using the topic-word H matrix, which can give us an idea of the content of the topic.

Since we were interested in the topics that are most discussed in our data set, hoping that they would also point to the main reasons for vaccination hesitancy, we ranked the obtained topics by importance by extracting the number of tweets in which each topic was dominant. This topic ranking is presented in Table 4, along with the original topic number. We analyzed each of the topics based on the ordering in this table.

Based on the number of tweets in Table 4, we can see that both methods generate 1 dominant topic. In the case of LDA, 692 (21.1%) tweets belonged to topic 1, and in the case of NMF, 606 (18.4%) tweets belonged to topic 13.

The 2 main ideas that appear in the first few dominant topics can be shortly summarized as concern over vaccine effectiveness and side effects. These ideas are often brought together into consideration. The dominant topic for the LDA method contains these topics mixed. NMF succeeded in extracting a dominant topic based on these 2 ideas, with an emphasis on the concern over the vaccination of children. Even though the word "child" appears among the LDA keywords, there were almost no tweets regarding the vaccination of children in the first 30 most important tweets for that topic. These 2 main ideas were later identified by LDA as several separate topics (topics 3, 4, 10, 11, and 13).

The second-most dominant topic for both LDA (n=420, 12.8%) and NMF (n=279, 8.5%) can be described as doubt about the effectiveness of COVID-19 vaccines. There are several subtopics with regard to vaccine effectiveness. NMF results point out the concern about the effectiveness in the context of new COVID-19 strains that appear rapidly due to the massive scale of the outbreak of the pandemic. In topic 3, LDA struggled with several subtopic mixtures. In the first subtopic, we can see the belief that the vaccine is less effective than natural immunity, while the second subtopic is more about side effects. Once again, we notice the failure of LDA to separate these topics. NMF extracted the topic of natural immunity as a separate topic (topic 8).

The third dominant topic for NMF is the negative sentiment toward government politics related to pandemics. This is not strictly an opinion or attitude toward vaccination, but it often happens to seem so due to the attitudes of Twitter users about government policies in general. As already mentioned in the Relevance Classifier section, it was difficult to draw a strict boundary between political opinions and vaccination itself.

Subtopics in this topic may include frustration over the freedom of movement and choice regarding vaccination, the belief that government institutions are not competent enough in the fight against pandemics, and the belief that their decisions are influenced by various global powers. The third dominant LDA topic pointed out users' frustration with the loss of freedom of movement and freedom of vaccination choice but again got mixed by the skepticism toward medical science, which formed a separate topic found by NMF (topic 5). Therefore, the fourth-most dominant topic found by NMF was skepticism toward vaccine effectiveness connected to the skepticism toward official scientific institutions and experts.

The next, fifth topic given by NMF is skepticism toward the effectiveness of the vaccines in the context of natural immunity. The thesis is that it is better to build immunity naturally than through the vaccination process. This was covered as a subtopic in the second dominant topic given by the LDA method.

The sixth and seventh topics given by the NMF method present a concern that vaccines were fast to appear and therefore could not have been sufficiently tested. This thesis appears in many topics given by the LDA method but was most pronounced in the topic 8.

The eighth dominant topic by NMF presents a pronounced fear of vaccination side effects, including death. Specific side effects are dominant in topic 10 found by LDA. Similarly, the next topic given by the NMF method outlines concerns about so many booster doses, which hints to users either that vaccines are not effective enough or that such a high number of doses may produce heavier side effects, which is the main concern in topic 11 in LDA.

Here, we must outline that the LDA method isolated a topic about the fear regarding messenger RNA (mRNA)–based vaccines (topic 13). The fear is connected with their effectiveness but mainly with the side effects, since in tweets, mRNA vaccines are often connected to genetic treatments. These types of vaccines are often connected with conspiracy theories that some center of power has a genetic mutation agenda for some kind of population control. This was a well-defined topic in both models (LDA topic 12, NMF topic 7).

For the NMF method, 4 last topics exposed fears that the entire pandemic and vaccination process are somehow conspired by various centers of power and for various reasons. The tenth topic postulates that COVID-19 exists only in the media, and topics 11 and 12 postulate that vaccines are a fraud for various different reasons (profit, population control, etc). These concerns appear in topics 6, 9, and 14 in the LDA method.

In the end, NMF extracted a general topic that encompasses frustration with key decision makers in the context of the pandemic. It is a more general version of topics 2, 5, and 6 in LDA.

XSL•FO
RenderX

Table 4. LDA[a] and NMF[b] topics by number of tweets (N=3286).

LDA			NMF		
Topic number	Topic name	Tweets, n (%)	Topic number	Topic name	Tweets, n (%)
1	General concern over vaccine effectiveness and side effects	692 (21.1)	13	Concern over vaccine side effects: negative attitude toward vaccination of children and anxiety about the effects on their health	606 (18.4)
3	Doubt about effectiveness: natural immunity is a better protection, and side effects over-weigh benefits	420 (12.8)	6	Doubt about effectiveness, especially for new strains	279 (8.5)
2	Mistrust of science and concern over violation of freedom of choice and movement	329 (10.0)	12	Linking vaccination with the negative attitude toward the country politics	272 (8.3)
8	Vaccines are an experiment	314 (9.6)	5	Mistrust of science and experts	271 (8.2)
4	Doubt about vaccine effectiveness: vaccines are no protection, especially regarding new strains	264 (8.0)	8	Doubt about vaccine effectiveness: natural immunity is better protection	263 (8.0)
7	Conspiracy theory: COVID-19 is a fraud; vaccines change the DNA	238 (7.2)	4	Vaccine is an experiment and is insufficiently tested	251 (7.6)
6	Vaccines and other measures are means of spreading fear and a money-making scheme	235 (7.2)	9	Anxiety over short vaccine development time and, consequently vaccine side effects	243 (7.4)
12	Conspiracy theory: vaccine as a means of population reduction and control	166 (5.1)	1	Pronounced fear of different vaccine side effects, primarily death	230 (7.0)
5	Mistrust of the government and institutions	146 (4.4)	10	Doubt about vaccine effectiveness and anxiety over side effects due to having to take boosters	218 (6.6)
13	Fear of side effects: vaccines are insufficiently tested, especially the mRNA[c] technology	119 (3.6)	11	Conspiracy theory: COVID-19 does not exist, and consequently, vaccines are a fraud	209 (6.4)
9	Conspiracy theory: Vaccines are a global fraud	100 (3.0)	2	Conspiracy theory: vaccines are a fraud	199 (6.1)
14	Conspiracy theory: linking vaccines with world powers and their agendas	95 (2.9)	7	Conspiracy theory: vaccine as a means of population reduction and control	134 (4.1)
10	Fear of specific side effects	91 (2.8)	3	General frustration over vaccines, institutions, and power players	111 (3.4)
11	Doubt about effectiveness: questioning the need for boosters	77 (2.3)	N/A[d]	N/A	N/A

[a]LDA: latent Dirichlet allocation.

[b]NMF: nonnegative matrix factorization.

[c]mRNA: messenger RNA.

[d]N/A: not applicable.

Discussion

Principal Findings

In this study, we demonstrated the application of several NLP techniques used in combination to find hidden concerns regarding COVID-19 vaccination to a data set of tweets in Serbian. We used BERT-based classifiers to augment the manually annotated data set and obtain the final data set of tweets expressing a negative sentiment toward the COVID-19 vaccination process. We then performed topic modeling on this subset using LDA and NMF and combined the topics obtained by both methods to compile a list of 5 overarching reasons for vaccine hesitancy in Serbia.

Automatic Tweet Classifier

In addition to being able to correctly classify tweets according to their relevance and sentiment, we also wanted to analyze human annotation errors. For both classifiers, we found that there were cases where human annotators made errors, which was to be expected, given the semantic complexity of the tweets. However, the algorithms proved to be resilient to this syndrome and statistically learned well from the majority of correctly labeled examples. To confirm this conclusion, we carefully revised annotations for the test set to the point where we could claim that the test set was almost fully correctly annotated. Nevertheless, we drew conclusions about the confused examples from the original test set.

Upon closer inspection, it was confirmed that this type of annotation task was difficult for people to perform and to decide objectively and with utmost certainty which labels to assign.

XSL•FO
RenderX

As mentioned earlier, the algorithm often outperformed its supervisor by about 12%. This led to the conclusion that annotation was an emotionally and mentally difficult process in which the annotator made typical human mistakes. BERTić, however, learned statistically from the majority of correctly labeled examples. Nevertheless, there was overfitting present in the fine-tuning process, indicated by extremely high training accuracy. This indicates that more data would improve the algorithm. The supervisor outperformed the algorithm in about 8% of the examples. These are the examples that usually contain complex emotional content and figurative language. For many of these examples, broader knowledge is required. Clearly mixed cases accounted for 12%. These examples are mostly long tweets with multiple contradictory statements. Any disagreement is therefore justified. Further inclusion of intermediate values would likely lead to improvement on this basis.

All this suggests that the algorithm would improve if we were to apply some revised annotations through the so-called active learning approach [64]. The already explained overfitting in combination with the annotators' mistakes may lead to a slight bias and degradation of the overall performance of the classifier. However, we expect this to produce a weak effect since most examples are correctly labeled and the algorithm learns robustly and statistically from most correctly labeled examples.

The most similar classifier in the literature for the English language was reported by To et al [9]. Several classifiers were analyzed and compared in this paper. The BERT-based model was reported to have the highest performance. Our metrics values are slightly lower. This is expected because our classifier is more complex as it categorized tweets into several classes according to relevancy and sentiment, whereas classifiers in Ref. [9] are trained in a binary fashion, dividing tweets into negative sentiments and others. Our approach may serve better future work that may encompass the analysis of positive-sentiment tweets.

Topic Modeling

Even though LDA is a generative model, in text mining it introduces a way to attach topical content to text documents. It views each document as a mixture of multiple distinct topics. Our tweets do not fulfill this requirement as they are usually short documents with 1 dominant topic. In addition, LDA suffers from order effects, meaning that different topics can be generated when the order of training data is shuffled. This error can lead to misleading results: the words that define the topic or the order of their importance can be different, which leads to a difference in defining the topic name. As a consequence, there is also a change in the distribution of topics in the documents.

NMF is a linear-algebraic model that factors high-dimensional vectors into a low-dimensional representation. Similar to principal component analysis, NMF takes advantage of the fact that the vectors are nonnegative. It works best with shorter texts, such as tweets or titles, because it does not predefine a document as a mixture of different topics but rather describes it through latent features, which are further clustered.

Having these short descriptions of the used models in mind, along with the analysis of the topics given in the previous section, we can conclude that NMF gave us clearer and more defined topics when looking at the output: keywords and most prominent tweets per topic. However, the LDA-specific results should not be omitted when considering the reasons for vaccination hesitancy, especially since they highlight some aspects that are not immediately seen in NMF topics. Therefore, we compiled the following list of reasons the users of Twitter in Serbia could be hesitant about COVID-19 vaccination by summarizing the topics in both models in the order of their importance:

- Concern over vaccine side effects: (1) general side effects, (2) side effects for children, (3) side effects due to many required doses
- Concern over vaccine effectiveness: (1) natural immunity is better protection, (2) vaccines are not effective against new COVID-19 strains, (3) vaccines are not effective since so many doses are required
- Concern over insufficiently tested vaccines: (1) side effects of such vaccines, (2) effectiveness of such vaccines, (3) violation of freedom by imposing the use of such vaccines
- Mistrust of authorities: (1) medical experts and institutions, (2) government and political decision makers
- Conspiracy theories: (1) vaccines are a money-making scheme; (2) vaccines, especially mRNA vaccines, change DNA; (3) COVID-19 does not exist; thus, vaccines are unnecessary; (4) vaccines are a means of population reduction and control; (5) vaccines are an instrument of world powers and their agendas

Both Table 4 and the table in Multimedia Appendix 1 remain insightful for anyone needing a more detailed overview of people's concerns regarding the vaccination process.

Conclusion

This paper presents a combination of NLP methods aimed at studying the reasons for vaccine hesitancy in Serbia. It focuses on information collected from Twitter and expressed by Twitter users. We first gathered tweets with keywords regarding COVID-19 vaccination. Some of the gathered tweets were used to build a BERT-based classifier for automatic detection of tweets with a relevant and negative opinion about the immunization process. We then used this classifier to automatically classify the second part of the tweets. The technology we used to build this classifier, based on the transformer encoder architecture BERTić, showed prominent and high-quality results. The classifier we built can be used effectively in future studies of public opinion and in particular the immunization process as the world is still unsure about the way pandemics will evolve. Our approach can be relatively easily extended to other world languages.

The second part of the analysis consisted of applying topic modeling methods, LDA and NMF, to negative-sentiment tweets. We considered using the resulting BERTić architecture to perform topic analysis. However, embeddings obtained in such a way did not behave as expected during clustering. In future work, we plan to consider the obtained sentiment classifier for the task of topic modeling. Specifically, our plan

XSL•FO
RenderX

is to use sentence-BERT [65] to obtain tweet embeddings and further cluster them into topics. Given that such resources have not yet been built for South Slavic languages, we opted for using the combination of more traditional techniques for topic analysis.

We isolated and listed the dominant topics in the tweets with a negative sentiment toward vaccination. The main result of this paper is seen in well-researched reasons behind the negative sentiments toward vaccination. Given these reasons, it is now possible to better understand the concerns of people regarding the vaccination process. This will allow the government and medical and pharmaceutical institutions to develop or redefine educational strategies that better address these issues. We hope this can significantly increase the effectiveness of the fight against the COVID-19 pandemic.

Acknowledgments

This work was jointly supported by the Government of Republic of Serbia, the United States Agency for International Development (USAID), and the United Nation's Development Programme (UNDP).

Data Availability

The data set we used to perform topic modeling is available on our GitHub repository [63].

Authors' Contributions

Conceptualization, methodology, and software tasks were performed by AL and NP; validation, formal analysis, investigation, and resources by AL, NP, DM, and BB; data curation and visualization by AL and DM; writing—original draft preparation by AL, NP, and DM; writing—review and editing by BB and JM; supervision and funding acquisition by JM; and project administration by BB. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topics and top 10 keywords detected by latent Dirichlet allocation (LDA) and nonnegative matrix factorization (NMF).
[DOCX File , 28 KB - jmir_v24i11e42261_app1.docx]

References

1. Mallory ML, Lindesmith LC, Baric RS. Vaccination-induced herd immunity: successes and challenges. J Allergy Clin Immunol 2018 Jul;142(1):64-66. [doi: 10.1016/j.jaci.2018.05.007] [Medline: 29803799]
2. Koirala A, Joo YJ, Khatami A, Chiu C, Britton PN. Vaccines for COVID-19: the current state of play. Paediatr Respir Rev 2020 Sep;35:43-49 [FREE Full text] [doi: 10.1016/j.prrv.2020.06.010] [Medline: 32653463]
3. Vergara RJD, Sarmiento PJD, Lagman JDN. Building public trust: a response to COVID-19 vaccine hesitancy predicament. J Public Health (Oxf) 2021 Jun 07;43(2):e291-e292 [FREE Full text] [doi: 10.1093/pubmed/fdaa282] [Medline: 33454769]
4. World Health Organization. Ten Threats to Global Health in 2019. 2019. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2021-01-31]
5. World Health Organization. Vaccination and Trust. 2017. URL: https://www.who.int/publications/i/item/vaccination-and-trust [accessed 2021-01-31]
6. Larson HJ, Cooper LZ, Eskola J, Katz SL, Ratzan S. Addressing the vaccine confidence gap. Lancet 2011 Aug;378(9790):526-535. [doi: 10.1016/s0140-6736(11)60678-8]
7. 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]
8. 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]
9. To QG, To KG, Huynh VN, Nguyen NTQ, Ngo DTN, Alley SJ, et al. Applying machine learning to identify anti-vaccination tweets during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Apr 12;18(8):4069 [FREE Full text] [doi: 10.3390/ijerph18084069] [Medline: 33921539]
10. 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]
11. Kunneman F, Lambooij M, Wong A, Bosch AVD, Mollema L. Monitoring stance towards vaccination in twitter messages. BMC Med Inform Decis Mak 2020 Feb 18;20(1):33 [FREE Full text] [doi: 10.1186/s12911-020-1046-y] [Medline: 32070334]

XSL•FO
RenderX

12. D'Andrea E, Ducange P, Bechini A, Renda A, Marcelloni F. Monitoring the public opinion about the vaccination topic from tweets analysis. Expert Syst Appl 2019 Feb;116:209-226. [doi: 10.1016/j.eswa.2018.09.009]
13. Hayawi K, Shahriar S, Serhani M, Taleb I, Mathew S. ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. Public Health 2022 Feb;203:23-30 [FREE Full text] [doi: 10.1016/j.puhe.2021.11.022] [Medline: 35016072]
14. Villavicencio C, Macrohon JJ, Inbaraj XA, Jeng J, Hsieh J. Twitter sentiment analysis towards COVID-19 vaccines in the Philippines using naïve Bayes. Information 2021 May 11;12(5):204. [doi: 10.3390/info12050204]
15. Wang Y, Chen Y. Characterizing discourses about COVID-19 vaccines on Twitter: a topic modeling and sentiment analysis approach. J Commun Healthc 2022 Mar 24:1-10. [doi: 10.1080/17538068.2022.2054196]
16. Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, et al. Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: spatiotemporal perspective. J Med Internet Res 2021 Sep 10;23(9):e30854 [FREE Full text] [doi: 10.2196/30854] [Medline: 34346888]
17. Cotfas L, Delcea C, Roxin I, Ioanas C, Gherai DS, Tajariol F. The longest month: analyzing COVID-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. IEEE Access 2021;9:33203-33223. [doi: 10.1109/access.2021.3059821]
18. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
19. Liu S, Li J, Liu J. Leveraging transfer learning to analyze opinions, attitudes, and behavioral intentions toward COVID-19 vaccines: social media content and temporal analysis. J Med Internet Res 2021 Aug 10;23(8):e30251 [FREE Full text] [doi: 10.2196/30251] [Medline: 34254942]
20. Huangfu L, Mo Y, Zhang P, Zeng DD, He S. COVID-19 vaccine tweets after vaccine rollout: sentiment–based topic modeling. J Med Internet Res 2022 Feb 8;24(2):e31726 [FREE Full text] [doi: 10.2196/31726] [Medline: 34783665]
21. Ma P, Zeng-Treitler Q, Nelson S. Use of Two Topic Modeling Methods to Investigate COVID Vaccine Hesitancy (presented at 14th International Conference on ICT, Society, and Human Beings, ICT 2021, 18th International Conference on Web Based Communities and Social Media, WBC 2021 and 13th International Conference on e-Health, EH 2021-Held at the 15th Multi-Conference on Computer Science and Information Systems, MCCSIS 2021). 2021. URL: https://www.ict-conf.org/wp-content/uploads/2021/07/04_202106C030_Ma.pdf [accessed 2022-07-11]
22. Monselise M, Chang C, Ferreira G, Yang R, Yang CC. Topics and sentiments of public concerns regarding COVID-19 vaccines: social media trend analysis. J Med Internet Res 2021 Oct 21;23(10):e30765 [FREE Full text] [doi: 10.2196/30765] [Medline: 34581682]
23. Sanandres E, Abello R, Madariaga C. Topic modeling of twitter conversations: the case of the National University of Colombia. In: Iezzi DF, Mayaffre D, Misuraca M, editors. Text Analytics. Cham: Springer; 2020.
24. Himelboim I, McCreery S, Smith M. Birds of a feather tweet together: integrating network and content analyses to examine cross-ideology exposure on Twitter. J Comput-Mediat Comm 2013 Jan 25;18(2):40-60. [doi: 10.1111/jcc4.12001]
25. Rathore AK, Kar AK, Ilavarasan PV. Social media analytics: literature review and directions for future research. Decis Anal 2017 Dec;14(4):229-249. [doi: 10.1287/deca.2017.0355]
26. Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. BMJ Open 2017 Oct 05;7(10):e016869 [FREE Full text] [doi: 10.1136/bmjopen-2017-016869] [Medline: 28982821]
27. Du J, Luo C, Shegog R, Bian J, Cunningham RM, Boom JA, et al. Use of deep learning to analyze social media discussions about the human papillomavirus vaccine. JAMA Netw Open 2020 Nov 02;3(11):e2022025 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22025] [Medline: 33185676]
28. Zhang L, Fan H, Peng C, Rao G, Cong Q. Sentiment analysis methods for HPV vaccines related tweets based on transfer learning. Healthcare (Basel) 2020 Aug 28;8(3):307 [FREE Full text] [doi: 10.3390/healthcare8030307] [Medline: 32872330]
29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A. Attention is all you need. arXiv 2017. [doi: 10.48550/arXiv.1706.03762]
30. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv 2019. [doi: 10.48550/arXiv.1810.04805]
31. Ljubešić N, Lauc D. BERTić: the transformer language model for Bosnian, Croatian, Montenegrin and Serbian. arXiv 2021. [doi: 10.4324/9781003154525]
32. Batanovic V. Semantic similarity and sentiment analysis of short texts in Serbian. 2021 Presented at: 29th Telecommunications Forum (TELFOR); November 23-24, 2021; Belgrade, Serbia. [doi: 10.1109/telfor52709.2021.9653390]
33. Mochtak M, Rupnik P, Ljubešić N. The ParlaSent-BCS dataset of sentiment-annotated parliamentary debates from Bosnia-Herzegovina, Croatia, and Serbia. arXiv 2022. [doi: 10.48550/arXiv.2206.00929]
34. Meaney C, Escobar M, Moineddin R, Stukel TA, Kalia S, Aliarzadeh B, et al. Non-negative matrix factorization temporal topic models and clinical text data identify COVID-19 pandemic effects on primary healthcare and community health in Toronto, Canada. J Biomed Inform 2022 Apr;128:104034 [FREE Full text] [doi: 10.1016/j.jbi.2022.104034] [Medline: 35202844]

XSL•FO
RenderX

35. Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: a call to action for strengthening vaccine confidence. J Infect Public Health 2021 Oct;14(10):1505-1512 [FREE Full text] [doi: 10.1016/j.jiph.2021.08.010] [Medline: 34426095]
36. Yan C, Law M, Nguyen S, Cheung J, Kong J. Comparing public sentiment toward COVID-19 vaccines across Canadian cities: analysis of comments on Reddit. J Med Internet Res 2021 Sep 24;23(9):e32685 [FREE Full text] [doi: 10.2196/32685] [Medline: 34519654]
37. Albalawi R, Yeap TH, Benyoucef M. Using topic modeling methods for short-text data: a comparative analysis. Front Artif Intell 2020;3:42 [FREE Full text] [doi: 10.3389/frai.2020.00042] [Medline: 33733159]
38. Nigam K, McCallum A, Thrun S, Mitchell T. Text classification from labeled and unlabeled documents using EM. Mach Learn 2000;39(2):134. [doi: 10.21236/ada350490]
39. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. 2013 Presented at: WWW '13: Proceedings of the 22nd International Conference on World Wide Web; May 13-17, 2013; Rio de Janeiro, Brazil. [doi: 10.1145/2488388.2488514]
40. Mehrotra R, Sanner S, Buntine W, Xie L. Improving LDA topic models for microblogs via tweet pooling and automatic labeling. 2013 Presented at: SIGIR '13: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 28-August 1, 2013; Dublin, Ireland p. 889-892. [doi: 10.1145/2484028.2484166]
41. Steinskog A, Therkelsen J, Gambäck B. Twitter topic modeling by tweet aggregation. 2017 Presented at: Proceedings of the 21st Nordic Conference on Computational Linguistics; May 2017; Gothenburg, Sweden.
42. Weng J, Lim E, Jiang J, He Q. Twitterrank: finding topic-sensitive influential twitterers. 2010 Presented at: Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM); February 3-6, 2010; New York City p. 261-270. [doi: 10.1145/1718487.1718520]
43. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022. [doi: 10.1162/jmlr.2003.3.4-5.993]
44. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999 Oct 21;401(6755):788-791. [doi: 10.1038/44565] [Medline: 10548103]
45. Arora S, Ge R, Moitra A. Learning topic models: going beyond SVD. arXiv 2012. [doi: 10.1109/focs.2012.49]
46. O'Callaghan D, Greene D, Carthy J, Cunningham P. An analysis of the coherence of descriptors in topic modeling. Expert Syst Appl 2015 Aug;42(13):5645-5657. [doi: 10.1016/j.eswa.2015.02.055]
47. Egger R, Yu J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. Front Sociol 2022 May 6;7:886498 [FREE Full text] [doi: 10.3389/fsoc.2022.886498] [Medline: 35602001]
48. Batanović V, Nikolić B. Sentiment classification of documents in Serbian: the effects of morphological normalization and word embeddings. Telfor 2017;9(2):104-109. [doi: 10.5937/telfor1702104b]
49. Ljajić A, Marovac U, Avdić A. Sentiment analysis of Twitter for the Serbian language. 2017 Presented at: 7th International Conference on Information Society and Technology ICIST 2017; 2017; Kopaonik, Serbia p. 7.
50. Mladenović M, Krstev C, Mitrović J, Stanković R. Using lexical resources for irony and sarcasm classification. 2017 Presented at: BCI '17: Proceedings of the 8th Balkan Conference in Informatics; September 20-23, 2017; Skopje, Macedonia p. 1-8. [doi: 10.1145/3136273.3136298]
51. Stanković R, Košprdić M, Nešic MI, Radović R. Sentiment analysis of sentences from Serbian ELTeC corpus. 2022 Presented at: Proceedings of the 2nd Workshop on Sentiment Analysis and Linguistic Linked Data (SALLD-2) @LREC 2020; June 24, 2022; Marseille, France p. 31-38.
52. Graovac J, Mladenović M, Tanasijević I. NgramSPD: Exploring optimal n-gram model for sentiment polarity detection in different languages. Intell Data Anal 2019 Apr 04;23(2):279-296. [doi: 10.3233/ida-183879]
53. twarc 2.12.0. URL: https://pypi.org/project/twarc/ [accessed 2022-11-07]
54. langid 1.1.6. URL: https://pypi.org/project/langid/ [accessed 2022-11-07]
55. Nguyen D, Vu T, Tuan NA. BERTweet: a pre-trained language model for English Tweets. 2020 Presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; November 8-12, 2020; Online p. 9-14. [doi: 10.18653/v1/2020.emnlp-demos.2]
56. srtools 0.1.13. URL: https://pypi.org/project/srtools/ [accessed 2022-11-07]
57. Ljubešić N, Štefanec V. The CLASSLA-StanfordNLP model for lemmatisation of non-standard Serbian 1.1. Slovenian Language Resource Repository CLARIN.SI. 2020. URL: https://www.clarin.si/repository/xmlui/handle/11356/1351 [accessed 2022-11-07]
58. Hoffman M, Bach F, Blei D. Online learning for latent Dirichlet allocation. In: Lafferty J, Williams C, Shawe-Taylor J, Zemel R, Culotta A, editors. Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates; 2010.
59. Rehurek, Radim, Sojka, Petr. Software framework for topic modelling with large corpora. 2010 Presented at: Proceedings of the LREC Workshop on New Challenges for NLP Frameworks (NLPFrameworks 2010); May 22, 2010; Valletta, Malta p. 45-50. [doi: 10.13140/2.1.2393.1847]
60. Cichocki A, Phan A. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. IEICE Trans Fundamentals 2009;E92-A(3):708-721. [doi: 10.1587/transfun.e92.a.708]

XSL•FO
RenderX

61. Marovac U, Avdić A, Ljajić A. Creating a stop word dictionary in Serbian. Sci Pub Univ Novi Pazar Ser A 2021;13(2):17-25. [doi: 10.5937/spsunp2101017m]
62. Belford M, Mac Namee B, Greene D. Stability of topic modeling via matrix factorization. Expert Systems with Appl 2018 Jan;91:159-169. [doi: 10.1016/j.eswa.2017.08.047]
63. COVID-19-Vaccine-Hesitancy-Tweets. URL: https://github.com/Adelija/COVID-19-vaccine-hesitancy-tweets [accessed 2022-11-07]
64. Ilić V, Tadić J. Active learning using a self-correcting neural network (ALSCN). Appl Intell 2021 Jun 01;52(2):1956-1968. [doi: 10.1007/s10489-021-02515-y]
65. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. arXiv 2019. [doi: 10.18653/v1/d19-1410]

Abbreviations

API: application programming interface
BERT: bidirectional encoder representations from transformers
LDA: latent Dirichlet allocation
mRNA: messenger RNA
NLP: natural language processing
NMF: nonnegative matrix factorization
RNN: recurrent neural network
TF-IDF: term frequency–inverse document frequency
WHO: World Health Organization

Edited by C Basch; submitted 29.08.22; peer-reviewed by G Muric, M Kapsetaki; comments to author 11.09.22; revised version received 29.09.22; accepted 29.09.22; published 17.11.22.

Please cite as:
Ljajić A, Prodanović N, Medvecki D, Bašaragin B, Mitrović J
Uncovering the Reasons Behind COVID-19 Vaccine Hesitancy in Serbia: Sentiment-Based Topic Modeling
J Med Internet Res 2022;24(11):e42261
URL: https://www.jmir.org/2022/11/e42261
doi:10.2196/42261
PMID:36301673

©Adela Ljajić, Nikola Prodanović, Darija Medvecki, Bojana Bašaragin, Jelena Mitrović. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Understanding the Social Mechanism of Cancer Misinformation Spread on YouTube and Lessons Learned: Infodemiological Study

Ho Young Yoon[1*], PhD; Kyung Han You[2*], PhD; Jung Hye Kwon[3,4,5], MD, PhD; Jung Sun Kim[3], MD, MSc; Sun Young Rha[6,7], MD, PhD; Yoon Jung Chang[8], MD, PhD; Sang-Cheol Lee[9], MD, PhD

[1]Division of Communication and Media, Ewha Womans University, Seoul, Republic of Korea

[2]Department of Media and Communication Studies, Jeonbuk National University, Jeonju, Republic of Korea

[3]Division of Hemato-Oncology, Department of Internal Medicine, Chungnam National University Sejong Hospital, Sejong-Si, Republic of Korea

[4]Division of Hematology and Oncology, Department of Internal Medicine, College of Medicine, Chungnam National University, Daejeon, Republic of Korea

[5]Daejeon Regional Cancer Center, Daejeon, Republic of Korea

[6]Division of Medical Oncology, Yonsei University College of Medicine, Seoul, Republic of Korea

[7]Yonsei Cancer Center, Yonsei University Health System, Seoul, Republic of Korea

[8]Division of Cancer Control and Policy, National Cancer Center Korea, Department of Cancer Control and Population Health, National Cancer Center Graduate School of Cancer Science and Policy, Goyang, Republic of Korea

[9]Division of Hematology and Oncology, Department of Internal Medicine, Soonchunhyang University Hospital Cheonan, Cheonan, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jung Hye Kwon, MD, PhD
Division of Hemato-Oncology
Department of Internal Medicine
Chungnam National University Sejong Hospital
20, Bodeum 7-Ro
Sejong-Si, 30099
Republic of Korea
Phone: 82 44 995 4781
Email: kwonjhye.onco@gmail.com

Related Article:

Correction of: https://www.jmir.org/2022/11/e39571

(J Med Internet Res 2022;24(11):e44334) doi:10.2196/44334

In "Understanding the Social Mechanism of Cancer Misinformation Spread on YouTube and Lessons Learned: Infodemiological Study" (J Med Internet Res 2022;24(11):e39571), the authors made the following changes.

1. The *Acknowledgments* section was inadvertently excluded in the originally published article.

In the corrected version, the following statement has been added under the new *Acknowledgments* section.

> *This study was supported by the National R&D Program for Cancer Control through the National Cancer Center (NCC) funded by the Ministry of Health & Welfare, Republic of Korea (HA21C0048).*

2. In the originally published article, Affiliations 7 and 8 were incorrectly presented as two separate affiliations.

In the corrected version, Affiliation 7 is revised as follows to present the original two affiliations:

> *Yonsei Cancer Center, Yonsei University Health System, Seoul, Republic of Korea*

The updated list of affiliations and its attribution to the authors are as follows:

> *Ho Young Yoon[1]; Kyung Han You[2]; Jung Hye Kwon[3,4,5]; Jung Sun Kim[3]; Sun Young Rha[6,7]; Yoon Jung Chang[8]; Sang-Cheol Lee[9]*
>
> *[1]Division of Communication and Media, Ewha Womans University, Seoul, Republic of Korea*
>
> *[2]Department of Media and Communication Studies, Jeonbuk National University, Jeonju, Republic of Korea*

XSL•FO
RenderX

[3]*Division of Hemato-Oncology, Department of Internal Medicine, Chungnam National University Sejong Hospital, Sejong-Si, Republic of Korea*

[4]*Division of Hematology and Oncology, Department of Internal Medicine, College of Medicine, Chungnam National University, Daejeon, Republic of Korea*

[5]*Daejeon Regional Cancer Center, Daejeon, Republic of Korea*

[6]*Division of Medical Oncology, Yonsei University College of Medicine, Seoul, Republic of Korea*

[7]*Yonsei Cancer Center, Yonsei University Health System, Seoul, Republic of Korea*

[8]*Division of Cancer Control and Policy, National Cancer Center Korea, Department of Cancer Control and Population Health, National Cancer Center Graduate School of Cancer Science and Policy, Goyang, Republic of Korea*

[9]*Division of Hematology and Oncology, Department of Internal Medicine, Soonchunhyang University Hospital Cheonan, Cheonan, Republic of Korea*

The correction will appear in the online version of the paper on the JMIR Publications website on November 25, 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 15.11.22; this is a non–peer-reviewed article;accepted 21.11.22; published 25.11.22.

Please cite as:
Yoon HY, You KH, Kwon JH, Kim JS, Rha SY, Chang YJ, Lee SC
Correction: Understanding the Social Mechanism of Cancer Misinformation Spread on YouTube and Lessons Learned: Infodemiological Study
J Med Internet Res 2022;24(11):e44334
URL: https://www.jmir.org/2022/11/e44334
doi:10.2196/44334
PMID:36427350

©Ho Young Yoon, Kyung Han You, Jung Hye Kwon, Jung Sun Kim, Sun Young Rha, Yoon Jung Chang, Sang-Cheol Lee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Digital Health Competencies Among Health Care Professionals: Systematic Review

Jessica Longhini[1], MSc; Giacomo Rossettini[2], DPhil; Alvisa Palese[1], DPhil

[1]Department of Medical Sciences, University of Udine, Udine, Italy
[2]School of Physiotherapy, University of Verona, Verona, Italy

Corresponding Author:
Alvisa Palese, DPhil
Department of Medical Sciences
University of Udine
Viale Ungheria, 20
Udine, 33100
Italy
Phone: 39 0432590926
Email: alvisa.palese@uniud.it

Related Article:

Correction of: https://www.jmir.org/2022/8/e36414/

(J Med Internet Res 2022;24(11):e43721) doi:10.2196/43721

In "Digital Health Competencies Among Health Care Professionals: Systematic Review" (J Med Internet Res 2022;24(8):e36414) the author of a study included in our review, namely "Kocher, A, et al (2021). Patient and healthcare professional eHealth literacy and needs for systemic sclerosis support: a mixed methods study. RMD open, 7(3), e001783" noted two errors:

1. In the "Main Characteristics of Studies Identified" sub-section of "Results", the maximum age of participants of studies included was incorrectly reported:

> *The sample size was variable across the studies, ranging from 36 [30] to 5209 participants [39] with a variable age range from 20 [36] to 68 years [27].*

Thus, the sentence has been updated as follows:

> *The sample size was variable across the studies, ranging from 36 [30] to 5209 participants [39] with a variable age mostly comprised between 30 [46] and 50 years [27].*

2. In Table 1, under the column "Sample and profession; age", referring to the study of Kocher et al, 2021 [27], the age of professionals was incorrectly reported:

> *47 professionals (registered nurses, physiotherapists, rheumatologists, occupational therapists, advanced practice nurses, general practitioners, psychologists, social workers, health policy); median age 60 (IQR 50-68) years*

Thus, the sentence has been updated as follows:

> *47 professionals (registered nurses, physiotherapists, rheumatologists, occupational therapists, advanced practice nurses, general practitioners, psychologists, social workers, health policy); median age 41 (IQR 31-51) years*

The correction will appear in the online version of the paper on the JMIR Publications website on November 29, 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.

XSL•FO
RenderX

Submitted 21.10.22; this is a non–peer-reviewed article;accepted 21.10.22; published 29.11.22.

Please cite as:
Longhini J, Rossettini G, Palese A
Correction: Digital Health Competencies Among Health Care Professionals: Systematic Review
J Med Internet Res 2022;24(11):e43721
URL: https://www.jmir.org/2022/11/e43721
doi:10.2196/43721
PMID:36446087

©Jessica Longhini, Giacomo Rossettini, Alvisa Palese. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predicting Emerging Themes in Rapidly Expanding COVID-19 Literature With Unsupervised Word Embeddings and Machine Learning: Evidence-Based Study

Ridam Pal[1], BTech; Harshita Chopra[2], BTech; Raghav Awasthi[1], BSc, MSc; Harsh Bandhey[1], BTech; Aditya Nagori[1,3], PhD; Tavpritesh Sethi[1], MBBS, PhD

[1]Department of Computational Biology, Indraprastha Institute of Information Technology Delhi, New Delhi, India
[2]Maharaja Surajmal Institute of Technology, Guru Gobind Singh Indraprastha University, New Delhi, India
[3]Council of Scientific & Industrial Research-Institute of Genomics and Integrative Biology, New Delhi, India

Corresponding Author:
Tavpritesh Sethi, MBBS, PhD
Department of Computational Biology
Indraprastha Institute of Information Technology Delhi
Third Floor, New Academic Block
Okhla Industrial Estate, Phase-III
New Delhi, 110020
India
Phone: 91 9779908630
Email: tavpriteshsethi@iiitd.ac.in

Abstract

Background: Evidence from peer-reviewed literature is the cornerstone for designing responses to global threats such as COVID-19. In massive and rapidly growing corpuses, such as COVID-19 publications, assimilating and synthesizing information is challenging. Leveraging a robust computational pipeline that evaluates multiple aspects, such as network topological features, communities, and their temporal trends, can make this process more efficient.

Objective: We aimed to show that new knowledge can be captured and tracked using the temporal change in the underlying unsupervised word embeddings of the literature. Further imminent themes can be predicted using machine learning on the evolving associations between words.

Methods: Frequently occurring medical entities were extracted from the abstracts of more than 150,000 COVID-19 articles published on the World Health Organization database, collected on a monthly interval starting from February 2020. Word embeddings trained on each month's literature were used to construct networks of entities with cosine similarities as edge weights. Topological features of the subsequent month's network were forecasted based on prior patterns, and new links were predicted using supervised machine learning. Community detection and alluvial diagrams were used to track biomedical themes that evolved over the months.

Results: We found that thromboembolic complications were detected as an emerging theme as early as August 2020. A shift toward the symptoms of long COVID complications was observed during March 2021, and neurological complications gained significance in June 2021. A prospective validation of the link prediction models achieved an area under the receiver operating characteristic curve of 0.87. Predictive modeling revealed predisposing conditions, symptoms, cross-infection, and neurological complications as dominant research themes in COVID-19 publications based on the patterns observed in previous months.

Conclusions: Machine learning–based prediction of emerging links can contribute toward steering research by capturing themes represented by groups of medical entities, based on patterns of semantic relationships over time.

(J Med Internet Res 2022;24(11):e34067) doi:10.2196/34067

KEYWORDS

COVID-19; named entity recognition; unsupervised word embeddings; machine learning; natural language preprocessing

XSL•FO
RenderX

Introduction

The COVID-19 pandemic is a global health threat and has proven to be an enigma, with its diverse clinical presentation, controversial evidence for treatment, fast-tracked vaccine development, and unclear systemic implications. Most countries have been affected by COVID-19, with around 187 million confirmed cases over a short span and more than 4 million deaths recorded until July 13, 2021 [1]. The literature around COVID-19 is growing exponentially, with more than 150,000 COVID-19 articles vetted by the World Health Organization (WHO) [2]. Understanding evolving themes in a context, such as COVID-19, is essential as knowledge synthesis from peer-reviewed literature becomes increasingly difficult for researchers, clinicians, and policymakers alike. Methods, such as topic modeling and sentiment analysis, have been previously carried out comparing preprint with peer-reviewed literature only over a short period. Ebadi et al [3] studied the temporal patterns of sentiments and the similarity between publications from different sources over time, using document embeddings. High-level research topics like oncology, personal protective equipment, analytics, rehabilitation panic, high-risk groups, and genomics were uncovered using structural topic modeling. Although such analyses reflect an abstract overview of the broad areas of research, they do not capture the evolving context between distinct domain-specific entities. The objective of our study was to analyze and track word-level semantic similarity among biomedical entities to uncover emerging themes.

Abstracts of articles hold a substantial amount of information in the literature. Named entities within abstracts play a crucial role in deducing valuable information from large amounts of text and influencing literature trends [4]. Models pretrained on biomedical, scientific, and clinical benchmark data sets have been used to extract various clinical entities, such as diseases, symptoms, chemicals, and adverse drug reactions, from continuous text. The relative context of these entities changes over time, leading to a shift in similarity with other words [5]. Unsupervised word embeddings have previously been used to capture complex science concepts using the semantic relationship signified by cosine similarity [6].

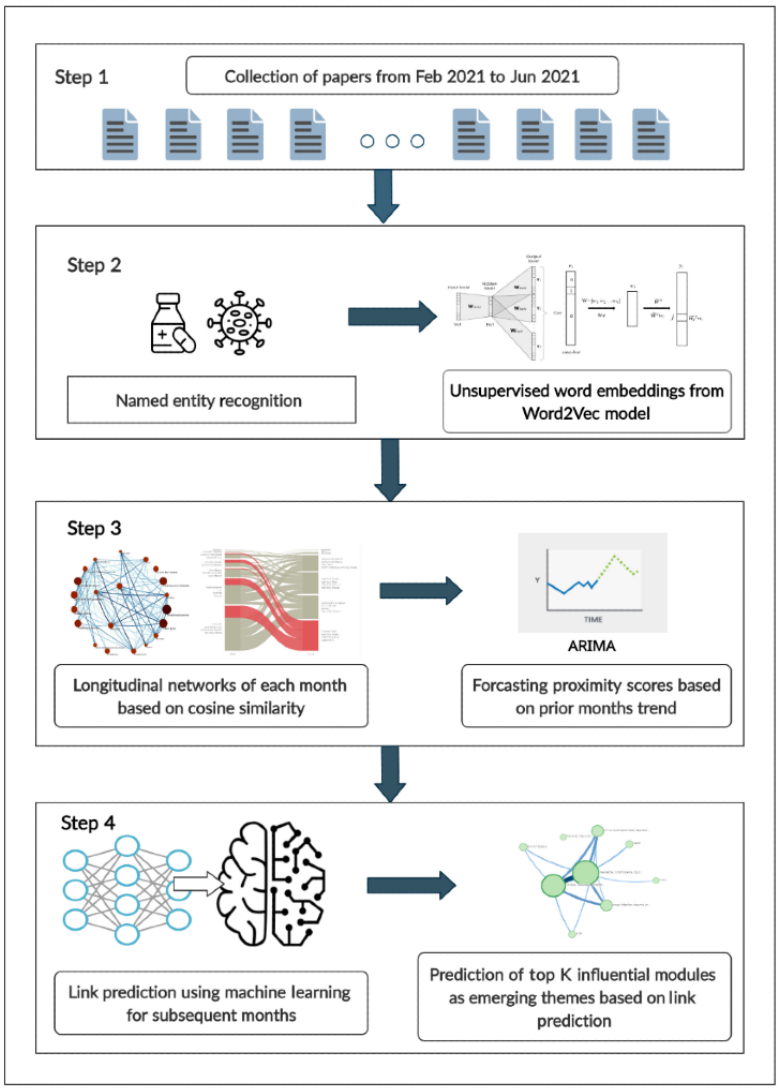
Predicting links between “medical terms” is of high significance to understand the underlying themes within the literature and the phenomenon. Link prediction is the task of predicting the existence of links between 2 nodes in a complex network based on a set of topological features. The problem of link prediction in real-world temporal networks has been explored a lot in recent years [7], primarily in online social media networks where nodes are represented by users and edges are represented by the relationship between them. Supervised learning methods based on topological proximity measures have been vastly used to capture the shifting of links across time within networks [8,9]. Our paper aims to fill these gaps through our proposed framework, EvidenceFlow [10], an interactive web application for tracking literature trends using alluvial diagrams, projection of influential entities, and network analysis across different months. We propose for the first time the use of diachronic word embeddings, link prediction in dynamic networks of entities, and machine learning to predict emerging theme literature and make these publicly available as a web application. This paper also studies the evolution of literature based on changing cosine similarity between extracted entities in weighted temporal networks and predicts future emerging trends using link prediction.

We have primarily focused on the fast emerging COVID-19 literature to train and validate our architecture for this study. We forecasted semantic and topological proximity features of named entity pairs generated from their temporal trends in prior months. Further, we used these forecasted features to predict links between clinical entities extracted from textual data over the forecasted time interval using machine learning algorithms. Furthermore, these links were used to create a network weighted by forecasted cosine similarity for detecting communities of entities that tend to reflect on the themes of the articles published in that month. To assess the efficacy of our predictive modeling, we validated the proximity features of entity pairs forecasted from autoregressive integrated moving average (ARIMA) using mean squared error (MSE). We also evaluated the machine learning algorithm’s performance for predicting the links over a time span of 3 months.

The schematic representation of workflow has been demonstrated (Figure 1). The interactive analysis and results of emerging themes are available publicly on our web application called EvidenceFlow. The details about its working can also be found in Multimedia Appendix 1. This study proposes a framework for capturing and tracking imminent themes formed by medical entities in the temporal space based on networks constructed using word embeddings trained upon the evolving COVID-19 literature.

XSL•FO
RenderX

Figure 1. Graphical representation of the proposed framework explaining the complete workflow. The pipeline takes abstracts as inputs from which entities are extracted using named entity recognition. Embeddings are generated, which are used as features for longitudinal networks. These networks are used for visualizing the trends using alluvial diagrams, link prediction, and predicting top k influential modules for theme prediction. ARIMA: autoregressive integrated moving average.



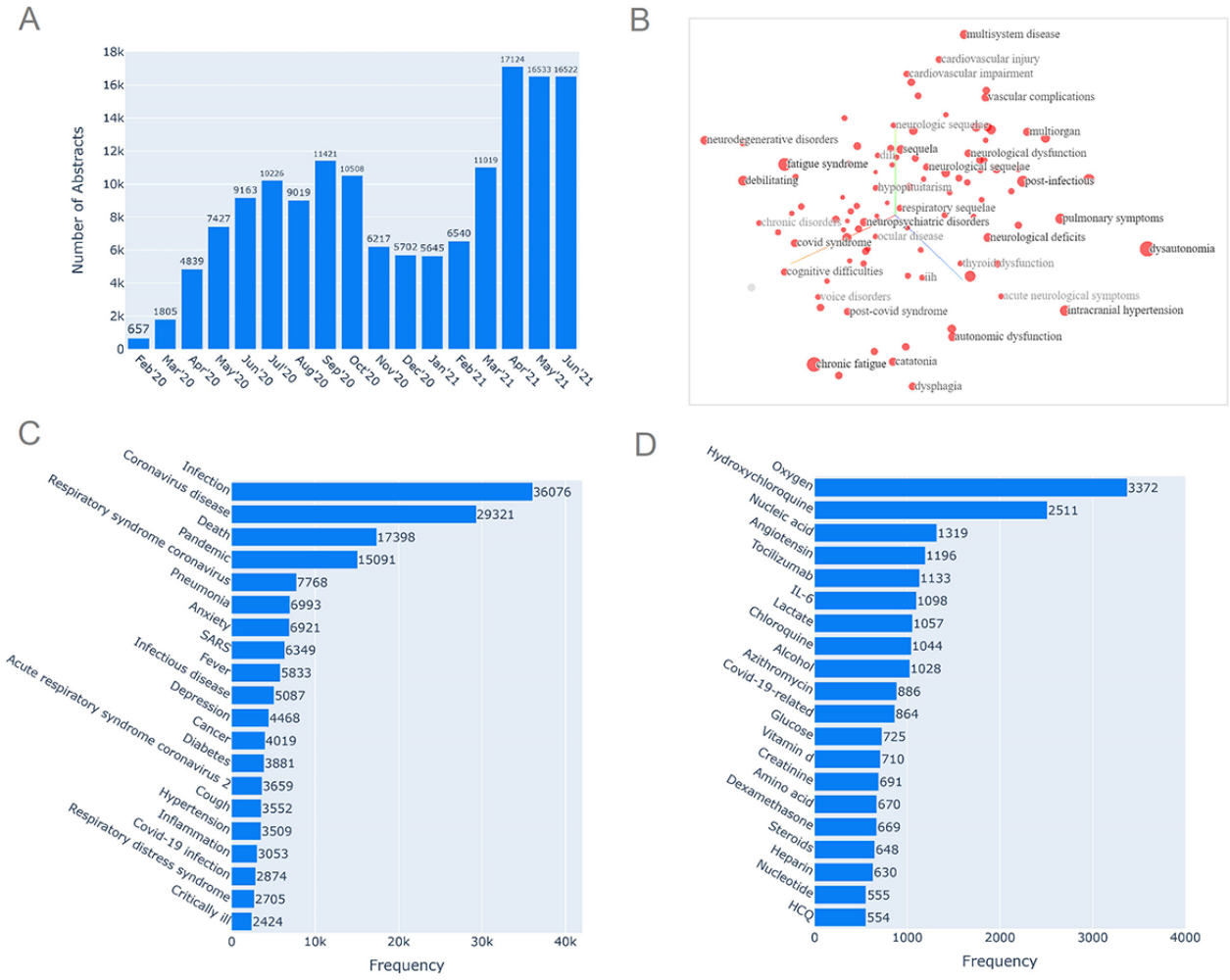
Methods

Data Set and Text Preprocessing

The data set was created from abstracts of approximately 150,000 COVID-19 articles published in the publicly available WHO Database [2] from February 2020 to June 2021 (Figure 2A). For every research article, the database contains the corresponding title, authors, source of publication, journal, database, language, type of publication, entry date, country, and full-text URL. We queried the database on all full-text articles in the English language, keeping the rest of the fields unfiltered. The frequency of articles concerning specific categories and keywords has been depicted in Multimedia Appendix 2. Formatting of text and removal of white spaces, punctuations, digits, and stop words were carried out on lower-case converted text using the Natural Language Toolkit (NLTK) package [11]. We list all the software and packages used in further analysis along with the corresponding versions and sources in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 2. (A) Graph showing the number of articles occurring each month. The curve depicts that there has been a rampant increase in the number of articles across each month since February 2020. (B) Latent space of word embeddings of diseases visualized around the keyword "post-covid syndrome," displaying 100 isolated points nearest to it. (C) Bar plot showing the frequency of the top diseases in the corpus of abstracts extracted using named entity recognition (NER). (D) Bar plot showing the frequency of the top chemicals in the corpus of abstracts extracted using NER. HCQ: hydroxychloroquine; IL: interleukin.



Named Entity Recognition

Named entity recognition (NER) was used to extract 2 types of entities (diseases and chemicals) from the original abstracts of vetted research articles using a model pretrained on the BC5CDR corpus by SciSpacy, an open-source project for biomedical natural language processing [12]. The model identifies entities with an F1 score of 84.49% [13]. The words extracted under the category of diseases also contained symptoms, adverse effects, conditions, disorders, and syndromes. All of these are collectively referred to as diseases in the other sections. Entities were further used to create networks to study the trends through alluvial diagrams and predict links between nodes across past and upcoming months.

Unsupervised Word Embeddings

Word embeddings were trained upon the abstracts obtained from the WHO database updated with new publications and preprints as these become available every month. A low-dimensional representation (d=100) for the words present in the corpus of abstracts was learned using the Word2Vec model with the skip-gram algorithm and a fixed window size of 5, implemented in Gensim [14-16]. Cosine distance between the word vectors of the extracted entities was calculated to analyze the dis(similarity) between entity pairs. Visualization of the word vectors was carried out using TensorFlow Embedding Projector [17] to allow interactive exploration of the relationships between diseases and chemicals. To create each month's network of entities, separate Word2Vec models were trained to capture shifts in word similarities in the literature published over time.

Longitudinal Entity Networks and Communities

High cosine similarity represents strong relationships between words. We used diachronic word embeddings to capture the evolving contextual similarities between various diseases and studied the evolution over time. Weighted networks were constructed using the similarity between word vectors of extracted entities as edge weights. From each month's corpus of abstracts, top N (=100) most frequently occurring diseases were extracted, and pairs having greater than the 90th percentile of cosine similarity based on the corresponding month's word embeddings were used to create a union set of entities across months, preserved as nodes in the temporal networks. Therefore,

XSL•FO
RenderX

every month's network had a fixed set of nodes with varying links, labeled as 0 or 1 based on the threshold of cosine similarity, and varying weights, calculated based on the evolving semantic closeness. The mentioned threshold has been chosen empirically based on experimentation; a high threshold has been selected to depict contextual similarity between 2 words present in the same latent space. For training and evaluation, a fixed set of entity pairs was created from the diseases identified in the abstracts of the papers published from February 2020 to February 2021, using the mentioned procedure. For the subsequent months, the word embedding models were trained on the respective corpora of abstracts, and the links between the fixed set of node pairs were assigned if they appeared in the vocabulary and were weighted by the cosine similarity between their word vectors. Community detection was performed over the monthly networks using the Infomap algorithm [18]. Semantic change in the word embeddings led to the formation of communities, which shifted as emerging themes over months. The importance of each node (entities) was tracked using an alluvial visualization based on PageRank values, which changed across different months [19]. Detailed steps with parameters are available in Multimedia Appendix 1.

Time Series Forecasting of Proximity Scores

In order to predict the existence of links between nodes in the networks of subsequent months, we computed 5 neighborhood proximity scores for the network of each month. Jaccard similarity, common neighbors, preferential attachment [20], and Adamic Adar similarity [21] were used as topology-based features, and cosine similarity between the entities represented by the nodes was used as a semantic feature. These proximity scores based upon network topology were calculated using the NetworkX package [22]. Adamic Adar similarity, common neighbors, and preferential attachment values lie between 0.00 and ∞, while Jaccard similarity and cosine similarity values lie between 0.00 and 1.00. To scale the values, we normalized the former 3 scores in each network to bring them in the range of 0.00 to 1.00.

Every proximity score was modeled as a time series for each node pair, and the value was predicted for the subsequent month using the ARIMA model [23]. Stationarity of the time series was assessed using the augmented Dickey-Fuller test. A first-order autoregressive model (p=1, d=0, q=0) was used for stationary series, and nonstationary time series were passed through the random walk order of the model (p=0, d=1, q=0). For validation, proximity scores for the network at timestamp $\tau+1$ were predicted based on their respective past values in the networks till timestamp τ. The model's performance was assessed by comparing the predictions with the original proximity scores in the $\tau+1$ time using MSE. MSE is one of the robust indicators to measure the closeness of forecast outputs to actual values in the time-series setting. To assess its sensitivity to outliers, we analyzed the distribution of errors (Multimedia Appendix 4). It was seen that the median of errors was close to zero, with minimal influence from outliers. Detailed steps with parameters are available in Multimedia Appendix 1.

Link Prediction Between Entities

The proximity scores predicted using the ARIMA model were further used to identify the occurrence of a link between entities in network G_{+1} based on the proximity scores and links in all previous networks (G_1, G_2, G_3, …, G), using supervised machine learning. We experimented with the proposed link prediction approach using logistic regression [24], random forest [25], support vector machine [26], AdaBoost [27], and XGBoost [28]. For training the models, 4 proximity scores (Jaccard coefficient, preferential attachment, Adamic Adar index, and common neighbors) were used as features of node pairs at each timestamp till . For validation, the forecasted proximity scores of the network at timestamp *+1* were used to predict links between nodes. Due to the high imbalance between the labels, the area under the receiver operating characteristic curve (AUROC) was evaluated to select the optimal threshold for binary classification. While training, validating, and testing the model, we did not use cosine similarity as a feature as it was the identifier variable for the link. Validation of the model was performed on the predicted proximity scores of April 2021 to June 2021. For logistic regression, evaluation of the key assumptions was done using the variance inflation factor for measuring the degree of multicollinearity, the Cook distance for detecting the presence of strongly influential outliers, and the scatter plot of log-odds for checking the linearity of independent variables. These tests were not satisfied for the data of most months; hence, logistic regression was not our preferred model, and we did not consider it further in the results. The Welch *t* test was performed for comparing the performance of the machine learning models, followed by Bonferroni correction [29]. The full details of the algorithm and features are available in Multimedia Appendix 1. We list the parameters set for all the models in Multimedia Appendix 5.

Community Detection on Predicted Networks

The links between node pairs predicted by the best performing model were used to create networks weighted by cosine similarity scores predicted by the ARIMA model. The Infomap algorithm was applied on the predicted and original test network to cluster the nodes into 10 modules. The modules were compared using intersection over union (IOU) with the following formula:

where A represents a set of nodes in the predicted ith module, i {1, 2, …, 10}, and B represents a set of nodes in the original jth module, j {1, 2, …, 10}.

Results

Overall, 46,885 distinct diseases and 53,375 unique chemicals were identified. The top entities are shown in Figure 2C and 2D. Anxiety, depression, and hypertension were found to be present in the top 20 most discussed medical conditions in the research articles. Oxygen and hydroxychloroquine were followed by nucleic acid and angiotensin, a peptide hormone that causes vasoconstriction, among the most discussed chemicals. The latent space of word embeddings around the

XSL•FO
RenderX

keyword "post-covid syndrome" visualized using a t-distributed stochastic neighbor embedding plot (Figure 2B) depicted "chronic fatigue," "debilitating," "neurodegenerative disorders," and "vascular complications" among the closest medical entities in terms of cosine distance. Similar visualization for the term "mental disorders" can be found in Multimedia Appendix 6, and the top 10 most similar entities with the selected keywords "vaccine," "comorbidity," "adverse effects," "social," and "psychological" can be found in Multimedia Appendix 7.

We conducted detailed inference of the alluvial diagram across different months to graphically explore the temporal trends in the literature based on dynamic and homogeneous networks of prevalent medical entities and their associated cosine similarities. Figure 3A represents the flow of themes found in the literature published in 2020. For March 2020, the dominant themes noted were chest pain, acute kidney injury, and lymphocytopenia. While there were lesser traces of "thromboembolic complications" in the literature of early months, it emerged as the most significant theme in August 2020 (Figure 3A). Myocardial injury and cardiovascular diseases surfaced as a crucial cluster of entities in December 2020. Mental health factors, such as depression, loneliness, anxiety, and burnout, gained significance in the literature of the last quarter of 2020. Figure 3B presents the flow of themes found in the literature published in 2021. While thromboembolism, hypoxemia, and myocardial infarction remained major concerns till January 2021, a significant transition toward long COVID symptoms was found as a major theme in March 2021. In June 2021, central modules, including posteffects and neurological complications, stroke, headache, and anosmia, were found to gain importance, along with newer themes around immunocompromised and chronic diseases. Cross-infection–related entities gained focus due to the second wave of COVID-19 cases in multiple countries around the world. The importance of mental health effects transitioned from lesser importance in the first quarter to more emerging and prominent links in the second quarter as highlighted in the alluvial diagram (Figure 3).

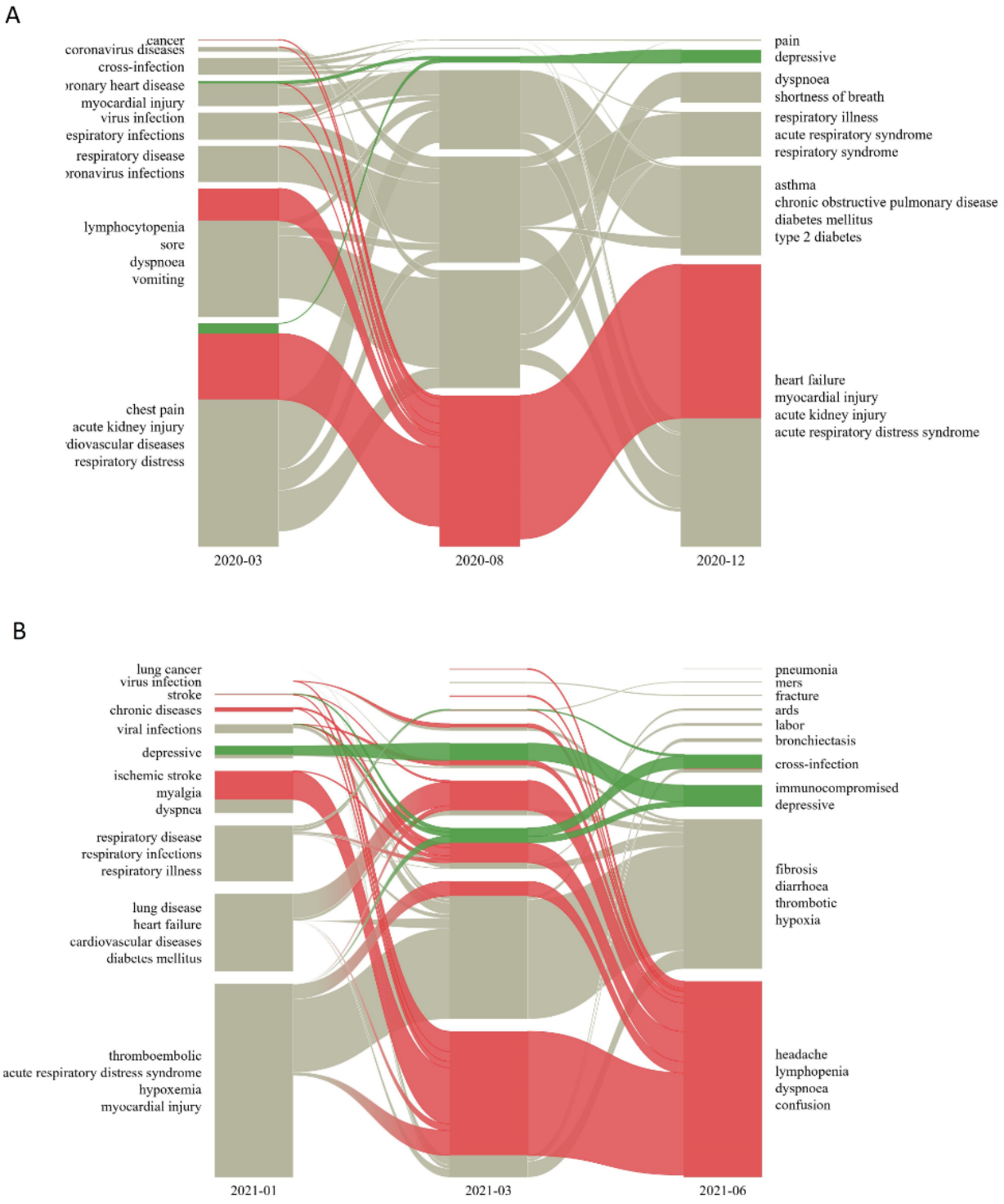
We further advanced the analysis of trends to predicting links between entity pairs for the upcoming months. Our proposed framework for temporal link prediction effectively forecasted 5 proximity scores, including semantic and topological measures, between node pairs by modeling the time series using the ARIMA model. The MSE in the prediction of each proximity score for April 2021, May 2021, and June 2021 is shown in Figure 4A (Multimedia Appendix 8). The associations between diseases for the successive month were predicted as links, using supervised learning based on dynamic networks belonging to the previous months. Our results showed that among the 4 classifiers (Multimedia Appendix 9), the AdaBoost model with 50 estimators and a learning rate of 0.1 classified links with a mean AUROC of 0.871 (all $P<.001$; statistically significant at a Bonferroni-corrected significance level of .02) in the test data of June 2021 (Figure 4B and 4C). Comparisons among other classifiers are shown in Multimedia Appendix 10. The predicted links weighted by forecasted cosine similarity showed a high intersection with the original modules, hence validating the proposed architecture. Multimedia Appendix 11 shows the clusters detected in the original network versus the predicted network. The ARIMA model was used for forecasting proximity scores for subsequent months based on the trends in node pair proximity measures retrieved from the prior months (February 2020 to June 2020). Our findings suggest that the themes of predisposing conditions and risk factors, and studies on cross-infection and neuropsychiatric manifestation will assume a higher centrality in the upcoming quarter of 2021 (Multimedia Appendix 12).

The intersection of nodes between the predicted and original modules was analyzed to prospectively validate the effectiveness of the proposed prediction framework. Table 1 depicts the top nodes in the different modules along with their respective IOU scores for January and June 2021. The collection of intersecting nodes has been interpreted to represent broad themes. Organ damage, like acute kidney injury and pulmonary embolism associated with COVID-19, was the most central theme in the literature from January 2021, followed by cardiovascular diseases, respiratory infections, and psychological effects. Interestingly, major themes in June 2021 shifted toward conditions related to long COVID and neurological symptoms. Headache, encephalitis, and confusion were predicted to be the central nodes, and showed a high IOU score when compared with the original network. The percentages of articles published in June 2021 mentioning entities from each module for the actual and predicted networks are presented in Multimedia Appendix 13. A subset of nodes belonging to different modules from both predicted and true networks has been presented in Multimedia Appendix 11.

Analysis of networks constructed upon chemical entities revealed the evolution of various drugs studied in the COVID-19 literature. During February 2020, the major module contained entities such as paracetamol, tofacitinib, thalidomide, vitamins, zinc, and other linked chemicals. Another relevant module included central entities, such as doxycycline, ruxolitinib, heparin, and ivermectin, which were discussed in the scientific research on the treatment and prevention of COVID-19. In contrast, our recently updated models showed the emergence of evidence for various immunosuppressive drugs, such as tacrolimus, and anti-inflammatory drugs, such as glucocorticoids and colchicine, during November 2021 (Multimedia Appendix 14). These relatively less important entities in earlier months started to become more prominent as the literature expanded. Evidence around "statins" also gained centrality over recent months. Our findings show that the proposed framework captures the dynamic changes in the importance of entities based on their evolving relationship with neighboring entities.

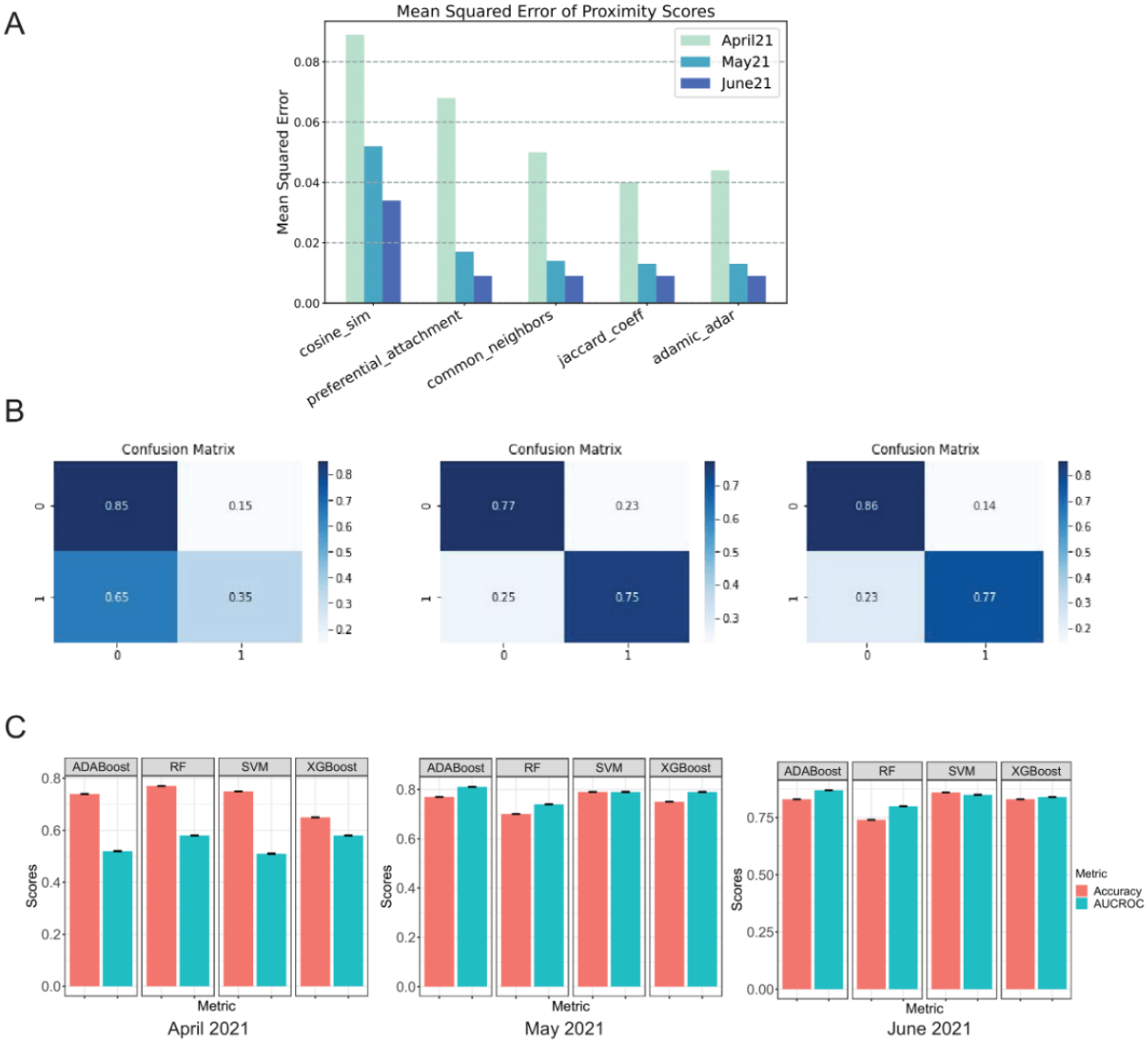
XSL•FO
RenderX

Figure 3. (A) Alluvial diagram for tracking the trends in 2020, from the networks of March, August, and December. (B) Alluvial diagram for monitoring the trends in 2021, from the networks of January, March, and June. The alluvial diagram eases tracing the temporal dynamics of the literature across different time intervals.



XSL•FO
RenderX

Figure 4. (A) Evaluation of the mean squared error (MSE) between the original and predicted proximity scores for the network of April 2021, May 2021, and June 2021. (B) Confusion matrix with normalized values of the results from the AdaBoost classifier across the months of April 2021, May 2021, and June 2021. AdaBoost has been the best performing model across all 3 months. (C) Results of link prediction between disease entities from March 2021 to June 2021, with a margin of error for 95% CIs. The mean value of metrics has been recorded by testing the models on a resampled test set. AUROC: area under the receiver operating characteristic curve; RF: random forest; SVM: support vector machine.



XSL•FO
RenderX

Table 1. Clusters or modules of diseases from the predicted network of January 2021 and June 2021.

Module ID	January 2021		June 2021	
	Top nodes[a]	IOU[b]	Top nodes	IOU
1	Acute kidney injury, ARDS[c], coagulopathy, myocardial injury, pulmonary embolism	0.45	Headache, lymphopenia, dyspnea, confusion, encephalitis, nausea	0.71
2	Cardiovascular disease, diabetes mellitus, COPD[d], hypertension	0.66	Fibrosis, coagulopathy, thrombotic, hypoxia, inflammation, delirium	0.70
3	Respiratory infection, MERS[e], respiratory diseases	0.55	Comorbidity, asthma, COPD, hypertension, dementia, diabetes	0.64
4	Depression, insomnia, anxiety, loneliness	0.71	Traumatic, anxiety, depression, loneliness, burnout, insomnia	0.81
5	Myalgia, lymphopenia, headache, anosmia, dyspnea	0.43	Immunocompromised, chronic diseases like tuberculosis	0.33

[a]A subset of top intersecting nodes in each cluster is mentioned, which collectively signify themes.

[b]The given intersection over union (IOU) was computed between clusters of predicted and original networks of the respective months.

[c]ARDS: acute respiratory distress syndrome.

[d]COPD: chronic obstructive pulmonary disease.

[e]MERS: Middle East respiratory syndrome.

Discussion

Principal Findings

In this paper, we demonstrate a computational approach, EvidenceFlow, in which a user interacts with the rapidly expanding COVID-19 literature to derive and predict emerging themes. The proposed framework tracks patterns of changing semantic and topological proximity between entity pairs across months. Further, it predicts links and network communities that may emerge in future months. Hence, users can follow the papers that contribute to emerging communities of themes, for example, literature around thromboembolic complications captured as early as August 2020 and mental health factors during the end of 2020. Interacting with the clusters on the interactive interface of the EvidenceFlow model revealed that symptoms of long COVID, such as fatigue, headache, myalgia, cough, and anosmia, were forming a central cluster during March 2021. This early signal for accumulating evidence was later validated in large prospective and retrospective cohorts of COVID-19 patients [30-32]. Another way in which users can interact with EvidenceFlow is to gain an understanding of the evolution of themes going beyond current approaches such as topic modeling and sentiment tracking [3]. An example is the early finding of imminent themes around neurological complications, such as confusion, psychiatric illness, and stroke, and mental health factors, such as anxiety, depression, posttraumatic stress disorder, burnout, and insomnia, in June 2021. Our violin plot analysis (Multimedia Appendix 4) showed that despite the mean error being centered on zero, there were some outlier node pairs whose predicted associations deviated from the ground truth. The future scope of this work will involve an analysis of such associations and insights gained by an interactive analysis of such pairs on the EvidenceFlow application.

Prediction of the themes represented by rising centrality of entities can assist in the formation of promising research hypotheses. The dynamics of the literature reveal the emergence of central themes as a combination of pre-existing themes in recent times [6]. For example, the alluvial diagram (Figure 3A) demonstrated how entities from multiple modules in March 2020 merged into a major cluster of thromboembolic complications. Similarly, the flow of importance of psychological disorders over the months indicates their contemporary relevance in the COVID-19 literature and their links with other entities in the cluster. Our framework can potentially help researchers in monitoring existing themes and directing their studies based on trends and predictions.

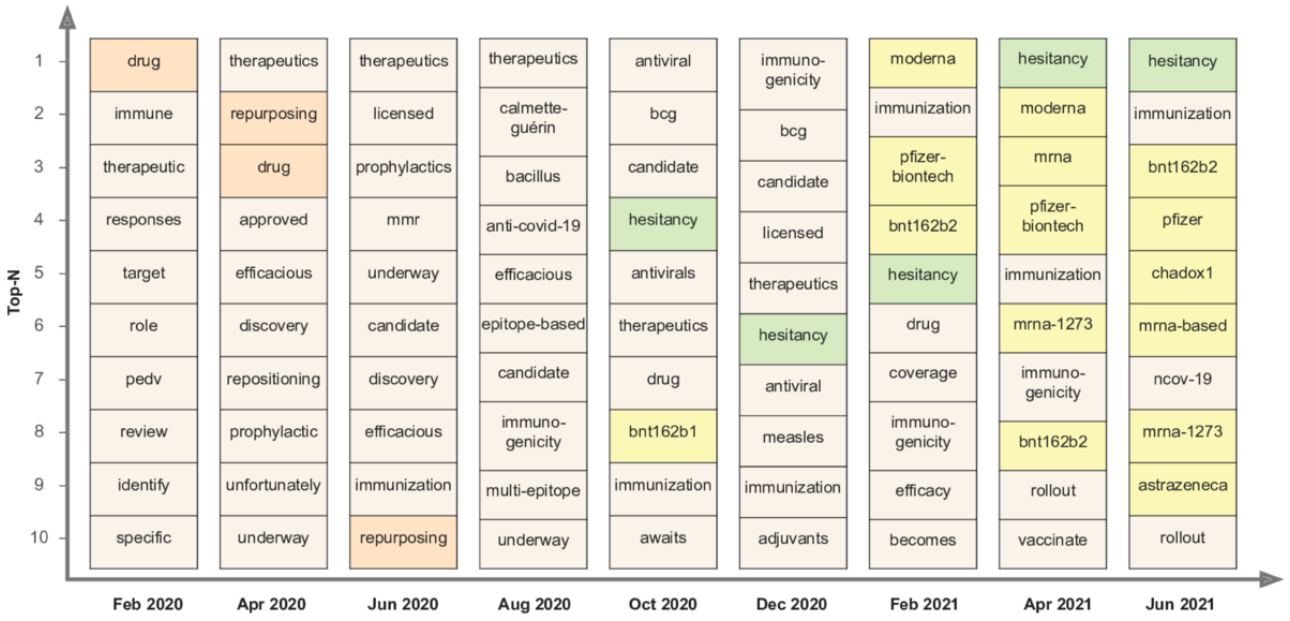
We conducted an analysis on the trends of the PageRank centrality of selected chemical and disease entities. Statins, a class of lipid-lowering medications, were found to be gaining centrality in late 2021 as compared to earlier values (Multimedia Appendix 15). Numerous studies discussed statins for having anti-inflammatory and immunomodulatory effects that may reduce the severity of COVID-19 [33,34]. Glucocorticoids, a class of steroid hormones that reduce inflammation and suppress the immune system, also emerged as a rising entity (Multimedia Appendix 15). Depression and other mental health disorders started becoming a prominent topic of research during the middle of 2020 and gained higher importance in subsequent months (Multimedia Appendix 15). COVID-19 has also been largely discussed in the context of a thromboembolism, and our model captured its emerging evidence as a theme till late 2020. However, the trends showed that its centrality in the literature relatively decreased in 2021 (Multimedia Appendix 15). Discovering such trends from a large corpus is indeed possible using manual curation and analysis by experts. However, our EvidenceFlow pipeline provides an efficient lens to discover, track, and predict emerging trends. This framework will enable faster synthesis of evidence, which then can be validated by experts.

To explore the potential of unsupervised word embeddings and changing cosine similarity among words, we analyzed the trends of terms having maximum similarity with selected keywords. For example, we analyzed the temporal shift in the context of

XSL•FO
RenderX

"vaccine" over the months by finding the top 10 terms most similar to *vaccine* in the latent space of word embeddings trained on the abstracts from each month (Figure 5). From February to August 2020, research on COVID-19 vaccines was underway, and the studies revolved around "therapeutics," "prophylactics," "drug repurposing," and associations with the MMR (measles-mumps-rubella) vaccine and BCG (Bacillus Calmette–Guérin) vaccine. As the clinical trials of certain vaccine candidates became prominent after August 2020, the theme of vaccine *hesitancy* emerged in October 2020 and gained higher similarity in subsequent months. Additionally, as the literature evolved in 2021, a wide range of COVID-19 vaccines, such as BNT162b1, Pfizer-BioNTech, AstraZeneca, ChAdOx1, mRNA-1273, and Moderna, were found to be majorly discussed in the context of research on vaccines. Terms, such as *immunogenicity* and *efficacy*, further suggested high association with vaccine trials and rollouts. Recently updated models showed the emergence of "booster" doses from August 2021 onwards. Such retrospective evaluation of the development of evidence from the literature over time can assist the research community in deriving detailed insights leveraging the applications of word embeddings.

Figure 5. Temporal evolution of the context of the term "vaccine" across alternate months. The top 10 most similar words based on cosine similarity using monthly Word2Vec embeddings are plotted. Origin and evolution of drug repurposing in the early months, hesitancy, and vaccine candidates in the later months are highlighted.



Limitations

Our study has some limitations. First, although the WHO database has been built using a detailed search strategy for COVID-19 literature, it does not explicitly report the exact purpose or accuracy of the search and decision process. The documentation [35] mentions screening done by expert reviewers and an attempt to remove duplicates, but further details are lacking. For example, the process does not clarify if redundancy across various publishers was taken care of. Further, the frequent use of the "OR" combination of keywords may have led to the inclusion of less relevant articles, while other forms of literature, such as patent applications, which can add value to the study, were not included in this database. Nonetheless, we chose the WHO COVID-19 database as it provides a large collection of articles that are updated regularly from searches of multiple bibliographic databases [2]. This, combined with curated expert-referred scientific articles, which would not be readily accessible on a custom search, was useful for building the EvidenceFlow pipeline. Future work with this framework will include potential extension to databases curated through both generic queries and expert vetting, thus facilitating targeted evidence synthesis from a variety of databases.

Further, we are currently using abstracts of research articles to extract named entities and may be missing on the details contained in the full text of the articles while training word embeddings. Therefore, future work may build upon the framework to include the full text of articles wherever available. The NER model used in our study has been reported to have achieved an F1 score of 84.49% on a benchmark data set [13]. Despite the limitations of the F1 score, such as equal weightage given to precision and recall [36,37], F1 remains one of the most widely reported performance indicators. We chose this metric in the absence of other metrics reported for this NER model. For forecasting, we used a relatively basic model (autoregressive approach), as our goal was to capture robust patterns. However, further research is possible for the use of more complex time-series approaches with higher-order difference and lags. Moreover, as the number of timestamps and data points increase, advanced architectures, such as recurrent neural network and long short-term memory [38,39], can be used for handling complex trends in the time series efficiently. Further experiments with larger networks can reveal themes that were not found with the top 100 entities. Importantly, our model is supporting the early detection of emerging trends, but it cannot capture themes on which no evidence has been accumulated.

XSL•FO
RenderX

Conclusion

Consortia across the globe were formed for the advancement of research related to COVID-19. The global attention has led to a widespread increase in the scientific literature to study and prevent the disease from spreading, resulting in an understanding of the disease from multiple perspectives. We introduced a framework built upon COVID-19–specific literature vetted by the WHO and deployed as a dashboard called EvidenceFlow [10]. The dashboard allows the user to unravel the literature with an interactive map of embeddings based on the visualization provided by Tensorboard. It aims to track literature trends using alluvial diagrams, multilevel community detection, and projection of influential entities through network analysis across different months. This study presented how machine learning–based prediction of emerging links can contribute toward analyzing research by capturing themes represented by groups of medical entities, based on patterns of semantic relationships over time.

Acknowledgments

We acknowledge support from the Center of Excellence in Healthcare and the Center of Excellence in Artificial Intelligence at Indraprastha Institute of Information Technology-Delhi.

Authors' Contributions

RP and HC designed and implemented the computational framework, interpreted the results, and wrote the paper. HB contributed to writing and created the associated dashboard. RA and AN interpreted the results and provided feedback on statistical methods. TS designed the study, analyzed the results, and contributed to writing. All authors read and approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary text.
[DOCX File , 17 KB - jmir_v24i11e34067_app1.docx]

Multimedia Appendix 2
Frequency of articles belonging to specific categories in the COVID-19 literature.
[DOCX File , 13 KB - jmir_v24i11e34067_app2.docx]

Multimedia Appendix 3
List of software and packages used for our study with their sources and identifiers for the reproducibility of this study.
[DOCX File , 14 KB - jmir_v24i11e34067_app3.docx]

Multimedia Appendix 4
Distribution of errors in the prediction of proximity scores between node pairs (used as features in model training) for the month of June 2021.
[DOCX File , 54 KB - jmir_v24i11e34067_app4.docx]

Multimedia Appendix 5
Models and respective parameters used for training.
[DOCX File , 13 KB - jmir_v24i11e34067_app5.docx]

Multimedia Appendix 6
Latent space of word embeddings of diseases and chemicals visualized around the keyword “mental disorders,” displaying 100 isolated points nearest to it.
[DOCX File , 151 KB - jmir_v24i11e34067_app6.docx]

Multimedia Appendix 7
The top 10 similar entities (diseases, conditions, or chemicals) with selected keywords (“vaccine,” “comorbidity,” “adverse effects,” “social,” and “psychological”) in descending order of cosine similarity calculated using the word embeddings generated from the Word2Vec model trained on the entire corpus.
[DOCX File , 15 KB - jmir_v24i11e34067_app7.docx]

XSL•FO
RenderX

Multimedia Appendix 8
Evaluation of the mean squared error between original and predicted proximity scores for the network of April 2021, May 2021, and June 2021.
[DOCX File , 13 KB - jmir_v24i11e34067_app8.docx]

Multimedia Appendix 9
Results of temporal link prediction between entities for the months of April 2021, May 2021, and June 2021, with a margin of error for 95% confidence intervals.
[DOCX File , 14 KB - jmir_v24i11e34067_app9.docx]

Multimedia Appendix 10
Welch t test results of the performance of algorithms for the test set of June 2021.
[DOCX File , 14 KB - jmir_v24i11e34067_app10.docx]

Multimedia Appendix 11
Community detection results from the predicted and actual networks for June 2021.
[DOCX File , 14 KB - jmir_v24i11e34067_app11.docx]

Multimedia Appendix 12
Results of community detection from the predicted subsequent network based on training data till June 2021.
[DOCX File , 14 KB - jmir_v24i11e34067_app12.docx]

Multimedia Appendix 13
Percentage of abstracts of articles published in June 2021 mentioning diseases belonging to each module in the actual (A) and predicted (B) networks.
[DOCX File , 59 KB - jmir_v24i11e34067_app13.docx]

Multimedia Appendix 14
Alluvial diagram for tracking the trends of chemical entities from the networks of February 2020 to November 2021.
[DOCX File , 148 KB - jmir_v24i11e34067_app14.docx]

Multimedia Appendix 15
Temporal trends of the PageRank centrality of (A) “statins,” (B) “glucocorticoids,” (C) “depressive,” and (D) “thromboembolic”.
[DOCX File , 167 KB - jmir_v24i11e34067_app15.docx]

References

1. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2022-01-19]
2. Global research on coronavirus disease (COVID-19). World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov [accessed 2022-01-19]
3. Ebadi A, Xi P, Tremblay S, Spencer B, Pall R, Wong A. Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. Scientometrics 2021 Nov 19;126(1):725-739 [FREE Full text] [doi: 10.1007/s11192-020-03744-7] [Medline: 33230352]
4. Cho H, Lee H. Biomedical named entity recognition using deep neural networks with contextual information. BMC Bioinformatics 2019 Dec 27;20(1):735 [FREE Full text] [doi: 10.1186/s12859-019-3321-4] [Medline: 31881938]
5. Kutuzov A, Øvrelid L, Szymanski T, Velldal E. Diachronic word embeddings and semantic shifts: a survey. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018 Presented at: 27th International Conference on Computational Linguistics; August 2018; Santa Fe, New Mexico, USA p. 1384-1397 URL: https://aclanthology.org/C18-1117//>
6. Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature 2019 Jul 3;571(7763):95-98. [doi: 10.1038/s41586-019-1335-8] [Medline: 31270483]
7. Bu Z, Wang Y, Li H, Jiang J, Wu Z, Cao J. Link prediction in temporal networks: Integrating survival analysis and game theory. Information Sciences 2019 Sep;498:41-61. [doi: 10.1016/j.ins.2019.05.050]

XSL•FO
RenderX

8. Özcan A, Öğüdücü Ş. Supervised temporal link prediction using time series of similarity measures. 2017 Presented at: Ninth International Conference on Ubiquitous and Future Networks (ICUFN); July 04-07, 2017; Milan, Italy p. 519-521. [doi: 10.1109/ICUFN.2017.7993838]
9. Güneş İ, Gündüz-Öğüdücü Ş, Çataltepe Z. Link prediction using time series of neighborhood-based node similarity scores. Data Min Knowl Disc 2015 Feb 27;30(1):147-180. [doi: 10.1007/s10618-015-0407-0]
10. What is EvidenceFlow? EvidenceFlow. URL: https://evidenceflow.tavlab.iiitd.edu.in/index [accessed 2022-01-19]
11. Bird S, Loper E. NLTK: The Natural Language Toolkit. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions. 2004 Presented at: ACL Interactive Poster and Demonstration Sessions; July 2004; Barcelona, Spain p. 214-217 URL: https://aclanthology.org/P04-3031/
12. Neumann M, King D, Beltagy I, Ammar W. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. In: Proceedings of the 18th BioNLP Workshop and Shared Task. 2019 Presented at: 18th BioNLP Workshop and Shared Task; August 2019; Florence, Italy p. 319-327 URL: https://aclanthology.org/W19-5034/
13. scispacy. GitHub. URL: https://allenai.github.io/scispacy/ [accessed 2022-01-19]
14. Ma L, Zhang Y. Using Word2Vec to process big text data. 2015 Presented at: 2015 IEEE International Conference on Big Data (Big Data); October 29, 2015-November 01, 2015; Santa Clara, CA, USA p. 2895-2897. [doi: 10.1109/bigdata.2015.7364114]
15. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality. Neural Information Processing Systems. URL: https://papers.nips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html [accessed 2022-01-19]
16. Rehurek R, Sojka P. Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University. URL: https://radimrehurek.com/gensim/ [accessed 2022-09-05]
17. Smilkov D, Thorat N, Nicholson C, Reif E, Viégas F, Wattenberg M. Embedding projector: Interactive visualization and interpretation of embeddings. arXiv. 2016. URL: https://arxiv.org/pdf/1611.05469.pdf [accessed 2022-09-05]
18. Bohlin L, Edler D, Lancichinetti A, Rosvall M. Community Detection and Visualization of Networks with the Map Equation Framework. In: Ding Y, Rousseau R, Wolfram D, editors. Measuring Scholarly Impact. Cham: Springer; 2014:3-34.
19. Rosvall M, Bergstrom CT. Mapping change in large networks. PLoS One 2010 Jan 27;5(1):e8694 [FREE Full text] [doi: 10.1371/journal.pone.0008694] [Medline: 20111700]
20. Barabasi AL, Albert R. Emergence of scaling in random networks. Science 1999 Oct 15;286(5439):509-512. [doi: 10.1126/science.286.5439.509] [Medline: 10521342]
21. Adamic LA, Adar E. Friends and neighbors on the Web. Social Networks 2003 Jul;25(3):211-230. [doi: 10.1016/s0378-8733(03)00009-1]
22. Hagberg A, Swart P, S Chult D. Exploring network structure, dynamics, and function using networkx. Office of Scientific and Technical Information. 2008. URL: https://www.osti.gov/biblio/960616 [accessed 2022-09-05]
23. Zhang G. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003 Jan;50:159-175. [doi: 10.1016/s0925-2312(01)00702-0]
24. Wright RE. Logistic regression. In: Grimm LG, Yarnold PR, editors. Reading and understanding multivariate statistics. Washington, DC: American Psychological Association; 1995:217-244.
25. Breiman L. Random forests. Machine Learning 2001;45:5-32. [doi: 10.1023/A:1010933404324]
26. Hearst M, Dumais S, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell. Syst. Their Appl 1998 Jul 10;13(4):18-28. [doi: 10.1109/5254.708428]
27. Freund Y, Schapire RE. A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence 1999;14(5):771-780 [FREE Full text]
28. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H. Xgboost: extreme gradient boosting. R package version 4–2. R Project. URL: https://cran.r-project.org/web/packages/xgboost/index.html [accessed 2022-09-05]
29. No authors. Etymologia: Bonferroni correction. Emerg Infect Dis 2015 Feb;21(2):289 [FREE Full text] [doi: 10.3201/eid2102.et2102] [Medline: 25786274]
30. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. PLoS Med 2021 Sep;18(9):e1003773 [FREE Full text] [doi: 10.1371/journal.pmed.1003773] [Medline: 34582441]
31. López-León S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo P, Cuapio A, et al. More than 50 Long-Term Effects of COVID-19: A Systematic Review and Meta-Analysis. SSRN. URL: https://ssrn.com/abstract=3769978 [accessed 2022-09-05]
32. Blomberg B, Mohn KG, Brokstad KA, Zhou F, Linchausen DW, Hansen B, Bergen COVID-19 Research Group, et al. Long COVID in a prospective cohort of home-isolated patients. Nat Med 2021 Sep;27(9):1607-1613 [FREE Full text] [doi: 10.1038/s41591-021-01433-3] [Medline: 34163090]
33. Daniels L, Ren J, Kumar K, Bui Q, Zhang J, Zhang X, et al. Relation of prior statin and anti-hypertensive use to severity of disease among patients hospitalized with COVID-19: Findings from the American Heart Association's COVID-19 Cardiovascular Disease Registry. PLoS One 2021;16(7):e0254635 [FREE Full text] [doi: 10.1371/journal.pone.0254635] [Medline: 34264974]

XSL•FO
RenderX

34. Peymani P, Dehesh T, Aligolighasemabadi F, Sadeghdoust M, Kotfis K, Ahmadi M, et al. Statins in patients with COVID-19: a retrospective cohort study in Iranian COVID-19 patients. Transl Med Commun 2021 Jan 25;6(1):3 [FREE Full text] [doi: 10.1186/s41231-021-00082-5] [Medline: 33521322]
35. WHO COVID-19 Sources Search Strategy. World Health Organization. URL: https://www.who.int/docs/default-source/coronaviruse/who-covid-19-database/who-covid-19_sources_searchstrategy_20211012.pdf [accessed 2022-09-05]
36. Hand D, Christen P. A note on using the F-measure for evaluating record linkage algorithms. Stat Comput 2017 Apr 19;28(3):539-547. [doi: 10.1007/s11222-017-9746-6]
37. Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv. URL: https://arxiv.org/ftp/arxiv/papers/2010/2010.16061.pdf [accessed 2022-09-05]
38. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena 2020 Mar;404:132306. [doi: 10.1016/j.physd.2019.132306]
39. Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. IEEE Trans. Neural Netw. Learning Syst 2017 Oct;28(10):2222-2232. [doi: 10.1109/tnnls.2016.2582924]

Abbreviations

ARIMA: autoregressive integrated moving average
AUROC: area under the receiver operating characteristic curve
IOU: intersection over union
MSE: mean squared error
NER: named entity recognition
WHO: World Health Organization

Edited by C Basch; submitted 06.10.21; peer-reviewed by A Rovetta, WD Dotson; comments to author 31.10.21; revised version received 17.12.21; accepted 11.02.22; published 02.11.22.

Please cite as:
Pal R, Chopra H, Awasthi R, Bandhey H, Nagori A, Sethi T
Predicting Emerging Themes in Rapidly Expanding COVID-19 Literature With Unsupervised Word Embeddings and Machine Learning: Evidence-Based Study
J Med Internet Res 2022;24(11):e34067
URL: https://www.jmir.org/2022/11/e34067
doi:10.2196/34067
PMID:36040993

©Ridam Pal, Harshita Chopra, Raghav Awasthi, Harsh Bandhey, Aditya Nagori, Tavpritesh Sethi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effect of the First UK COVID-19 Lockdown on Users of the Drink Less App: Interrupted Time Series Analysis of Sociodemographic Characteristics, Engagement, and Alcohol Reduction

Melissa Oldham[1], PhD; Olga Perski[1], PhD; Gemma Loebenberg[1], MSc; Jamie Brown[1], PhD; Claire Garnett[1], PhD

Department of Behavioural Science and Health, University College London, London, United Kingdom

Corresponding Author:
Melissa Oldham, PhD
Department of Behavioural Science and Health
University College London
1-19 Torrington Pl
London, WC1E 7HB
United Kingdom
Phone: 44 7507712530
Email: m.oldham@ucl.ac.uk

Abstract

Background: The first UK COVID-19 lockdown had a polarizing impact on drinking behavior and may have impacted engagement with digital interventions to reduce alcohol consumption.

Objective: We examined the effect of lockdown on engagement, alcohol reduction, and the sociodemographic characteristics of users of the popular and widely available alcohol reduction app Drink Less.

Methods: This was a natural experiment. The study period spanned 468 days between March 24, 2019, and July 3, 2020, with the introduction of UK lockdown measures beginning on March 24, 2020. Users were 18 years or older, based in the United Kingdom, and interested in drinking less. Interrupted time series analyses using generalized additive mixed models (GAMMs) were conducted for each outcome variable (ie, sociodemographic characteristics, app downloads and engagement levels, alcohol consumption, and extent of alcohol reduction) for existing (downloaded the app prelockdown) and new (downloaded the app during the lockdown) users of the app.

Results: Among existing users of the Drink Less app, there were increases in the time spent on the app per day (B=0.01, *P*=.01), mean units of alcohol recorded per day (B>0.00 *P*=.02), and mean heavy drinking (>6 units) days (B>0.00, *P*=.02) during the lockdown. Previous declines in new app downloads plateaued during the lockdown (incidence rate ratio [IRR]=1.00, *P*=.18). Among new app users, there was an increase in the proportion of female users (B>0.00, *P*=.04) and those at risk of alcohol dependence (B>0.00, *P*=.01) and a decrease in the proportion of nonmanual workers (B>–0.00, *P*=.04). Among new app users, there were step increases in the mean number of alcohol units per day (B=20.12, *P*=.03), heavy-drinking days (B=1.38, *P*=.01), and the number of days the app was used (B=2.05, *P*=.02), alongside a step decrease in the percentage of available screens viewed (B=–0.03, *P*=.04), indicating users were using less of the intervention components within the app.

Conclusions: Following the first UK lockdown, there was evidence of increases in engagement and alcohol consumption among new and existing users of the Drink Less app.

(J Med Internet Res 2022;24(11):e42320) doi:10.2196/42320

KEYWORDS

alcohol reduction; COVID-19; digital intervention; smartphone app; United Kingdom; alcohol; app; Drink Less; engagement; users; lockdown; female

XSL•FO
RenderX

Introduction

Alcohol consumption is a dose-dependent [1], leading risk factor for preventable cases of cancer [2] and is linked with many other chronic and acute conditions [3]. Restrictions introduced as a result of the COVID-19 pandemic impacted drinking behavior, with rises in increasing and higher-risk drinking (defined by standard cut-offs of ≥8 on the full Alcohol Use Disorders Identification Test [AUDIT]) [4-6] and heavy episodic drinking [6] in the United Kingdom. Reducing alcohol-related harms is a public health priority [7]. This study reports the possible impact of COVID-19 restrictions on app engagement, alcohol reduction, and the sociodemographic characteristics of existing and new users of the moderately popular alcohol reduction app *Drink Less* [8].

The first lockdown was introduced in the United Kingdom in response to the COVID-19 pandemic, in March 2020. Following initial government advice from March 16, 2020, to avoid group gatherings and to work from home, where possible, the first national lockdown was announced with behavioral restriction measures coming into force in the United Kingdom from the March 24, 2020, remaining in place until July 4, 2020 [9]. Two subsequent national lockdowns and other social distancing measures followed. During the lockdown, all nonessential stores and licensed premises were closed, social gathering was prohibited, and the opportunity to drink alcohol outside the home was limited. Although pubs, clubs, and bars were closed, people could still purchase alcohol for consumption at home and domestic alcohol expenditure increased in the first lockdown [10,11]. Evidence suggests that the first lockdown had a polarizing impact on drinking patterns [6,12], with 26% of drinkers drinking less and 26% drinking more [12] and people replacing on-trade consumption with increases in own-home drinking [13]. The prevalence of increasing and higher-risk drinking increased significantly, with 1.8 times greater odds during the lockdown relative to the prelockdown period [4].

In addition to leading to increased alcohol consumption among some groups, the first lockdown also led to an increase in self-reported alcohol reduction attempts by increasing and higher-risk drinkers (28.5% during the lockdown vs 15.3% prelockdown) [4]. Many nonemergency, in-person National Health Service (NHS) facilities were temporarily closed during the lockdown, which could have made it more difficult for those motivated to cut down to access advice from health care professionals or specialist support services. As the opportunity to engage with physical health services was limited during the lockdown, we might expect to see an increase in the use of digital interventions during this period.

Digital interventions, such as websites and apps, are convenient and low cost [14] and accessible to most people in the United Kingdom [15]. Despite this, alcohol reduction apps are less frequently used than more traditional support services, with only around 4% of drinkers using one when attempting to cut down between 2015 and 2018 [16]. The context in which a digital intervention is used affects engagement [17], and it is unclear how the lockdown affected engagement with alcohol reduction apps. Being furloughed from work or having limited socialization opportunities could have resulted in more new users downloading apps or existing app users having a greater opportunity to engage with them. This could be reflected in the frequency, amount, and depth of use among existing users. Alternatively, competing priorities, such as adjusting to working from home, childcare, or other caring responsibilities, may have negatively impacted the time spent using alcohol reduction apps. Although many apps are available, few are informed by evidence and theory [18,19]. Drink Less is a free app containing multiple intervention components informed by behavior change theory and evidence, including a drinking diary and goal setting, that aims to support users in reducing their alcohol consumption [20,21] and has over 70,000 unique users, with a 4.5 star rating in the App Store [8]. The full process of developing and refining Drink Less, along with the intervention content, is reported elsewhere [20,22].

Understanding how the initial lockdown affected engagement with digital alcohol interventions and subsequent alcohol reduction could inform the targeting of public health messaging and the provision of alcohol support in the future. It is also important to consider the consistency of these effects across the population. Drinkers who drink more heavily [23], are more deprived [24,25], and are younger [26,27] are more likely to experience alcohol-related harms and may also have been differentially affected by the lockdown [28-30]. For example, more deprived groups are more likely to have unstable incomes and greater financial concerns and may have been isolated in less comfortable, more crowded accommodation [31]. This likely impacted the motivation, opportunity, and capability to engage with digital interventions and reduce drinking. Before the lockdown, users of Drink Less tended to be of higher socioeconomic status (SES) and heavier drinkers [32]. As such, it is important to consider the characteristics of those engaging with digital interventions designed to support alcohol reduction and whether this changed in the lockdown in order to monitor the possibility of emerging or worsened inequalities.

The aims of this paper are threefold. First, to understand how the use of Drink Less and drinking behavior may have changed during the lockdown, we examined whether the lockdown affected engagement and recorded drinking behavior among existing Drink Less users (ie, those who downloaded the app prelockdown). Second, we examined whether the lockdown led to a change in new app downloads. Third, to understand how the lockdown may have impacted the characteristics of users downloading the Drink Less app, we analyzed whether the sociodemographic characteristics of users, engagement with the app, and drinking behaviors recorded in the app differed between new users downloading the app prelockdown and during the lockdown.

This study addresses the following 3 research questions (RQs):

RQ1: Was the first UK lockdown associated with an immediate change in existing users of the Drink Less app in terms of:

- The depth of use (percent of available screens viewed)
- The amount of use (mean time spent on the app)
- The frequency of use (number of sessions)
- The number of alcohol units recorded each day
- The number of alcohol-free days recorded each day

XSL•FO
RenderX

- The number of heavy-drinking days recorded each day

RQ2: Was the first UK lockdown associated with a change in the number of new Drink Less downloads per day?

RQ3: Was the first UK lockdown associated with an immediate change in new users of the Drink Less app in terms of:

- Sociodemographic and drinking characteristics at baseline and in the 28 days following app download
- The depth of use (percentage of available screens viewed)
- The amount of use (mean time spent on the app)
- The frequency of use (number of sessions and number of days used)
- The number of alcohol units recorded
- The number of alcohol-free days recorded
- The number of heavy-drinking days recorded

Methods

Design

This was a natural experiment without active recruitment.

Intervention and Study Period

For all RQs, the interruption was conceptualized as the introduction of national lockdown measures in the United Kingdom on March 24, 2020. Due to differences in the way that the independent variables were operationalized (see the Analysis section), the time periods for RQ1, RQ2, and RQ3a differed from RQ3b-g. Specifically, for RQ1, each of the outcome variables was aggregated at a daily level by the number of active users in that week. For RQ2, the number of new downloads each day was attributed to the day of download. For RQ3a, baseline AUDIT scores and sociodemographic variables were measured once and attributed to the day of download. As such, data were collected for the 468 days between March 24, 2019, and July 3, 2020. This captured the period of time up until pubs in England reopened on July 4, 2020, which could have had a stepped effect on alcohol consumption. This period was divided into pre- (March 24, 2019-March 23, 2020; 366 days) and during-lockdown (March 24, 2020-July 3, 2020; 102 days) segments. For RQ3b-g, the depth, amount, and frequency of use, along with the number of alcohol units, alcohol-free days, and heavy-drinking days, recorded were aggregated over the 28-day period following app download by the number of users who downloaded the app on that day. Therefore, to limit potential confounding after pubs reopened, only respondents downloading the app a full 28 days prior to when pubs reopened (up to June 6, 2020) were included. Therefore, the study period for RQ3b-g was March 24, 2019-June 6, 2020 (441 days) and was divided into pre- (March 24, 2019-March 23, 2020; 366 days) and during-lockdown (March 24, 2020-June 6, 2020; 75 days) segments.

Study Population

The sample was UK users who downloaded the Drink Less app from Apple App Store, where it is freely available. To be eligible for inclusion, users had to be aged 18 years or older, based in the United Kingdom, interested in drinking less (specified when downloading the app), and have agreed to the privacy policy and terms and conditions within the app, as well as completing the AUDIT. For RQ1, existing users were defined as all those who downloaded the app between March 24, 2019, and the March 23, 2020 (prelockdown). RQ1 focused on existing, regular users, defined as use of the app at least once a week for a minimum of 4 weeks. For RQ2 and RQ3a, new users were defined as those who downloaded the app between March 24, 2020, and July 3, 2020, with no limits on regularity of use for new or existing users. Finally, for RQ3b-g, new users were defined as those who downloaded the app between March 24, 2020, and June 6, 2020, with no limits on regularity of use for new or existing users.

Measures

Sociodemographic and Drinking Characteristics

Three sociodemographic characteristics were measured at download. These were age (in years, continuous), sex (percentage female), and employment type (percentage nonmanual). The AUDIT was asked of all users providing both an AUDIT score (continuous) and the percentage of increasing and higher-risk drinkers (AUDIT score>=8).

Number of Downloads

The number of new app downloads each day was recorded.

Engagement Indicators

Three indicators of user engagement were derived from screen view records for each user: (1) number of sessions (where a new session is defined as a new screen view after 30 minutes of inactivity), (2) time spent on the app in minutes, and (3) percentage of available screens viewed. For RQ1, these measures were aggregated at a daily level across active users and attributed to the day of engagement. For RQ3b-g, each measure was aggregated over the 28-day period following app download for each user and was attributed to the date of download. Due to the differences in aggregation, the number of days used was also included as a measure of engagement for RQ3.

Drinking Measures

In the app, users were prompted to fill in a daily drinking calendar, where they either marked days as “alcohol free” or entered any alcoholic drinks they drank that day. Three drinking variables were calculated: (1) number of alcohol units (UK unit=10 mL of ethanol) consumed (aggregated daily), (2) number of alcohol-free days, and (3) number of heavy-drinking days (defined as >6 alcohol units). As described before, these measures were operationalized differently for RQ1 and RQ3.

Analysis

All analyses were conducted in R Studio. The engagement measures were derived using Pandas, a Python framework, within a Jupyter Notebook, an open source web application.

Models 1a-1f (for RQ1) examined whether the lockdown was associated with the percentage of available screens viewed, the mean time spent on the app, the number of sessions on the app, the number of alcohol units, the number of alcohol-free days, and the number of heavy-drinking days among existing, regular users of the Drink Less app.

XSL•FO
RenderX

Model 2 (for RQ2) examined whether the lockdown was associated with a change in the number of new daily downloads of the Drink Less app.

Models 3a-3e (for RQ3a) examined whether the lockdown was associated with changes in age, the proportion of female users, the proportion of nonmanual users, the proportion of users who were at risk of alcohol dependence, and the AUDIT scores among new downloaders of the Drink Less app.

Models 3f-3l (for RQ3b-g) examined whether there were changes in the percentage of available screens viewed, the mean time spent on the app, the number of sessions on the app, the number of days the app was used, the number of units, the number of alcohol-free days, and the number of heavy-drinking days recorded among new users following the first UK lockdown.

To estimate the associations between the lockdown and each of the outcomes, we conducted separate interrupted time series analyses using generalized additive mixed models (GAMMs). Analyses were conducted at the daily aggregated level while controlling for day of the week and month of the year. Smoothing "splines" were fitted in order to account for seasonal nonlinear variations in, for example, drinking behavior. To account for differences in the trends prelockdown and during the lockdown, the regression models included terms for the baseline level for each outcome prelockdown, the trend in the prelockdown period, the level change in the outcome immediately after the lockdown, and the trend in the during-lockdown period.

Plots of the autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) were used to test for both autoregressive (AR) and moving average (MA) autocorrelation over time. The ACF plots were used to identify plausible values for AR and MA terms for the baseline model. Models with various plausible AR and MA terms were compared with our baseline model using the Akaike information criterion (AIC), where smaller values indicate a better model fit.

As little was known about how the trends in each of the outcome variables during the lockdown, secondary analyses assessed whether regression models with nonlinear trends (cubic and quadratic) provided a better fit to the data. Best-fitting models were selected with the AIC, and where appropriate, cubic and quadratic models were reported.

All continuous variables were normally distributed, but a negative binomial distribution was used for the number of new downloads (RQ2), as the outcome variable was operationalized as a discrete (rather than continuous) variable and overdispersion was present. Data and details for each model (ie, AR and MA terms, AICs) are available in the annotated R code and can be accessed through GitHub [33].

Sensitivity Analyses

Preplanned sensitivity analyses (adjusting the date of the interruption to that on which social distancing measures were introduced, March 17, 2020, and controlling for potential confounders) were not conducted due to the complexity of the paper, the practical constraints associated with running additional analyses, and the robust model selection approach already taken.

Ethical Considerations

Ethical approval was obtained from University College London's Research Ethics Committee (CEHP/2016/556; CEHP/2020/579), and participants provided online consent to having their anonymous data used for scientific research purposes.

Results

Descriptive Statistics

Table 1 reports descriptive statistics for the outcome variables for each of the RQs.

XSL•FO
RenderX

Table 1. Descriptive statistics for the aggregated outcome variables of interest, stratified by period (prelockdown vs during the lockdown).

RQs[a]	Entire period (441 days)	Entire period (468 days)	Prelockdown (366 days)	During the lockdown (75 days)[b]	During the lockdown (102 days)
RQ1, mean (SD)					
Mean percentage screens viewed	N/A[c]	4.66 (0.80)	4.80 (0.83)	N/A	4.16 (0.30)
Mean time spent on app (minutes)	N/A	1.58 (0.72)	1.73 (0.73)	N/A	1.05 (0.25)
Mean number of logins	N/A	0.95 (0.11)	0.97 (0.11)	N/A	0.89 (0.06)
Mean alcohol units per day	N/A	3.29 (1.52)	3.38 (1.55)	N/A	2.98 (1.36)
Mean heavy-drinking days	N/A	0.21 (0.10)	0.21 (0.10)	N/A	0.18 (0.09)
Mean alcohol-free days	N/A	0.46 (0.16)	0.48 (0.16)	N/A	0.42 (0.16)
RQ2, median (IQR)[d]					
Number of new downloads per day	N/A	14.50 (13.00)	18.00 (13.00)	N/A	7.5 (6.00)
RQ3a, mean (SD)					
Gender (proportion female)	N/A	0.54 (0.16)	0.53 (0.15)	N/A	0.57 (0.21)
Age (years)	N/A	44.28 (4.07)	44.16 (3.64)	N/A	44.73 (5.34)
Employment type (proportion nonmanual)	N/A	0.71 (0.15)	0.71 (0.13)	N/A	0.69 (0.19)
AUDIT score	N/A	16.47 (2.41)	16.40 (2.06)	N/A	16.71 (3.40)
At risk of alcohol dependence (proportion at risk)	N/A	0.91 (0.09)	0.91 (0.08)	N/A	0.90 (0.11)
RQ3b-e, mean (SD)					
Number of logins	15.18 (6.60)	N/A	14.80 (5.80)	17.05 (9.42)	N/A
Number of days used	9.66 (3.41)	N/A	9.46 (2.96)	10.65 (4.97)	N/A
Percentage screens viewed	0.31 (0.04)	N/A	0.30 (0.04)	0.31 (0.06)	N/A
Time spent on app (minutes)	39.08 (21.64)	N/A	38.73 (20.28)	40.81 (27.51)	N/A
Alcohol units	72.21 (36.65)	N/A	70.60 (31.51)	80.15 (55.10)	N/A
Heavy-drinking days	4.44 (2.19)	N/A	4.35 (1.63)	4.90 (3.94)	N/A
Alcohol-free days	10.70 (5.42)	N/A	10.74 (5.31)	10.48 (5.93)	N/A

[a]RQ: research question.

[b]Shorter during-lockdown period as outcome variables 3b-e were aggregated over 28 days rather than at a monthly level.

[c]N/A: not applicable.

[d]Median (IQR) presented here; a large variance could lead to a skewed mean.

Association of the First UK Lockdown With Engagement With the Drink Less App Among Existing, Regular Users (RQ1)

There was an overall decline in the mean time spent on the app, the mean units recorded per day, and the mean heavy-drinking days recorded during the study period, with no step changes following the first COVID-19 lockdown. However, there was a change in slope during the lockdown, with a significant increase in the trajectory in the mean time spent on the app, the mean units recorded per day, and the mean heavy-drinking days recorded following the lockdown, though the magnitude of the change in these daily trends appeared small (Table 2).

There was an overall decline in the mean percentage of screens viewed and the mean number of sessions during the study period, with no step change following the first COVID-19 lockdown in the United Kingdom. The declining trend plateaued during the lockdown, with no significant trend in the mean percentage of screens viewed or the mean number of sessions (Table 2).

There was no significant trend in the mean alcohol-free days over the study period and no step change or change in slope following introduction of the first lockdown in the United Kingdom (Table 2 and Figure 1).

XSL•FO
RenderX

Table 2. Results of the best-fitting model for each outcome variable for RQ1[a] (N=468 days, range 9-598 users per day).

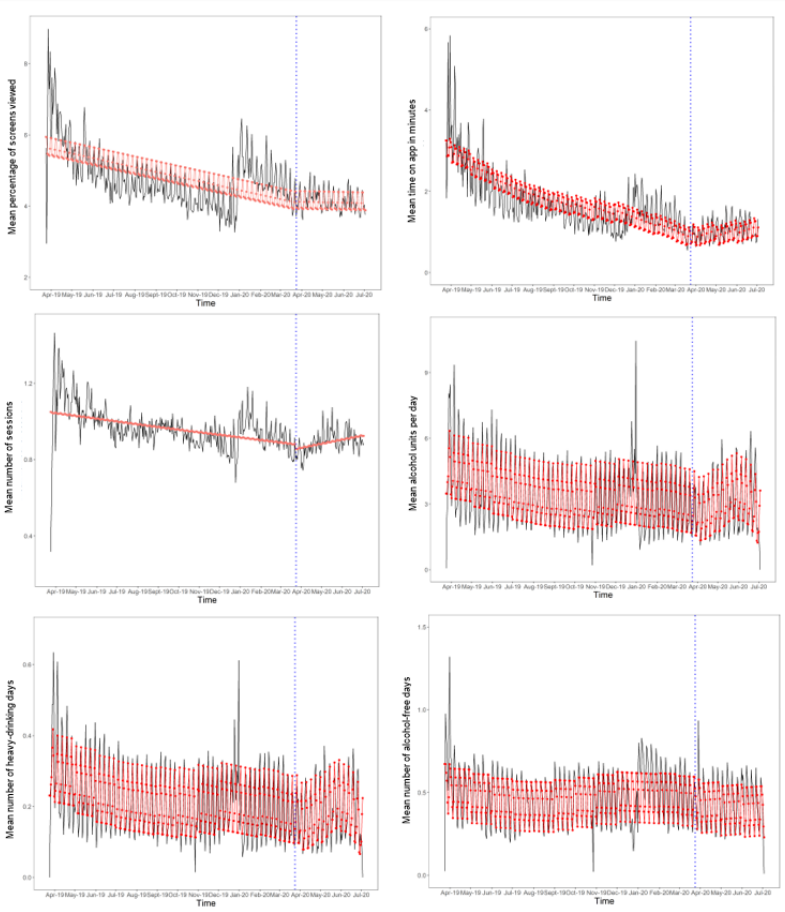
Outcome variables	B (95% CI)	*P* value
Mean percentage screens viewed[b], linear model		
Trend	–0.0042 (–0.0074 to –0.0010)	.01
Level	0.0204 (–0.5541 to 0.5949)	.95
Slope	0.0038 (–0.0105 to 0.0181)	.60
Mean time spent on app[a], linear model		
Trend	–0.0061 (–0.0081 to –0.0041)	.00
Level	0.0604 (–0.3230 to 0.4438)	.76
Slope	0.0118 (0.0027-0.0209)	.01
Mean number of sessions[b], linear model		
Trend	–0.0005 (–0.0007 to –0.0003)	.00
Level	–0.0247 (–0.1083 to 0.0589)	.56
Slope	0.0012 (–0.0001 to 0.0025)	.09
Mean alcohol units per day[b], cubic model		
Trend	–0.0049 (–0.0066 to –0.0032)	.00
Level	0.0351 (–0.9218 to 0.9920)	.94
Slope	–0.0297 (–0.1065 to 0.0471)	.45
$Slope^2$	0.0016 (–0.0001 to 0.0033)	.07
$Slope^3$	0.0000 (0.0000-0.0000)	.02
Mean heavy-drinking days[b], cubic model		
Trend	–0.0004 (–0.0005 to –0.0003)	.00
Level	0.0106 (–0.0498 to 0.0710)	.73
Slope	–0.0019 (–0.0068 to 0.0030)	.44
$Slope^2$	0.0001 (0.0000-0.0002)	.06
$Slope^3$	0.0000 (0.0000-0.0000)	.02
Mean alcohol-free days[b], linear model		
Trend	–0.0002 (–0.0004 to 0.0000)	.10
Level	–0.0183 (–0.1101 to 0.0735)	.70
Slope	0.0005 (–0.0009 to 0.0019)	.44

[a]RQ: research question.

[b]Adjusted for month of the year (cubic spline), day of the week (cubic spline), and autocorrelation.

XSL•FO
RenderX

Figure 1. Engagement indicators among existing, regular users of the Drink Less app over the study period (RQ1a-f). The red line indicates fitted values, the gray area indicates the 95% CI, and the dashed blue line indicates the interruption (ie, the first national UK lockdown). RQ: research question. Higher-resolution version of this figure is available in Multimedia Appendix 1.



Effect of the First UK Lockdown on the Number of Drink Less Downloads er Day (RQ2)

There was an overall declining trend in the number of downloads per day over the full study period, with no step change detected following the first COVID-19 lockdown in the United Kingdom. However, the declining trend plateaued during the lockdown, with no significant trend in downloads per day (Table 3 and Figure 2).

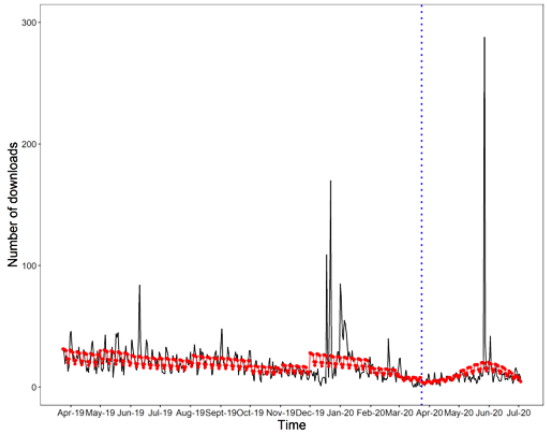
Table 3. Results of the best-fitting model for the number of downloads (N=468 days; range 0-288 downloads per day).

Number of downloads[a], cubic model	Incidence rate ratio (95% CI)	*P* value
Trend	0.9962 (0.9945-0.9979)	.00
Level	0.5365 (0.2332-1.2342)	.14
Slope	1.0161 (0.9495-1.0875)	.64
Slope^2	1.0006 (0.9991-1.0021)	.45
Slope^3	1.0000 (1.0000-1.0000)	.18

[a]Adjusted for month of the year (cubic spline), day of the week (cubic spline), and autocorrelation.

XSL•FO
RenderX

Figure 2. Number of new Drink Less downloads per day over the study period (RQ2). The red line indicates fitted values, the gray area indicates the 95% CI, and the dashed blue line indicates the interruption. RQ: research question.



Effect of the First UK Lockdown on Sociodemographic and Drinking Characteristics Among New Users of the Drink Less App (RQ3a)

There was no significant overall trend detected in the proportion of female users over the study period or the proportion of nonmanual (vs manual) workers, with no step change following the lockdown. However, there was a change in slope following the introduction of the first UK lockdown to a significant upward trajectory in the proportion of female users and a significant negative trajectory in the proportion of nonmanual workers (Table 4 and Figure 3).

There was no significant overall trend in the proportion of new users who were at risk of alcohol dependence over the study period. However, there was a step decrease during the lockdown, followed by a change in slope to an upward trend, whereby the proportion of new users who were at risk of alcohol dependence increased after the first lockdown in the United Kingdom, though the magnitude of this trend appeared small (Table 4).

We did not detect a significant trend over the study period or a step change or change in slope following the introduction of the first UK lockdown in age or AUDIT scores of new app users (Table 4).

XSL•FO
RenderX

Table 4. Results of the best-fitting model for each outcome variable for RQ3a[a] (N=440 days; range 1-245 users per day).

Outcome variables	B (95% CI)	*P* value
Gender[b,c], linear model		
Trend	–0.0001 (–0.0003 to 0.0001)	.10
Level	0.0109 (–0.0692 to 0.0910)	.79
Slope	0.0013 (0.0001-0.0025)	.04
Age[b,c], linear model		
Trend	0.0006 (–0.0036 to 0.0048)	.77
Level	0.3648 (–1.5329 to 2.2625)	.71
Slope	0.0015 (–0.0272 to 0.0302)	.92
Employment type[b], linear model		
Trend	0.0000 (–0.0001 to 0.0001)	.65
Level	0.0367 (–0.0271 to 0.1005)	.26
Slope	–0.0010 (–0.0020 to 0.0000)	.04
AUDIT[d] score[b], linear model		
Trend	0.0020 (–0.0003 to 0.0043)	.09
Level	–0.3994 (–1.4571 to 0.6583)	.46
Slope	0.0044 (–0.0116 to 0.0204)	.59
At risk of alcohol dependence[b,c], quadratic model		
Trend	0.0000 (–0.0001 to 0.0001)	.73
Level	–0.0596 (–0.1060 to –0.0132)	.01
Slope	0.0028 (0.0009-0.0047)	.004
Slope^2	0.0000 (0.0000-0.0000)	.01

[a]RQ: research question.

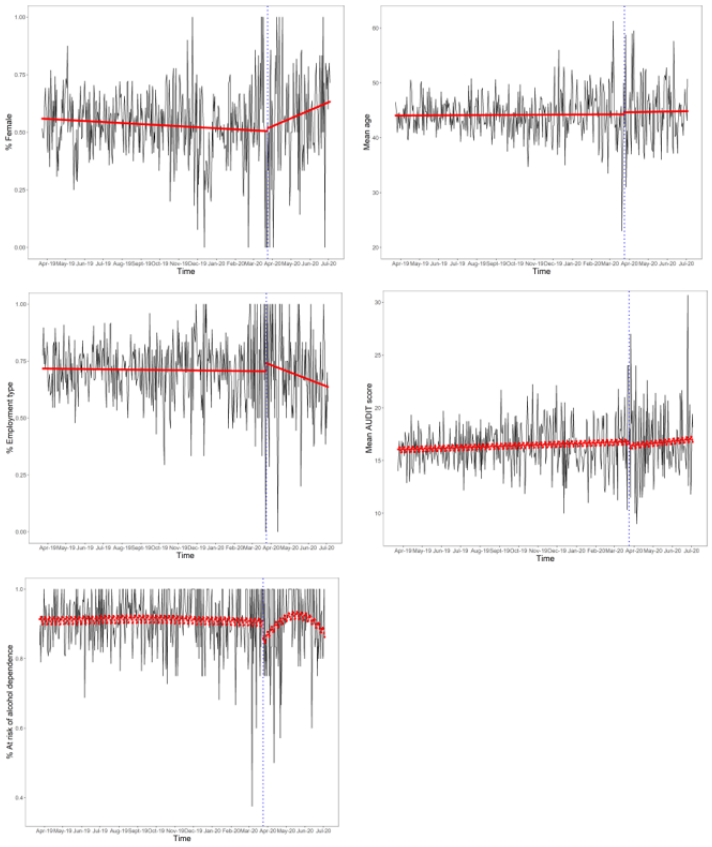
[b]Adjusted for month of the year (cubic spline) and day of the week (cubic spline).

[c]Adjusted for autocorrelation.

[d]AUDIT: Alcohol Use Disorders Identification Test.

XSL•FO
RenderX

Figure 3. Sociodemographic and drinking characteristics of new users of the Drink Less app over the study period (RQ3a). The red line indicates fitted values, the gray area indicates the 95% CI, and the dashed blue line indicates interruption. AUDIT: Alcohol Use Disorders Identification Test; RQ: research question. Higher-resolution version of this figure is available in Multimedia Appendix 1.



Effect of the First UK Lockdown on Engagement With the Drink Less App Among New Users (RQ3b-g)

There was no significant trend in the number of days used by new users across the study period and no change in slope. However, there was a step increase in the number of days used by new users immediately following the introduction of the first UK lockdown (Table 5).

There was an overall upward trend in terms of the percentage of available screens viewed across the study period by new users. There was a step decrease immediately following the introduction of the first UK lockdown, and the upward trend stabilized, with no significant trend during the lockdown (Table 5 and Figure 4).

There were no significant trends across the whole study period for mean alcohol units or heavy-drinking days reported by new users. However, there was a step increase for both following the introduction of the first UK lockdown but no significant change in slope.

There was an overall upward trend in alcohol-free days reported over the whole study period by new users, with no significant step change. The upward trend stabilized, with no significant trend during the lockdown.

There was no overall trend in the number of logins or time spent on the app in minutes across the whole study period for new users, with no significant step change and no significant change in slope following the introduction of the first UK lockdown.

XSL•FO
RenderX

Table 5. Results of the best-fitting model for each outcome variable for RQ3b-g[a] (N=440 days; range 1-245 users per day).

Outcome variables	B (95% CI)	*P* value
Number of logins[b], linear model		
Trend	–0.0063 (–0.0126 to 0.0000)	0.05
Level	2.6500 (–0.6376 to 5.9376)	0.12
Slope	0.0262 (–0.0437 to 0.0961)	0.46
Number of days used[b], linear model		
Trend	–0.0022 (–0.0055 to 0.0011)	0.19
Level	2.0484 (0.3449-3.7519)	0.02
Slope	–0.0099 (–0.0461 to 0.0263)	0.59
Percentage screens viewed[b], linear model		
Trend	0.0001 (0.0000-0.0002)	0.00
Level	–0.0265 (–0.0511 to –0.0019)	0.04
Slope	0.0000 (–0.0005 to 0.0005)	0.87
Time spent on app (minutes)[b,c], linear model		
Trend	–0.0188 (–0.0408 to 0.0032)	0.09
Level	0.9802 (–10.4248 to 12.3852)	0.87
Slope	0.1412 (–0.1013 to 0.3837)	0.25
Alcohol units[b], linear model		
Trend	–0.0314 (–0.0668 to 0.0040)	0.08
Level	20.1247 (1.7605-38.4889)	0.03
Slope	–0.0975 (–0.4880 to 0.2930)	0.62
Heavy-drinking days[b,c], linear model		
Trend	–0.0019 (–0.0039 to 0.0001)	0.06
Level	1.3845 (0.3318-2.4372)	0.01
Slope	–0.0105 (–0.0329 to 0.0119)	0.36
Alcohol-free days[b,c], linear model		
Trend	0.0083 (0.0028-0.0138)	0.003
Level	–1.7637 (–4.3836 to 0.8562)	0.19
Slope	–0.0144 (–0.0692 to 0.0404)	0.61

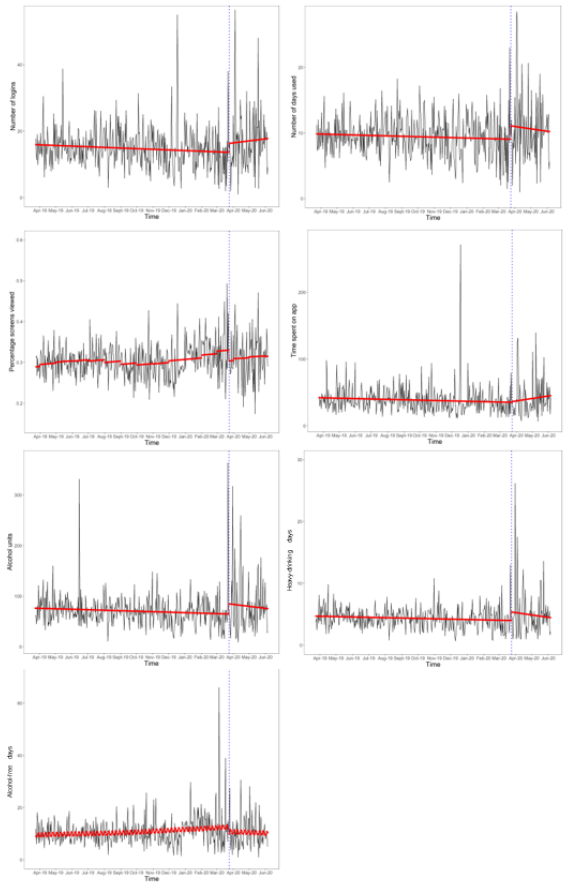
[a]RQ: research question.

[b]Adjusted for month of the year (cubic spline) and day of the week (cubic spline).

[c]Adjusted for autocorrelation.

XSL•FO
RenderX

Figure 4. Aggregated engagement indicators among new users of the Drink Less app over the study period (RQ3b-e). The red line indicates fitted values, the gray area indicates the 95% CI, and the dashed blue line indicates the interruption. RQ: research question. Higher-resolution version of this figure is available in Multimedia Appendix 1.



Discussion

Principal Findings

Following the first COVID-19 lockdown in the United Kingdom, there was a significant increase in the time spent on the app and in the mean alcohol units per day and the number of heavy-drinking days recorded by existing, regular users of the Drink Less app, although no change was detected in the percentage of screens viewed, the number of sessions logged, or in the number of alcohol-free days recorded. There was no increase in downloads per day following the lockdown, although the overall negative trend in new daily app downloads plateaued following the introduction of the first UK lockdown. Among the new users of Drink Less, there were increases in the proportion of female users, manual workers, and those at risk of alcohol dependence following the first UK lockdown. With regard to changes in engagement indicators, there was a step increase in the number of days the app was used but a step decrease in the percentage of available screens viewed within Drink Less, suggesting that users engaged with the app for a longer period but with less of the available content. In terms of drinking characteristics, new users reported step increases in the mean number of alcohol units and heavy-drinking days aggregated over 28 days after app download following the first UK lockdown.

Strengths and Limitations

A strength of this study is that it was a natural experiment based on longitudinal data exploring self-motivated engagement with a freely available alcohol reduction app in the real world in a large sample against the backdrop of a global pandemic. However, there are also limitations associated with this approach. The period here reflects the immediate effects of COVID-19, which may not have been consistent over a longer period. These findings were also isolated to the United Kingdom. The app is reliant on self-reported alcohol consumption data. It is possible that changes in drinking contexts during the lockdowns could have affected the accuracy of self-report data. People were likely to be drinking in smaller groups in private, rather than public, settings, where they may have been more willing or less likely to forget to log drinks. Conversely, logged drinks during the lockdown might be more likely to be underestimated as individuals pouring their own drinks at home may be less likely to use standard measures than in on-trade settings. Finally, although the GAMMs incorporated cyclic cubic terms for day and month, additional seasonality terms may have further improved the model fit. This should be explored in future research involving app data. A recent study showed that COVID-19 had different effects on health behaviors in different countries, increases in alcohol consumption during the early months of the pandemic were recorded in the United Kingdom and Ireland, and decreases in consumption were recorded among 20 other European countries within the same

XSL•FO
RenderX

period [36]. As such, these findings are unlikely to be generalizable to other countries. Furthermore, we were unable to account for differences in living situations, such as whether an individual was furloughed or a parent, which have both been shown to be associated with changes in drinking in lockdown [6].

Avenues for Future Research

This research focused on 1 digital intervention, and it would be of interest to attempt to triangulate these findings across other forms of digital support available in the United Kingdom and internationally. This would aid in building a comprehensive overview of the effect of the COVID-19 pandemic on the use of digital alcohol reduction support and drinking patterns. The longer-term impact of the ongoing pandemic on engagement with digital interventions is also of interest. Furthermore, although increased engagement is positive, it is also necessary to examine the success of alcohol reduction attempts and whether they can be better supported.

Implications for Policy and Practice

This study indicated increases in units of alcohol consumed and heavy-drinking days among both existing and new users of the Drink Less app following the first UK lockdown. This is in line with other research outlining the polarizing impact of the first UK lockdown on alcohol consumption [6,12], with increases in consumption seen in increasing and higher-risk drinkers [4] and increases in the frequency of heavy episodic drinking [6]. There was no change in alcohol-free days recorded, suggesting increases in the amount consumed but not the frequency of drinking. Alongside this, there was some indication of increased engagement, with increases in the time spent on the app among existing users and increases in the number of days used but declines in the number of screens used among new users. This supports research outlining increases in self-reported reduction attempts by high-risk drinkers [4]. Increased engagement with the Drink Less app could have been partly due to reductions in the availability of health care services throughout the first lockdown, though this is speculative.

There were also increases in the proportion of female users, manual workers, and those at risk of alcohol dependence following the first UK lockdown. Shifts in engagement with the Drink Less app may be linked to more dramatic changes in lifestyles throughout the lockdown. There is evidence that women were disproportionately affected by additional caring responsibilities during the lockdowns [34,35] and reported greater reductions in well-being [29,37]. Those working in manual professions may have been more likely to have been furloughed than those working in office jobs who could work remotely, which could have resulted in more time to engage with the Drink Less app. Increases in engagement among those at risk of alcohol dependence could be particularly promising as heavier drinking is associated with increased risks of harm [1]. Increases in the proportion of heavier drinkers during the pandemic has persisted in the longer-term postlockdown in the United Kingdom, resulting in significantly increased health and economic burden in England [38,39]. As such, it is important to capitalize on increased interest in alcohol reduction by increasing funding and resources for alcohol support services. Alcohol reduction campaigns specifically targeted at those who have increased their consumption throughout the pandemic may also be useful.

Conclusion

Following the first COVID-19 lockdown in the United Kingdom, there is some evidence of increased engagement with the alcohol reduction app Drink Less, the previously negative trend in new downloads plateaued, and there was an increase in the time spent on the app among regular, existing users and a step increase in the number of days used among new users during the first lockdown. However, there was also evidence of increased alcohol consumption in the first lockdown, with increases in units consumed and heavy-drinking days among existing users and step increases in units consumed and heavy-drinking days among new users.

Acknowledgments

MO, GL, and CG are funded by the National Institute for Health Research (NIHR) Public Health Research Programme (project reference NIHR127651). CG and OP are funded by Cancer Research UK (CRUK; C1417/A22962). JB is a member of the SPECTRUM consortium. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any author-accepted manuscript version arising.

Conflicts of Interest

MO and CG are paid scientific consultants for the behavior change and lifestyle organization One Year No Beer. JB has received unrestricted research funding to study smoking cessation from companies who manufacture smoking cessation medications (Pfizer and J&J). OP and GL declare no conflicts of interest.

Multimedia Appendix 1
Higher resolution versions of Figure 1-Figure 4.
[DOCX File , 2242 KB - jmir_v24i11e42320_app1.docx]

References

XSL•FO
RenderX

1. Corrao G, Bagnardi V, Zambon A, Arico S. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: a meta-analysis. Addiction 1999 Oct;94(10):1551-1573. [doi: 10.1046/j.1360-0443.1999.9410155111.x] [Medline: 10790907]
2. Connor J. Alcohol consumption as a cause of cancer. Addiction 2017 Feb 21;112(2):222-228. [doi: 10.1111/add.13477] [Medline: 27442501]
3. Griswold M, Fullman N, Hawley C, GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018 Sep;392(10152):1015-1035 [FREE Full text] [doi: 10.1016/S0140-6736(18)31310-2]
4. Jackson SE, Garnett C, Shahab L, Oldham M, Brown J. Association of the COVID-19 lockdown with smoking, drinking and attempts to quit in England: an analysis of 2019-20 data. Addiction 2021 May 26;116(5):1233-1244 [FREE Full text] [doi: 10.1111/add.15295] [Medline: 33089562]
5. Jackson S, Beard E, Angus C, Field M, Brown J. Moderators of changes in smoking, drinking and quitting behaviour associated with the first COVID-19 lockdown in England. Addiction 2022 Mar;117(3):772-783 [FREE Full text] [doi: 10.1111/add.15656] [Medline: 34431577]
6. Oldham M, Garnett C, Brown J, Kale D, Shahab L, Herbec A. Characterising the patterns of and factors associated with increased alcohol consumption since COVID-19 in a UK sample. Drug Alcohol Rev 2021 Sep;40(6):890-899 [FREE Full text] [doi: 10.1111/dar.13256] [Medline: 33657644]
7. Brundtland GH. From the World Health Organization. Reducing risks to health, promoting healthy life. JAMA 2002 Oct 23;288(16):1974. [doi: 10.1001/jama.288.16.1974] [Medline: 12387638]
8. App Store Preview. Drink Less. URL: https://apps.apple.com/gb/app/drink-less/id1020579244 [accessed 2022-10-17]
9. Institute for Government. Timeline of UK Lockdowns, March 2020 to March 2021. URL: https://www.instituteforgovernment.org.uk/sites/default/files/timeline-lockdown-web.pdf [accessed 2022-10-17]
10. Behavioural Risk Factors. URL: https://analytics.phe.gov.uk/apps/covid-19-indirect-effects/# [accessed 2022-10-17]
11. Anderson P, O'Donnell A, Jané Llopis E, Kaner E. The COVID-19 alcohol paradox: British household purchases during 2020 compared with 2015-2019. PLoS One 2022;17(1):e0261609 [FREE Full text] [doi: 10.1371/journal.pone.0261609] [Medline: 35045099]
12. Garnett C, Jackson S, Oldham M, Brown J, Steptoe A, Fancourt D. Factors associated with drinking behaviour during COVID-19 social distancing and lockdown among adults in the UK. Drug Alcohol Depend 2021 Feb 01;219:108461 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108461] [Medline: 33454159]
13. Hardie I, Stevely AK, Sasso A, Meier PS, Holmes J. The impact of changes in COVID-19 lockdown restrictions on alcohol consumption and drinking occasion characteristics in Scotland and England in 2020: an interrupted time-series analysis. Addiction 2022 Jun;117(6):1622-1639 [FREE Full text] [doi: 10.1111/add.15794] [Medline: 35108758]
14. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Jul;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
15. Deloitte. Global Mobile Consumer Survey: UK Cut; Plateauing at the Peak - The state of the smartphone. URL: https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-plateauing-at-the-peak-the-state-of-the-smartphone.pdf [accessed 2022-06-10]
16. Perski O, Jackson SE, Garnett C, West R, Brown J. Trends in and factors associated with the adoption of digital aids for smoking cessation and alcohol reduction: a population survey in England. Drug Alcohol Depend 2019 Dec 01;205:107653 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.107653] [Medline: 31675544]
17. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun 13;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
18. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
19. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]
20. Garnett C, Crane D, West R, Brown J, Michie S. The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers. Transl Behav Med 2019 Mar 01;9(2):296-307 [FREE Full text] [doi: 10.1093/tbm/iby043] [Medline: 29733406]
21. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384-4307 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
22. Garnett C, Perski O, Michie S, West R, Field M, Kaner E, et al. Refining the content and design of an alcohol reduction app, Drink Less, to improve its usability and effectiveness: a mixed methods approach. F1000Res 2021 Jun 29;10:511. [doi: 10.12688/f1000research.51416.1]

XSL•FO
RenderX

23. Lewer D, Meier P, Beard E, Boniface S, Kaner E. Unravelling the alcohol harm paradox: a population-based study of social gradients across very heavy drinking thresholds. BMC Public Health 2016 Jul 19;16(1):599 [FREE Full text] [doi: 10.1186/s12889-016-3265-9] [Medline: 27430342]
24. Bloomfield K, Grittner U, Kramer S, Gmel G. Social inequalities in alcohol consumption and alcohol-related problems in the study countries of the EU concerted action 'Gender, Culture and Alcohol Problems: a Multi-national Study'. Alcohol Alcohol Suppl 2006;41(1):i26-i36. [doi: 10.1093/alcalc/agl073] [Medline: 17030500]
25. Jones L, Bates G, McCoy E, Bellis MA. Relationship between alcohol-attributable disease and socioeconomic status, and the role of alcohol consumption in this relationship: a systematic review and meta-analysis. BMC Public Health 2015 Apr 18;15(1):400 [FREE Full text] [doi: 10.1186/s12889-015-1720-7] [Medline: 25928558]
26. Selin K. Predicting alcohol-related harm by sociodemographic background: high prevalence versus high risk. Contemp Drug Probl 2018 Apr 24;32(4):547-588 [FREE Full text] [doi: 10.1177/009145090503200404]
27. Livingston M, Room R. Variations by age and sex in alcohol-related problematic behaviour per drinking volume and heavier drinking occasion. Drug Alcohol Depend 2009 May 01;101(3):169-175. [doi: 10.1016/j.drugalcdep.2008.12.014] [Medline: 19233572]
28. Bu F, Steptoe A, Fancourt D. Loneliness during a strict lockdown: trajectories and predictors during the COVID-19 pandemic in 38,217 United Kingdom adults. Soc Sci Med 2020 Nov;265:113521 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113521] [Medline: 33257177]
29. Fancourt D, Bu F, Mak HW, Steptoe A. COVID-19 Social Study: Results Release 16. URL: https://www.covidsocialstudy.org/_files/ugd/3d9db5_dc64263647624fd3842e6521c186aa69.pdf [accessed 2022-10-17]
30. Fancourt D, Bu F, Mak H, Steptoe A. Covid-19 Social Study: Results Release 29. URL: https://b6bdcb03-332c-4ff9-8b9d-28f9c957493a.filesusr.com/ugd/3d9db5_59d1b940054440bbb52a72b6bd0b0a06.pdf [accessed 2022-10-17]
31. Patel J, Nielsen F, Badiani A, Assi S, Unadkat V, Patel B, et al. Poverty, inequality and COVID-19: the forgotten vulnerable. Public Health 2020 Jun;183:110-111 [FREE Full text] [doi: 10.1016/j.puhe.2020.05.006] [Medline: 32502699]
32. Garnett C, Perski O, Beard E, Michie S, West R, Brown J. The impact of celebrity influence and national media coverage on users of an alcohol reduction app: a natural experiment. BMC Public Health 2021 Jan 06;21(1):30 [FREE Full text] [doi: 10.1186/s12889-020-10011-0] [Medline: 33407283]
33. OlgaPerski / DrinkLess_TSA. URL: https://github.com/OlgaPerski/DrinkLess_TSA [accessed 2022-10-18]
34. Farré L, Fawaz Y, González L, Graves J. Gender inequality in paid and unpaid work during covid‐19 times. Rev Income Wealth 2021 Dec 17;68(2):323-347 [FREE Full text] [doi: 10.1111/roiw.12563]
35. Craig L, Churchill B. Dual-earner parent couples' work and care during COVID-19. Gend Work Organ 2021 Jan 24;28(Suppl 1):66-79 [FREE Full text] [doi: 10.1111/gwao.12497] [Medline: 32837023]
36. Kilian C, Rehm J, Allebeck P, Braddick F, Gual A, Barták M, European Study Group on Alcohol UseCOVID-19. Alcohol consumption during the COVID-19 pandemic in Europe: a large-scale cross-sectional study in 21 countries. Addiction 2021 Dec 09;116(12):3369-3380. [doi: 10.1111/add.15530] [Medline: 34109685]
37. Etheridge B, Spantig L. The gender gap in mental well-being at the onset of the Covid-19 pandemic: evidence from the UK. Eur Econ Rev 2022 Jun;145:104114 [FREE Full text] [doi: 10.1016/j.euroecorev.2022.104114] [Medline: 35464832]
38. Boniface S. The COVID Hangover. URL: https://www.ias.org.uk/wp-content/uploads/2022/07/The-COVID-Hangover-report-July-2022.pdf [accessed 2022-10-17]
39. Angus C, Henney M, Pryce R. Modelling the Impact of Changes in Alcohol Consumption during the COVID-19 Pandemic on Future Alcohol-Related Harm in England. URL: https://tinyurl.com/27a5fr58 [accessed 2022-10-17]

Abbreviations

ACF: autocorrelation function
AIC: Akaike information criterion
AR: autoregressive
AUDIT: Alcohol Use Disorders Identification Test
GAMM: generalized additive mixed model
MA: moving average
PACF: partial autocorrelation function
RQ: research question

XSL•FO
RenderX

Edited by C Basch; submitted 31.08.22; peer-reviewed by C Kilian, M Kapsetaki; comments to author 25.09.22; revised version received 10.10.22; accepted 13.10.22; published 10.11.22.

Please cite as:
Oldham M, Perski O, Loebenberg G, Brown J, Garnett C
The Effect of the First UK COVID-19 Lockdown on Users of the Drink Less App: Interrupted Time Series Analysis of Sociodemographic Characteristics, Engagement, and Alcohol Reduction
J Med Internet Res 2022;24(11):e42320
URL: https://www.jmir.org/2022/11/e42320
doi:10.2196/42320
PMID:36240461

©Melissa Oldham, Olga Perski, Gemma Loebenberg, Jamie Brown, Claire Garnett. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Natural Language Processing to Explore "Dry January" Posts on Twitter: Longitudinal Infodemiology Study

Alex M Russell[1], PhD; Danny Valdez[2], PhD; Shawn C Chiang[1], MPH; Ben N Montemayor[3], PhD; Adam E Barry[3], PhD; Hsien-Chang Lin[2], PhD; Philip M Massey[1], PhD

[1]Center for Public Health and Technology, Department of Health, Human Performance, and Recreation, University of Arkansas, Fayetteville, AR, United States

[2]Department of Applied Health Science, School of Public Health, Indiana University, Bloomington, IN, United States

[3]Department of Health and Kinesiology, Texas A&M University, College Station, TX, United States

Corresponding Author:
Alex M Russell, PhD
Center for Public Health and Technology
Department of Health, Human Performance, and Recreation
University of Arkansas
HPER Building, 308-V
Fayetteville, AR, 72701
United States
Phone: 1 479 575 8672
Email: ar117@uark.edu

Abstract

Background: Dry January, a temporary alcohol abstinence campaign, encourages individuals to reflect on their relationship with alcohol by temporarily abstaining from consumption during the month of January. Though Dry January has become a global phenomenon, there has been limited investigation into Dry January participants' experiences. One means through which to gain insights into individuals' Dry January-related experiences is by leveraging large-scale social media data (eg, Twitter chatter) to explore and characterize public discourse concerning Dry January.

Objective: We sought to answer the following questions: (1) What themes are present within a corpus of tweets about Dry January, and is there consistency in the language used to discuss Dry January across multiple years of tweets (2020-2022)? (2) Do unique themes or patterns emerge in Dry January 2021 tweets after the onset of the COVID-19 pandemic? and (3) What is the association with tweet composition (ie, sentiment and human-authored vs bot-authored) and engagement with Dry January tweets?

Methods: We applied natural language processing techniques to a large sample of tweets (n=222,917) containing the term "dry january" or "dryjanuary" posted from December 15 to February 15 across three separate years of participation (2020-2022). Term frequency inverse document frequency, k-means clustering, and principal component analysis were used for data visualization to identify the optimal number of clusters per year. Once data were visualized, we ran interpretation models to afford within-year (or within-cluster) comparisons. Latent Dirichlet allocation topic modeling was used to examine content within each cluster per given year. Valence Aware Dictionary and Sentiment Reasoner sentiment analysis was used to examine affect per cluster per year. The Botometer automated account check was used to determine average bot score per cluster per year. Last, to assess user engagement with Dry January content, we took the average number of likes and retweets per cluster and ran correlations with other outcome variables of interest.

Results: We observed several similar topics per year (eg, Dry January resources, Dry January health benefits, updates related to Dry January progress), suggesting relative consistency in Dry January content over time. Although there was overlap in themes across multiple years of tweets, unique themes related to individuals' experiences with alcohol during the midst of the COVID-19 global pandemic were detected in the corpus of tweets from 2021. Also, tweet composition was associated with engagement, including number of likes, retweets, and quote-tweets per post. Bot-dominant clusters had fewer likes, retweets, or quote tweets compared with human-authored clusters.

Conclusions: The findings underscore the utility for using large-scale social media, such as discussions on Twitter, to study drinking reduction attempts and to monitor the ongoing dynamic needs of persons contemplating, preparing for, or actively pursuing attempts to quit or cut down on their drinking.

XSL•FO
RenderX

(J Med Internet Res 2022;24(11):e40160) doi:10.2196/40160

KEYWORDS

alcohol; drinking; social media; Twitter; Dry January; infodemiology; infoveillance; natural language processing

Introduction

Background

"Dry January"—a public health campaign aimed at encouraging individuals to reflect on their relationship with alcohol by temporarily abstaining from consumption during the month of January—originated in the United Kingdom in 2013 [1,2]. Those who register to participate in the month-long challenge via the Alcohol Change UK website are provided added accountability and support through access to interactive online resources (eg, TryDry mobile application) and health communication messaging highlighting the benefits of temporary alcohol abstinence (eg, emails and social media messaging about financial health, physical health, and mental health benefits) [3]. Dry January is theorized to confer benefits to participants via social contagion, which suggests widespread changes in health beliefs and behaviors are more likely to occur when a supportive community or subgroup of people endorse similar motivations and goals [4-6].

Prior research evaluating the characteristics of Dry January participants and the efficacy for the campaign in terms of reducing alcohol consumption and enhancing quality of life indicators has primarily focused on official Dry January registrants (ie, those who reside in the United Kingdom and officially registered for the challenge on the Alcohol Change UK website) [7-9]. Most of these studies have demonstrated that official participation in the temporary abstinence initiative is associated with numerous short- and long-term benefits, including reductions in alcohol consumption, increases in alcohol-refusal skills, saving money, improved sleep, increased energy, weight loss, and enhanced psychological well-being [5,7-9]. However, Case et al [10] found that increased participation in Dry January in England between 2015 and 2018 was not associated with population-level reductions in alcohol consumption over the 4-year period.

One potential explanation for these mixed findings could be that, although the number of officially registered Dry January participants in the United Kingdom has risen from 4000 in 2013 to 130,000 in 2021 [1], this represents only a small minority of the public who are informally participating in the temporary alcohol abstinence initiative (an estimated 6.5 million Britons reported planning to give up alcohol during the month of January in 2021) [11]. Additionally, the reach of the Dry January campaign has extended beyond the United Kingdom and has become a global cultural phenomenon with millions of informal participants worldwide [12]. For example, an estimated 15% to 19% of American adults reported going alcohol-free during January 2022 [13,14]. This has coincided with increasing news media attention [15,16], social media engagement, and Dry January–related alcohol industry promotional efforts (eg, marketing of nonalcoholic alternatives) [17]. For the millions of individuals who *unofficially* participate in alcohol abstinence during the month of January, there remains a paucity of investigations and a need to better understand their experiences in attempting to abstain from alcohol during the month of January. One such means through which to gain insights into individuals' Dry January-related experiences is by leveraging large-scale social media data (eg, Twitter chatter) to explore and characterize public discourse concerning Dry January.

Infodemiology

Infodemiology (the epidemiology of online information, such as using search result data or social media posts to inform public health and policy) and infoveillance (longitudinal tracking of online information for surveillance purposes) are emerging fields [18-21]. The last decade has witnessed a proliferation in Twitter and other social media platform usage, and many individuals rely on these platforms for health information [22-24]. Along these lines, infodemiology methods have been used to systematically monitor public sentiment and characterize communication concerning various health topics using publicly available social media data, such as Twitter posts [21]. Though not intended to replace, but rather complement, more traditional methods, infodemiology offers several advantages, including the ease and rapidity with which data can be collected, allowing for the ability to detect changes in public attention and attitudes in real time [18-20]. Previous studies leveraging Twitter as a data source have provided insights into a variety of health topics, including alcohol-related behaviors [25-28], tobacco use and cessation [29-32], drug use [33,34], mental health [35,36], vaccination [37,38], and the spread of health-related misinformation [39]. Moreover, Twitter has been used as a real-time surveillance tool to monitor reactions to public health prevention campaigns [40] and public policy changes [41,42], providing timely information to public health researchers, practitioners, and policy makers.

Alcohol Use Infodemiology on Twitter

A growing number of studies have explored alcohol-related, user-generated content posted on Twitter [25-28]. For instance, Cavazos-Rehg et al [25] was among the first to characterize a large sample of alcohol-related tweets, finding that the vast majority of such tweets expressed positive sentiment toward alcohol and frequently glamorized heavy drinking, while rarely portraying any alcohol-related negative consequences. Other studies have examined tweets concerning alcohol-related blackouts [26,28,43]; increases in alcohol-related blackout tweets in early 2020 were in line with population-level increases in alcohol consumption observed during the COVID-19 pandemic [28]. Weitzman et al [44] compared state-level alcohol use–related Twitter posts and Google Trends search data with 3 years of national epidemiological survey data, providing support for using search activity and social media data to complement epidemiological approaches to monitor alcohol use and inform prevention efforts. However, there has been a dearth of infodemiology studies focused on efforts to quit or cut down on drinking, such as drinking reduction attempts

XSL•FO
RenderX

associated with the Dry January temporary alcohol abstinence campaign [8,9].

This Study

The purpose of this study was to identify and describe a corpus of Dry January–related tweets authored by the public and social bots across 3 years of participation (2020-2022) and to evaluate whether there were changes in themes and sentiment from year to year in response to the COVID-19 pandemic. We sought to compare conversational themes over time to demonstrate the potential use for social media platforms—such as Twitter—to be used to study drinking reduction attempts and to monitor the ongoing dynamic needs of persons actively involved in or thinking about attempts to quit or cut down on drinking. To achieve this objective, we applied natural language processing (NLP) techniques to a large sample of Twitter data (n=222,917), spanning 3 distinct years (2020-2022), to answer the following research questions (RQs):

1. (RQ1) What themes are present within a corpus of tweets about Dry January, and is there consistency in the language used to discuss Dry January across multiple years of tweets (2020-2022)?
2. (RQ2) Do unique themes or patterns emerge in Dry January 2021 tweets after the onset of the COVID-19 pandemic?
3. (RQ3) What is the association between tweet composition (ie, sentiment and human-authored vs bot-authored) and engagement with Dry January tweets?

Methods

Data Collection

Tweets associated with this study, including metadata (eg, number of likes, retweets, replies) were extracted using the Twitter application programming interface (API) v2 and Python 3.9. After obtaining approval for access to the Academic Research product track of Twitter's API v2, we identified and extracted all tweets containing the term "dry january" or "dryjanuary" posted from December 15 to February 15 across 3 separate years of participation (12/15/2019 to 02/15/2020, 12/15/2020 to 02/15/2021, and 12/15/2021 to 02/15/2022). Capturing the 2 weeks prior to and after the month of January allowed us to analyze conversations related to anticipation of Dry January, as well as those reflecting on completed Dry January attempts (whether successful or unsuccessful). We excluded all retweets, defined as the same tweet appearing multiple times in the corpus, and non-English tweets, defined as any tweets not originally written in the English language. Note, eliminating duplicate tweets and non-English tweets was done to enhance the interpretability of the NLP analyses undertaken herein [45]. Overall, 70,215 tweets were extracted from 12/15/2019 to 2/15/2020, 86,378 tweets from 12/15/2020 to 2/15/2021, and 66,324 tweets from 12/15/2021 to 2/15/2022, resulting in a final sample of 222,917 tweets. All tweets collected for this study, inclusive of nonpersonally identifiable metadata, were saved into a secure repository only accessible by the research team, strictly conforming to standards for ethical data use and online privacy.

Ethical Considerations

Research procedures were deemed exempt by the appropriate institutional review board prior to data collection from Twitter.

Analyses

Our research questions were exploratory in nature. As such, we strategically selected several classes of computational informatics methods designed to extract overall themes in the corpus and project relative similarity and dissimilarity across themes. These methods can be classified into those used for data visualization (term frequency inverse document frequency [TF-IDF], k-means clustering, and principal component analysis [PCA]) and for data interpretation (latent Dirichlet allocation [LDA] topic models, Valence Aware Dictionary and Sentiment Reasoner [VADER] sentiment analysis, and Botometer automated account check).

Data Visualization (Research Questions 1 and 2)

Term Frequency Inverse Document Frequency

TF-IDF refers to an information retrieval technique used to transform text data into numeric data [46,47]. Specifically, the TF-IDF algorithm creates weights for each word in a corpus, such that weights implicate (1) how important a word is in a singular tweet relative to (2) the number of times the same word was used in the entirety of the corpus. Weights per term can be interpreted as greater values equating higher word importance and lower values equating lower term importance. These weights are then transposed into a sparse matrix for further analysis.

K-means Clustering

K-means clustering is an unsupervised machine learning tool used to group text content into themes, or clusters. This analysis relies on the sparse matrix created by the TF-IDF calculations to categorize tweets into one of the k-clusters. The optimal number of k clusters is identified by calculating the sums of squared differences for a range of possible clusters (ie, 1 cluster to 10 clusters). The sums of squared differences for a range of k clusters are plotted along an elbow scree plot, where breaks in a plotted line indicate a possible clusters solution. For more information on k-means clustering, please see Na et al [48].

PCA

PCA, a commonly used analysis in exploratory factor analysis, is a dimensionality technique used to reduce the complexity, or components, of data while still maintaining the integrity of the data [49,50]. For text mining analysis, all words assigned weights by TF-IDF that have been assigned into one of the k-clusters are reduced into simple X and Y coordinates. These coordinates are transposed onto a vector map and color coded along the predetermined optimal k-clusters. For this analysis, we examined data shape, which simply refers to the way in which data are presented on a vector map.

Data Interpretation (Research Questions 2 and 3)

LDA Topic Models

LDA refers to an unsupervised NLP method that uses probabilistic inferencing to identify latent topics within a corpus of similar content. LDA is widely acknowledged as the most effective and precise topic modeling algorithm and has been

XSL•FO
RenderX

widely applied for a variety of research areas and social issues [51,52].

VADER

VADER is a rule-based sentiment analysis attuned to social media vernacular [53,54]. VADER specifically examines the polarity of words in each tweet by feeding text data through a lexicon that is precoded with values for all positive and negative words in the English language. VADER scores can range from –.99 to .99. High values typically denote higher affect, or greater positivity, and lower values typically denote lower affect, or greater negativity.

Botometer

Botometer is a proprietary algorithm developed by the Indiana University Network Science Institute [55]. Botometer is widely used to determine if content in a tweet originates from an account that is principally human-authored or principally bot-authored. Users can leverage the Botometer API and search for specific user IDs or usernames and immediately receive a score from .01 to .99. Lower scores indicate that the account likely belongs to a human; higher scores, typically above .70, indicate that the account likely belongs to an automated bot. Note that, due to limitations with the Botometer API, we were only able to subsample 500 posts per cluster per year as a rough approximation of bot activity. Our decision to use a general .70 cutoff as a delineator between likely bot and likely human account is supported by Botometer validation literature and other studies leveraging Botometer for bot detection and removal [56,57].

Simple Inductive Coding and Validation (Research Questions 1, 2, and 3)

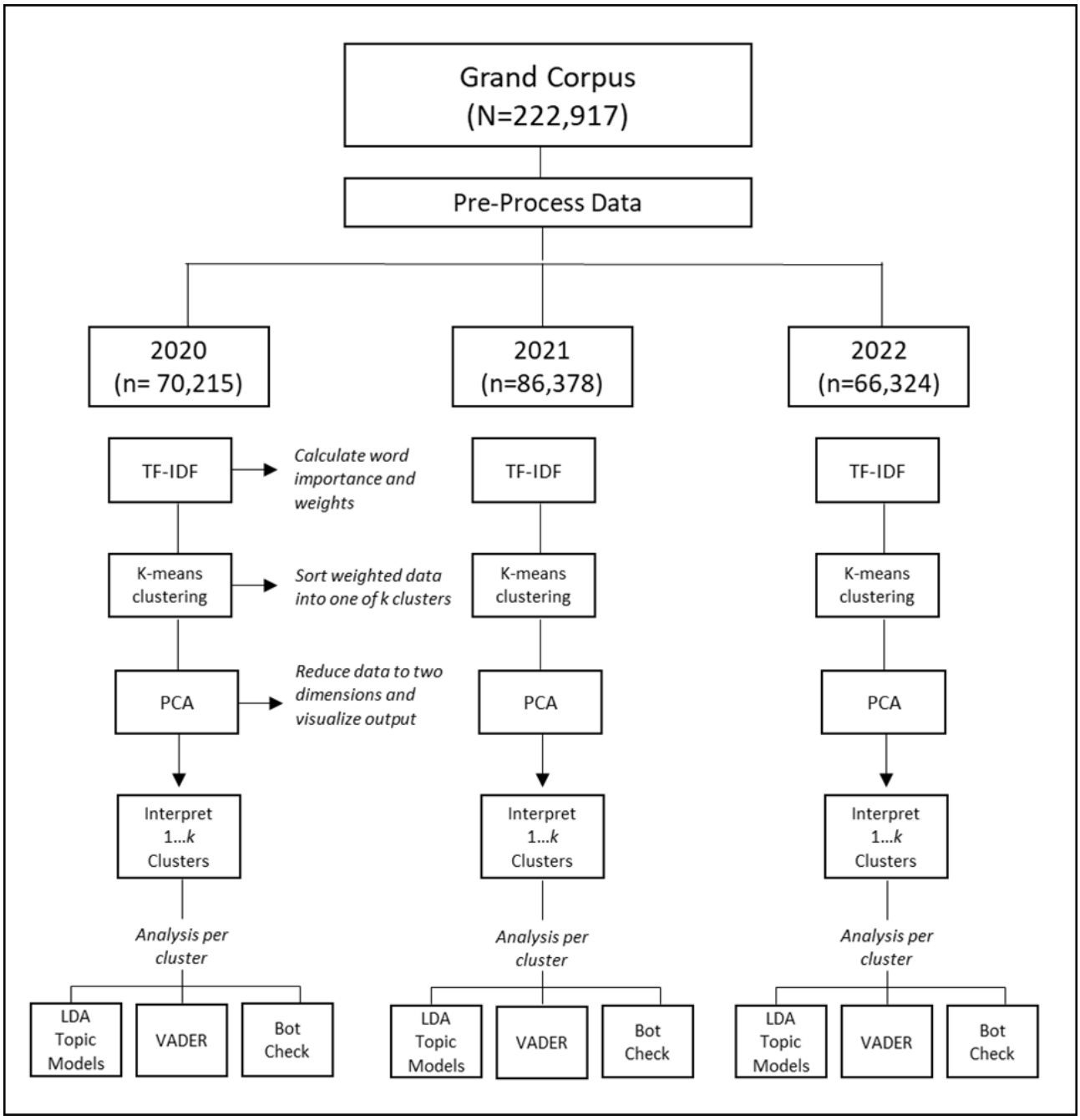
Although NLP methods can analyze language data en masse, a computer cannot ascribe meaning to themes derived from such analyses nor detect certain facets of human speech such as sarcasm [51]. As such, we invoked a simple inductive coding procedure in which 3 authors affiliated with this study independently reviewed approximately 50 posts per cluster per year. Authors were asked to describe the cluster in 3 or 4 words, and upon completion, the authors met to discuss overlap and differences. Key questions asked of the authors were to determine the overall content of each cluster, whether clusters were serious or humorous (ie, sarcasm), and whether the cluster seemed to promote a Dry January–related product. For humorous or sarcastic posts, we specifically looked for indicators, such as the presence of emojis, references to jokes, or exaggerated claims styled for likes. In circumstances in which unanimous consensus could not be reached, we repeated this process with 50 more randomly selected tweets until agreement was met. This process is generally deemed sufficient when dealing with mixed methods topic models on large-scale documents [58], though more research on uniform mixed methods topic modeling guidelines is needed.

Procedure

Our workflow is depicted in Figure 1. To prepare data for analysis, we initiated a series of preprocessing steps, including removing numbers, punctuation, and parts of speech that would detract from the readability of our models, including articles, prepositions, and contractions. Once all data were processed and cleaned, we divided our grand corpus into yearly iterations to afford content comparisons between years (RQ1). We ran a TF-IDF across every year (ie, 2020, 2021, and 2022), then used k-means clustering with elbow scree plots to identify the optimal number of clusters per year. We then applied a PCA to visualize our 2020, 2021, and 2022 data along a vector map. Once data were visualized, we ran interpretation models to afford within-year (or within-cluster) comparisons, including to determine the extent that a natural experiment, such as the COVID-19 pandemic, affected yearly Dry January–related content (RQ2). For example, we used LDA to examine content within each cluster per given year. We used VADER to examine affect per cluster per year. We used the Botometer to determine average bot score per cluster per year. Last, to assess user engagement with Dry January content (RQ3), we took the average number of likes and retweets per cluster and ran correlations with other outcome variables of interest including VADER and Botometer scores.

XSL•FO
RenderX

Figure 1. Study workflow detailing visualization and interpretation analyses per year. LDA: latent Dirichlet allocation; PCA: principal component analysis; TF-IDF: term frequency inverse document frequency; VADER: Valence Aware Dictionary and Sentiment Reasoner.



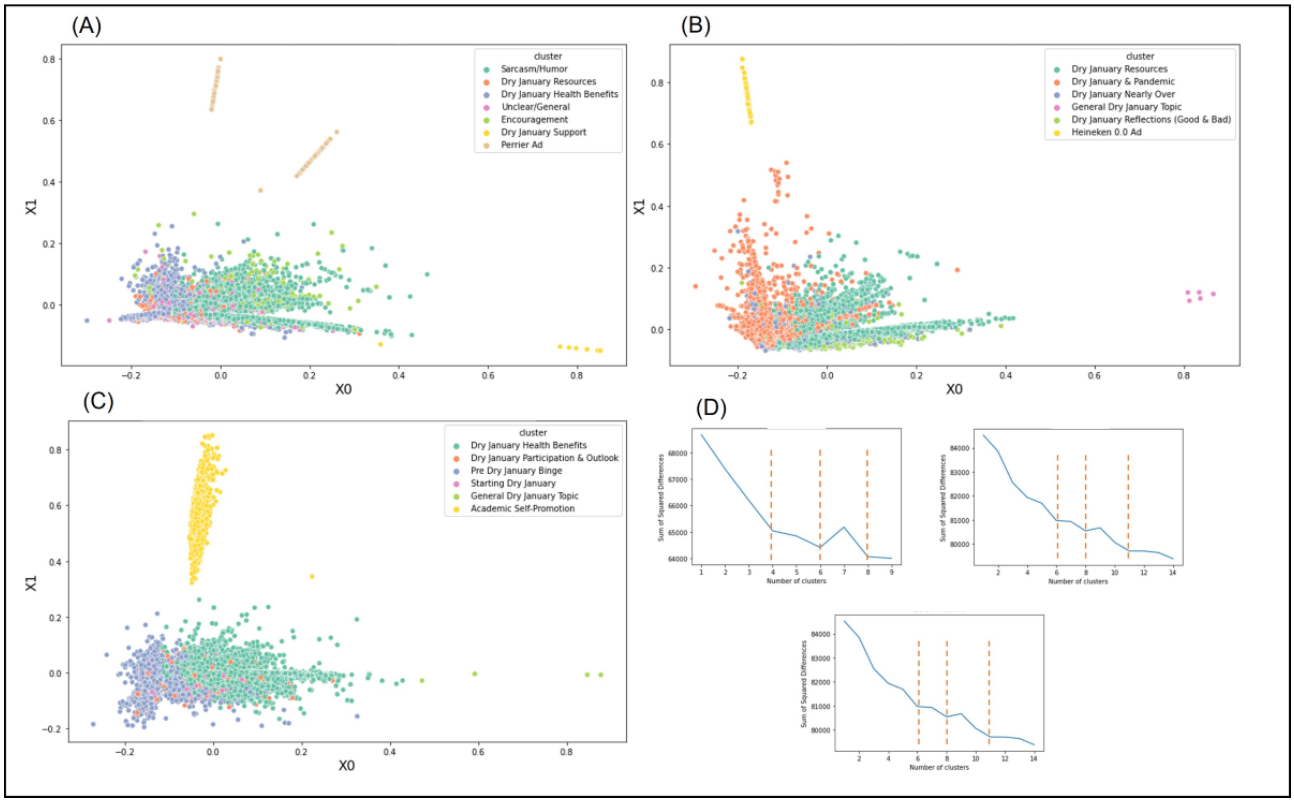
Results

RQ1. What Themes Are Present Within a Corpus of Tweets About Dry January, and Is There Consistency in the Language Used to Discuss Dry January Across Multiple Years of Tweets (2020-2022)?

First, we observed general consistency in topics over time. We used 2 measures to determine consistency of topics: (1) data shape (from the PCA) and (2) overlap in yearly topics (or repeating topics across each year of analysis). Figure 2 provides a visualization of our data per year and model fit summaries; Table 1 similarly provides general information for each year of data collection, topics per year and associated names, the number of tweets per cluster, engagement variables, and other indicators.

XSL•FO
RenderX

Figure 2. Composite figure with principal component analysis (PCA) visualization by year with model fit: (A) 2020 Dry January Twitter dialogue, (B) 2021 Dry January Twitter dialogue, (C) 2022 Dry January Twitter dialogue, (D) elbow method graphs.



XSL•FO
RenderX

Table 1. Content cluster themes and associated summary statistics (n=222,917).

Year and topic	Results, n (%)	VADER[a], mean[b]	Retweets, mean[c]	Likes, mean[c]	Quotes, mean[c]	Botometer score[d]
2020 (n=70,215)						
Sarcasm/humor	38,242 (54.5)	0.16	0.82	9.10	0.12	0.37
DJ[e] health benefits	5804 (8.3)	0.37	1.17	5.39	0.21	0.52
Perrier ad	1320 (1.9)	–0.93	0.00	0.12	0.01	0.88
Unclear/general	1458 (2.1)	0.03	0.32	4.28	0.07	0.37
DJ progress	3372 (4.8)	0.24	0.85	9.04	0.10	0.48
Perrier ad II	1334 (1.9)	0.93	0.00	0.13	0.01	0.88
DJ resources	16,390 (24.1)	0.36	0.77	4.18	0.10	0.44
Support & engagement	1755 (2.5)	0.29	0.50	7.80	0.08	0.39
Entire 2020 data set	N/A[f]	0.18	0.55	5.01	0.01	0.54
2021 (n=86,378)						
DJ nearly over	6190 (7.2)	0.2	0.72	12.39	0.17	0.49
Heineken 0.0. ad	953 (1.1)	0.61	0.007	0.07	0.003	0.9
DJ reflections	56,823 (65.8)	0.14	0.78	13.76	0.14	0.49
DJ resources	17,374 (20.1)	0.35	0.76	8.16	0.18	0.55
DJ & pandemic	3305 (3.8)	0.19	0.455	13.98	0.11	0.47
DJ general topic	1733 (2.0)	0.02	2.8	29.32	0.27	0.44
Entire 2021 data set	N/A	0.25	0.92	12.95	0.15	0.56
2022 (n=66,324)						
Starting DJ	2242 (3.4)	0.24	1.03	16.81	0.27	0.5
Academic self-promotion	1254 (1.9)	0.533	0.02	0.04	0.005	0.82
DJ health benefits	42,894 (64.7)	0.17	0.88	14.03	0.13	0.52
Pre-DJ binge drinking	15,183 (22.9)	0.37	0.7	5.85	0.09	0.67
General DJ topic	1447 (2.2)	0.03	0.4	7.97	0.07	0.52
DJ participation & outlook	3304 (5.0)	0.23	0.79	13.38	0.11	0.49
Entire 2022 data set	N/A	0.26	0.64	9.6	0.11	0.59
Total	N/A	0.23	0.70	9.19	0.09	0.56

[a]VADER: Valence Aware Dictionary and Sentiment Reasoner.

[b]Mean scores were derived from scores ranging from –.99 (high negative affect) to .99 (high positive affect).

[c]A score of 1 indicates 1 retweet, like, or quote.

[d]Botometer scores range from .01 (low bot account likelihood) to .90 (high bot account likelihood).

[e]DJ: Dry January.

[f]Not applicable.

Using a coding procedure outlined in the previous sections, 3 authors affiliated with this study manually named each cluster using a series of representative tweets. Language in representative tweets posted by individual users subsequently included as exemplar tweets was slightly modified to capture original sentiment while preserving anonymity. Per each year, we observed several similar topics that suggest relative consistency in Dry January content over time. These topics include: (1) a general Dry January topic (eg, “Dry January yes, or no?”), (2) Dry January resources (eg, “Have you considered our app to help you maintain your #DryJanuary Goals?”), (3) Dry January health benefits (eg, “Here’s what one alcohol-free month can do for your mind and body”), and (4) updates (positive and negative) related to Dry January progress (eg, “Well, I only lasted a week of Dry January before I drank!”). In 2 of the 3 years included for analysis, we also observed corporate ads targeting Dry January participants, though similar ads were not apparent in 2022.

To support that yearly Dry January content was consistent, we also examined data shape (Figure 2). Indeed, our combined k-means and PCA approach demonstrates relative similarity and dissimilarity of clusters for each year of analysis. Clusters that are proximal contain similar content; clusters that are distal

XSL•FO
RenderX

indicate dissimilar content. Though we acknowledge certain variation across each year, the data shape was relatively similar, which may indicate limited change in content over time. For example, in each year included for analysis, we observed 2 dominant clusters and several smaller clusters dispersed throughout the diagram. Additionally, for each year, we consistently observed at least 2 topics that were far removed and disconnected from the rest of the diagram. Topics, or clusters, that do not overlap with other clusters suggest pockets of conversation that are related to, but not necessarily embedded, within the larger conversation. A secondary explanation for consistent data shape may also be the cohesive theme of the grand corpus or subcorpora (ie, alcohol abstention during the month of January).

RQ2. Do Unique Themes or Patterns Emerge in Dry January 2021 Tweets After the Onset of the COVID-19 Pandemic?

Our findings also indicate that Dry January was affected by emerging news cycles, most notably the COVID-19 pandemic. In the 2020 subcorpora, for example, we did not observe any tweets related to COVID-19, which would not become prevalent in the United States and Europe until March the same year. However, in the following year, we observed 1 cluster containing humorous content about Dry January's cancellation due to the ongoing global pandemic (eg,"Bro, how can we do Dry January during a pandemic?" and "#DryJanuary is officially CANCELLED"). We also observed a small portion of tweets related to the January 6, 2021, US Capitol insurrection, though this content was less prevalent than COVID-19–related tweets. We did not observe a similar cluster related to COVID-19, or similarly disruptive news cycles, during 2022. Yearly news cycle changes may also explain variation in yearly data shape.

RQ3. How Does Tweet Composition (ie, Sentiment and Human-Authored vs Bot-Authored) Affect Engagement With Dry January Tweets?

Tweet composition was associated with engagement, including number of likes, retweets, and quote-tweets per post. We used the Botometer and VADER sentiment analysis to test (1) whether bot-authored and human-authored posts had observed differences in engagement and (2) whether sentiment, which is calculated using the VADER lexicon, similarly affected tweet engagement.

For each year included in our analysis, we observed at least one bot-dominant cluster or an otherwise automated account that posts prewritten content. Per year, bot-dominant clusters were typically comprised of ads, such as Perrier Water and Heineken 0.0 beer, and to a smaller extent, paid or free resources to promote Dry January adherence. Bot-dominant clusters also had fewer likes, retweets, or quote tweets compared with human-authored clusters. Similarly, bot-dominant clusters also had the highest observed positive affect, or greatest amount of positivity per post (eg, "Ready to crush Dry January...with Perrier in your hands you are going to #MakeDryFly!!"). By contrast, human-authored accounts typically had greater engagement and contained lower affect, or greater amount of negativity (eg, "Bro I'm gonna DIE if I have to do another week of Dry January. LOL"). We note that lower affect may reflect sarcasm, though more research on this area is needed.

Discussion

Principal Findings and Implications

Our study characterized online content about Dry January, assessing trends, themes, and general attitudes toward the challenge. We used NLP tools to analyze and visualize a yearly series of tweets related to Dry January over the course of 3 years of participation. Our findings highlight that there is consistency in discussion themes about Dry January across multiple years of tweets, yet we were still able to detect unique themes that emerged in 2021 in response to the COVID-19 global pandemic. Additionally, tweet composition, or whether a tweet was bot-authored or human-authored and the sentiment of the tweet, was associated with user engagement (number of likes, retweets, and quote-tweets).

In the content cluster analysis of the corpus of Dry January tweets, several common themes emerged across multiple years of Dry January participation. For example, the promotion of Dry January resources—such as blogs with tips for help with sustaining Dry January efforts, mobile applications facilitating additional support and accountability, and recipes for nonalcoholic "mocktails"—was a consistent theme each year. Additionally, we observed a cluster associated with Dry January health benefits (eg, drinking reductions, weight loss, healthier dietary choices, reflecting on relationship with alcohol). These findings are consistent with prior work on Dry January that similarly highlighted reductions in alcohol consumption and weight loss as Dry January benefits, in addition to increases in alcohol refusal skills, saving money, improved sleep, increased energy, and enhanced psychological well-being [5,7-9]. Finally, a topic related to sharing about Dry January progress emerged across multiple years of data (eg, no desire to participate in Dry January, intention to participate in Dry January, failed attempts to abstain during Dry January, successful ongoing attempts, successful completion of Dry January). Although some tweets in this cluster referenced successful Dry January experiences and positive associations with these experiences, a large number of these tweets used humor and sarcasm to make light of Dry January participation and voiced an overall lack of desire to participate in the temporary abstinence initiative. This finding is in line with prior work examining alcohol-related content on social media platforms, such as Twitter and TikTok [25,26,59]; the vast majority of alcohol-related posts on these social media platforms portray drinking in a positive manner and often depict hazardous drinking behaviors, such as intoxication and blacking out, in a favorable manner. Similarly, alcohol-related negative consequences are rarely portrayed in alcohol-related social media posts, and when such portrayals are present, they are often depicted in a humorous manner that serves to downplay the severity of alcohol-related problems [25,59].

Content cluster analysis also detected unique themes related to Dry January across years, most notably a cluster of tweets related to Dry January participation in the context of the ongoing COVID-19 global pandemic during January 2021. Many of these tweets referenced individuals experiencing increased

XSL•FO
RenderX

difficulty or a lack of desire to participate in Dry January in the context of the pandemic and social distancing restrictions and increased psychological stressors. Yet, others made reference to having an easier time abstaining during January due to the lack of access to social drinking activities. Humor was commonly used to make light of Dry January in the context of the pandemic. Subthemes within this cluster of tweets were consistent with prior research on alcohol consumption during the peak of the pandemic [60,61]. In addition to millions of COVID-19–related deaths, the COVID-19 pandemic has been associated with increased psychological stressors due to social isolation and higher unemployment rates, among numerous other factors [60,61]. Many have coped with COVID-19 pandemic stressors in the form of self-medication by increasing alcohol consumption [60,61]. Real-time infoveillance of social media posts may prove a valuable means through which to complement health behavior surveillance efforts and to detect public discourse and communication about unique health needs in response to big events, such as coping with the increased psychological stressors associated with the COVID-19 pandemic and how this may negatively impact efforts to quit or cut down on drinking [62].

Finally, we found that tweet composition, most namely whether a tweet was bot-authored versus human-authored affected online engagement with posts. That is to say, bot-dominant clusters (eg, Perrier and Heineken 0.0 promotional efforts) had fewer likes, retweets, and quote-tweets compared to primarily human-authored clusters. This finding has implications for public health messaging and intervention on social media platforms. Although there may be public health benefits from the development and facilitation of social bot-oriented online interventions [63], investigation is warranted into how best to tailor such intervention efforts to enhance engagement, as it appears many individuals in this study largely ignore posts from automated accounts with prewritten content. That said, without knowing the goals or intended outcomes of the bot creators (ie, generating content vs sharing content or raising awareness vs generating engagement), we are unable to determine the effectiveness of social bot presence in Dry January content on Twitter. Our findings do support the presence of social bots and their potential to create, share, and engage with online content.

Limitations

This work is subject to limitations we hope to address in future work. First, although a combined k-means and PCA approach has been extensively validated as an effective way to analyze and visualize abundant social media content, this approach is exploratory and relies on unsupervised algorithms to arrive at findings. As such, there is a possibility that a small proportion of tweets may have been miscategorized by the algorithms. Second, given financial limitations with the Botometer API, we were unable to calculate Botometer scores for all tweets included in the analysis. Instead, we relied on generalizing the Botometer scores from a random subsample of 500 tweets per cluster. It is possible that a full Botometer analysis with the entire sample would alter our findings slightly, particularly for larger clusters comprised of tens of thousands of tweets; however, significant cost barriers associated with the Botometer API prohibited access to a full analysis of tweets. Finally, we also acknowledge that we did not perform a full qualitative analysis with these data. Although we maintain our blinded coding procedure to name clusters was sufficient to determine cluster names, there is also a possibility that a full review of all tweets in a given cluster would yield marginally different cluster names. Through the limitations outlined, we offer several compelling research opportunities to continue this study. For example, a comparative study contrasting our findings from those generated using supervised NLP algorithms, for example the Sentence Bidirectional Encoder from Transformers (S-BERT), could help validate our findings particularly if there is strong overlap across analyses.

Conclusions

We explored themes within and across 3 separate years of Twitter posts about the Dry January temporary alcohol abstinence challenge. Although there was overlap in themes across multiple years of tweets, unique themes related to individuals' experiences with alcohol during the midst of the COVID-19 global pandemic were detected in the corpus of tweets from 2021. Findings underscore the utility for using large-scale social media, such as discussions on Twitter, to study drinking reduction attempts and to monitor the ongoing dynamic needs of persons contemplating, preparing for, or actively pursuing attempts to quit or cut down on their drinking.

Acknowledgments

AMR was supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award number K01AA030614. HCL was supported by the National Institute on Drug Abuse of the National Institutes of Health under award number R01DA049154. PMM was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA229324. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

AMR, DV, SCC, AEB, HCL, and PMM conceptualized and designed the study. AMR, DV, SCC, and BNM contributed to writing the initial draft of the manuscript. DV performed the data analysis for this study with support from SCC. PMM, AEB, and HCL provided mentorship throughout and helped with interpretation of findings and critical reviews of the manuscript. All authors contributed to and have approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. The Dry January story. Alcohol Change UK. URL: https://alcoholchange.org.uk/get-involved/campaigns/dry-january/about-dry-january/the-dry-january-story [accessed 2022-11-13]
2. Why do Dry January? Alcohol Change UK. URL: https://alcoholchange.org.uk/get-involved/campaigns/dry-january/why-do-dry-january-1/why-do-dry-january [accessed 2022-11-13]
3. Take part in Dry January. Alcohol Change UK. URL: https://alcoholchange.org.uk/get-involved/campaigns/dry-january/sign-up-for-dry-january [accessed 2022-11-13]
4. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. Stat Med 2013 Feb 20;32(4):556-577 [FREE Full text] [doi: 10.1002/sim.5408] [Medline: 22711416]
5. de Visser RO, Nicholls J. Temporary abstinence during Dry January: predictors of success; impact on well-being and self-efficacy. Psychol Health 2020 Nov 27;35(11):1293-1305. [doi: 10.1080/08870446.2020.1743840] [Medline: 32216557]
6. Yeomans H. New Year, New You: a qualitative study of Dry January, self-formation and positive regulation. Drugs: Education, Prevention and Policy 2018 Dec 31;26(6):460-468. [doi: 10.1080/09687637.2018.1534944]
7. de Visser RO, Robinson E, Bond R. Voluntary temporary abstinence from alcohol during "Dry January" and subsequent alcohol use. Health Psychol 2016 Mar;35(3):281-289. [doi: 10.1037/hea0000297] [Medline: 26690637]
8. de Visser RO, Robinson E, Smith T, Cass G, Walmsley M. The growth of 'Dry January': promoting participation and the benefits of participation. Eur J Public Health 2017 Oct 01;27(5):929-931. [doi: 10.1093/eurpub/ckx124] [Medline: 28957493]
9. de Visser RO, Piper R. Short- and longer-term benefits of temporary alcohol abstinence during 'Dry January' are not also observed among adult drinkers in the general population: prospective cohort study. Alcohol Alcohol 2020 Jun 25;55(4):433-438. [doi: 10.1093/alcalc/agaa025] [Medline: 32391879]
10. Case P, Angus C, De Vocht F, Holmes J, Michie S, Brown J. Has the increased participation in the national campaign 'Dry January' been associated with cutting down alcohol consumption in England? Drug Alcohol Depend 2021 Oct 01;227:108938 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108938] [Medline: 34392050]
11. Press release: 5 million people plan to do Dry January 2021, up from 3.9 million in 2020. Alcohol Change UK. URL: https://alcoholchange.org.uk/blog/2020/press-release-6-5-million-people-plan-to-do-dry-january-2021-up-from-3-9-million-in-2020 [accessed 2022-11-13]
12. de Ternay J, Leblanc P, Michel P, Benyamina A, Naassila M, Rolland B. One-month alcohol abstinence national campaigns: a scoping review of the harm reduction benefits. Harm Reduct J 2022 Mar 04;19(1):24 [FREE Full text] [doi: 10.1186/s12954-022-00603-x] [Medline: 35246148]
13. Sheffey A, Lalljee J. Sober-curious Millennials and Gen Z are driving Dry January's comeback after a stressful 2021. Insider. 2022 Jan 5. URL: https://www.businessinsider.com/dry-january-sober-curious-taking-break-from-drinking-stress-2021-12 [accessed 2022-11-13]
14. Moquin E. Dry January movement grows in 2022, but for many it's more damp than dry. Morning Consult. 2022 Jan 10. URL: https://morningconsult.com/2022/01/10/dry-january-movement-grows-in-2022/ [accessed 2022-11-13]
15. Asmelash L. How Dry January's continued presence reflects society's evolving – and divisive – relationship with alcohol. CNN. 2022 Jan 19. URL: https://www.cnn.com/2022/01/19/us/dry-january-less-people-drinking-wellness-cec/index.html [accessed 2022-11-13]
16. Furnari C. New Surveys indicate increasing interest in Dry January. Forbes. 2021 Jan 11. URL: https://www.forbes.com/sites/chrisfurnari/2021/01/11/new-surveys-indicate-increasing-interest-in-dry-january/?sh=22f87caf6f57 [accessed 2022-11-13]
17. Miller M, Pettigrew S, Wright CJC. Zero-alcohol beverages: Harm-minimisation tool or gateway drink? Drug Alcohol Rev 2022 Mar 09;41(3):546-549. [doi: 10.1111/dar.13359] [Medline: 34370881]
18. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
19. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
20. Mackey T, Baur C, Eysenbach G. Advancing infodemiology in a digital intensive era. JMIR Infodemiology 2022 Feb 14;2(1):e37115. [doi: 10.2196/37115]
21. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
22. Social media fact sheet. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2022-11-13]
23. Majority of adults look online for health information. Pew Research Center. 2013 Feb 01. URL: https://www.pewresearch.org/fact-tank/2013/02/01/majority-of-adults-look-online-for-health-information/ [accessed 2022-11-13]

XSL•FO
RenderX

24. Shearer E, Mitchell A. News Use Across Social Media Platforms in 2020. Pew Research Center. 2021 Jan 12. URL: https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/ [accessed 2022-11-13]
25. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. "Hey Everyone, I'm Drunk." An evaluation of drinking-related Twitter chatter. J Stud Alcohol Drugs 2015 Jul;76(4):635-643 [FREE Full text] [doi: 10.15288/jsad.2015.76.635] [Medline: 26098041]
26. Riordan BC, Merrill JE, Ward RM. "Can't Wait to Blackout Tonight": An analysis of the motives to drink to blackout expressed on Twitter. Alcohol Clin Exp Res 2019 Aug 10;43(8):1769-1776 [FREE Full text] [doi: 10.1111/acer.14132] [Medline: 31373703]
27. Russell A, Colditz J, Barry A, Davis RE, Shields S, Ortega JM, et al. Analyzing Twitter chatter about tobacco use within intoxication-related contexts of alcohol use: "Can Someone Tell Me Why Nicotine is So Fire When You're Drunk?". Nicotine Tob Res 2022 Jul 13;24(8):1193-1200 [FREE Full text] [doi: 10.1093/ntr/ntab195] [Medline: 34562100]
28. Ward RM, Riordan BC, Merrill JE, Raubenheimer J. Describing the impact of the COVID-19 pandemic on alcohol-induced blackout tweets. Drug Alcohol Rev 2021 Feb 06;40(2):192-195 [FREE Full text] [doi: 10.1111/dar.13186] [Medline: 33020967]
29. Allem J, Dharmapuri L, Leventhal AM, Unger JB, Boley Cruz T. Hookah-related posts to Twitter from 2017 to 2018: thematic analysis. J Med Internet Res 2018 Nov 19;20(11):e11669 [FREE Full text] [doi: 10.2196/11669] [Medline: 30455162]
30. Sidani JE, Colditz JB, Barrett EL, Shensa A, Chu K, James AE, et al. I wake up and hit the JUUL: analyzing Twitter for JUUL nicotine effects and dependence. Drug and Alcohol Dependence 2019 Nov;204:107500. [doi: 10.1016/J.DRUGALCDEP.2019.06.005]
31. Sidani JE, Colditz JB, Barrett EL, Chu K, James AE, Primack BA. JUUL on Twitter: analyzing tweets about use of a new nicotine delivery system. J Sch Health 2020 Feb 11;90(2):135-142 [FREE Full text] [doi: 10.1111/josh.12858] [Medline: 31828791]
32. Unger JB, Rogers C, Barrington-Trimis J, Majmundar A, Sussman S, Allem J, et al. "I'm using cigarettes to quit JUUL": An analysis of Twitter posts about JUUL cessation. Addict Behav Rep 2020 Dec;12:100286 [FREE Full text] [doi: 10.1016/j.abrep.2020.100286] [Medline: 32637562]
33. Allem J, Escobedo P, Dharmapuri L. Cannabis surveillance with Twitter data: emerging topics and social bots. Am J Public Health 2020 Mar;110(3):357-362. [doi: 10.2105/ajph.2019.305461]
34. Kalyanam J, Katsuki T, R G Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav 2017 Feb;65:289-295. [doi: 10.1016/j.addbeh.2016.08.019] [Medline: 27568339]
35. Budenz A, Klassen A, Purtle J, Yom Tov E, Yudell M, Massey P. Mental illness and bipolar disorder on Twitter: implications for stigma and social support. J Ment Health 2020 Apr;29(2):191-199. [doi: 10.1080/09638237.2019.1677878] [Medline: 31694433]
36. 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]
37. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
38. Muric G, Wu Y, Ferrara E. COVID-19 vaccine hesitancy on social media: building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. JMIR Public Health Surveill 2021 Nov 17;7(11):e30642 [FREE Full text] [doi: 10.2196/30642] [Medline: 34653016]
39. Mackey TK, Purushothaman V, Haupt M, Nali MC, Li J. Application of unsupervised machine learning to identify and characterise hydroxychloroquine misinformation on Twitter. The Lancet Digital Health 2021 Feb;3(2):e72-e75. [doi: 10.1016/s2589-7500(20)30318-6]
40. Allem J, Escobedo P, Chu K, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. Tob Control 2017 Mar 08;26(2):226-229 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052757] [Medline: 26956467]
41. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. J Med Internet Res 2014 Oct 16;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]
42. Lazard AJ, Wilcox GB, Tuttle HM, Glowacki EM, Pikowski J. Public reactions to e-cigarette regulations on Twitter: a text mining analysis. Tob Control 2017 Dec;26(e2):e112-e116. [doi: 10.1136/tobaccocontrol-2016-053295] [Medline: 28341768]
43. Merrill JE, Ward RM, Riordan BC. Posting post-blackout: a qualitative examination of the positive and negative valence of tweets posted after "blackout" drinking. J Health Commun 2020 Feb 01;25(2):150-158 [FREE Full text] [doi: 10.1080/10810730.2020.1719242] [Medline: 31986999]
44. Weitzman ER, Magane KM, Chen P, Amiri H, Naimi TS, Wisk LE. Online searching and social media to detect alcohol use risk at population scale. Am J Prev Med 2020 Jan;58(1):79-88. [doi: 10.1016/j.amepre.2019.08.027] [Medline: 31806270]

XSL•FO
RenderX

45. Krumm J, Horvitz E. Eyewitness: identifying local events via space-time signals in twitter feeds. SIGSPATIAL '15: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems 2015 Nov 03;20:1-10 [FREE Full text] [doi: 10.1145/2820783.2820801]
46. Qaiser S, Ali R. Text mining: use of TF-IDF to examine the relevance of words to documents. IJCA 2018 Jul 16;181(1):25-29. [doi: 10.5120/ijca2018917395]
47. Wu HC, Luk RWP, Wong KF, Kwok KL. Interpreting TF-IDF term weights as making relevance decisions. ACM Trans. Inf. Syst 2008 Jun 01;26(3):1-37. [doi: 10.1145/1361684.1361686]
48. Shi N, Liu X, Guan Y. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. 2010 Presented at: Third International Symposium on Intelligent Information Technology and Security Informatics; April 2-4, 2010; Jian, China URL: https://ieeexplore.ieee.org/document/5453745 [doi: 10.1109/iitsi.2010.74]
49. Drikvandi R, Lawal O. Sparse principal component analysis for natural language processing. Ann. Data. Sci 2020 May 18:1. [doi: 10.1007/s40745-020-00277-x]
50. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci 2016 Apr 13;374(2065):20150202 [FREE Full text] [doi: 10.1098/rsta.2015.0202] [Medline: 26953178]
51. Valdez D, Picket A, Young B, Golden S. On mining words: the utility of topic models in health education research and practice. Health Promot Pract 2021 May;22(3):309-312 [FREE Full text] [doi: 10.1177/1524839921999050] [Medline: 33759597]
52. Valdez D, Pickett AC, Goodson P. Topic modeling: latent semantic analysis for the social sciences. Social Science Quarterly 2018 Sep 07;99(5):1665-1679. [doi: 10.1111/ssqu.12528]
53. Bathina KC, Ten Thij M, Valdez D, Rutter LA, Bollen J. Declining well-being during the COVID-19 pandemic reveals US social inequities. PLoS One 2021 Jul 8;16(7):e0254114 [FREE Full text] [doi: 10.1371/journal.pone.0254114] [Medline: 34237087]
54. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In: ICWSM. 2014 May 16 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 216-225 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109 [doi: 10.1609/icwsm.v8i1.14550]
55. Yang K, Varol O, Hui P, Menczer F. Scalable and generalizable social bot detection through data selection. AAAI 2020 Apr 03;34(01):1096-1103. [doi: 10.1609/aaai.v34i01.5460]
56. Luceri L, Badawy A, Deb A, Ferrara E. Red bots do it better: Comparative analysis of social bot partisan behavior. 2019 Presented at: WWW '19: Companion Proceedings of The 2019 World Wide Web Conference; May 13-17, 2019; San Francisco, CA. [doi: 10.1145/3308560.3316735]
57. Yang K, Ferrara E, Menczer F. Botometer 101: social bot practicum for computational social scientists. J Comput Soc Sci 2022 Aug 20:1-18 [FREE Full text] [doi: 10.1007/s42001-022-00177-5] [Medline: 36035522]
58. Eickhoff M, Wieneke R. Understanding topic models in context: a mixed-methods approach to the meaningful analysis of large document collections. Proceedings of the 51st Hawaii International Conference on System Sciences 2018:903-912. [doi: 10.24251/hicss.2018.113]
59. Russell AM, Davis RE, Ortega JM, Colditz JB, Primack B, Barry AE. #Alcohol: portrayals of alcohol in top videos on TikTok. J. Stud. Alcohol Drugs 2021 Sep;82(5):615-622. [doi: 10.15288/jsad.2021.82.615]
60. Nordeck CD, Riehm KE, Smail EJ, Holingue C, Kane JC, Johnson RM, et al. Changes in drinking days among United States adults during the COVID-19 pandemic. Addiction 2022 Feb 12;117(2):331-340 [FREE Full text] [doi: 10.1111/add.15622] [Medline: 34159674]
61. Rodriguez LM, Litt DM, Stewart SH. Drinking to cope with the pandemic: The unique associations of COVID-19-related perceived threat and psychological distress to drinking behaviors in American men and women. Addict Behav 2020 Nov;110:106532 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106532] [Medline: 32652385]
62. Bunting AM, Frank D, Arshonsky J, Bragg MA, Friedman SR, Krawczyk N. Socially-supportive norms and mutual aid of people who use opioids: An analysis of Reddit during the initial COVID-19 pandemic. Drug Alcohol Depend 2021 May 01;222:108672 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108672] [Medline: 33757708]
63. Deb A, Majmundar A, Seo S, Matsui A, Tandon R, Yan S, et al. Social bots for online public health interventions. 2018 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); August 28-31, 2018; Barcelona, Spain. [doi: 10.1109/ASONAM.2018.8508382]

Abbreviations

API: application programming interface
DJ: Dry January
LDA: latent Dirichlet allocation
NLP: natural language processing
PCA: principal component analysis
RQ: research question

XSL•FO
RenderX

S-BERT: Sentence Bidirectional Encoder from Transformers
TF-IDF: term frequency inverse document frequency
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by C Basch; submitted 08.06.22; peer-reviewed by JP Allem, M Field; comments to author 07.10.22; revised version received 13.10.22; accepted 25.10.22; published 18.11.22.

Please cite as:
Russell AM, Valdez D, Chiang SC, Montemayor BN, Barry AE, Lin HC, Massey PM
Using Natural Language Processing to Explore "Dry January" Posts on Twitter: Longitudinal Infodemiology Study
J Med Internet Res 2022;24(11):e40160
URL: https://www.jmir.org/2022/11/e40160
doi:10.2196/40160
PMID:36343184

©Alex M Russell, Danny Valdez, Shawn C Chiang, Ben N Montemayor, Adam E Barry, Hsien-Chang Lin, Philip M Massey. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Opinion Leaders and Structural Hole Spanners Influencing Echo Chambers in Discussions About COVID-19 Vaccines on Social Media in China: Network Analysis

Dandan Wang[1,2,3,4], PhD; Yadong Zhou[5], MSc; Feicheng Ma[1,3,4], Prof Dr

[1]School of Information Management, Wuhan University, Wuhan, China

[2]School of Data Science, City University of Hong Kong, Hong Kong, China (Hong Kong)

[3]Center for Studies of Information Resources, Wuhan University, Wuhan, China

[4]Big Data Institute, Wuhan University, Wuhan, China

[5]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, United States

Corresponding Author:
Feicheng Ma, Prof Dr
School of Information Management
Wuhan University
No.16 Luojia Mountain
Wuchang District
Wuhan, 430072
China
Phone: 86 1 350 711 9710
Email: fchma@whu.edu.cn

Abstract

Background: Social media provide an ideal medium for breeding and reinforcing vaccine hesitancy, especially during public health emergencies. Algorithmic recommendation–based technology along with users' selective exposure and group pressure lead to online echo chambers, causing inefficiency in vaccination promotion. Avoiding or breaking echo chambers largely relies on key users' behavior.

Objective: With the ultimate goal of eliminating the impact of echo chambers related to vaccine hesitancy on social media during public health emergencies, the aim of this study was to develop a framework to quantify the echo chamber effect in users' topic selection and attitude contagion about COVID-19 vaccines or vaccinations; detect online opinion leaders and structural hole spanners based on network attributes; and explore the relationships of their behavior patterns and network locations, as well as the relationships of network locations and impact on topic-based and attitude-based echo chambers.

Methods: We called the Sina Weibo application programming interface to crawl tweets related to the COVID-19 vaccine or vaccination and user information on Weibo, a Chinese social media platform. Adopting social network analysis, we examined the low echo chamber effect based on topics in representational networks of information, according to attitude in communication flow networks of users under different interactive mechanisms (retweeting, commenting). Statistical and visual analyses were used to characterize behavior patterns of key users (opinion leaders, structural hole spanners), and to explore their function in avoiding or breaking topic-based and attitude-based echo chambers.

Results: Users showed a low echo chamber effect in vaccine-related topic selection and attitude interaction. For the former, the homophily was more obvious in retweeting than in commenting, whereas the opposite trend was found for the latter. Speakers, replicators, and monologists tended to be opinion leaders, whereas common users, retweeters, and networkers tended to be structural hole spanners. Both leaders and spanners tended to be "bridgers" to disseminate diverse topics and communicate with users holding cross-cutting attitudes toward COVID-19 vaccines. Moreover, users who tended to echo a single topic could bridge multiple attitudes, while users who focused on diverse topics also tended to serve as bridgers for different attitudes.

Conclusions: This study not only revealed a low echo chamber effect in vaccine hesitancy, but further elucidated the underlying reasons from the perspective of users, offering insights for research about the form, degree, and formation of echo chambers, along with depolarization, social capital, stakeholder theory, user portraits, dissemination pattern of topic, and sentiment. Therefore, this work can help to provide strategies for public health and public opinion managers to cooperate toward avoiding or correcting echo chamber chaos and effectively promoting online vaccine campaigns.

XSL•FO
RenderX

(J Med Internet Res 2022;24(11):e40701) doi:10.2196/40701

KEYWORDS

COVID-19; COVID-19 vaccine; echo chamber; opinion leader; structural hole spanner; topic; sentiment; social media; vaccine hesitancy; public health; vaccination; health promotion; online campaign; social network analysis

Introduction

Background

Despite scientific consensus that COVID-19 vaccines are safe and effective [1], there is still widely circulated controversial information on social media, with statements such as "while vaccinations offer good protection, they do not provide full immunity, and the extent to which they would be effective against new variants of the virus remains uncertain," which damages public confidence [2]. This misinformation leads to vaccine hesitancy, which has been recognized by the World Health Organization as a major global health threat [3]. Social media platforms such as Twitter, Facebook, TikTok, and YouTube provide an ideal medium for spreading and reinforcing antivaccine ideas [4-7]. First, the information-filtering mechanism based on algorithmic recommendation technology mediates and facilitates content promotion by considering users' interest and attitudes [8]. Second, online users have access to a wealth of information and narratives. Affected by individual and social factors such as selective exposure and group pressure, users prefer to select information that fits their belief system, while ignoring dissident information. Gradually, echo chambers emerge, in which like-minded people continue to frame and strengthen shared narratives [9]. In the vaccine promotion campaign, the trend of simplification of users' vaccine-information sources is strengthened and the flow of information between groups with different ideologies toward vaccines is blocked, which widens the knowledge gap and assimilates value cognition [10]. The accompanying group polarization and social fragmentation blind the public to preconceived misconceptions and undermine authorities' efforts to improve the public's information literacy [11], causing inefficiency in the vaccine campaign [12,13].

Users in a social network can be divided into three roles: opinion leaders, structural hole spanners, and ordinary users [14]. Lou and Tang [15] pointed out that the top 1% of users acting as structural hole spanners control almost 80% of information diffusion between communities and 25% of information diffusion on Twitter. Wu et al [16] revealed that 50% of URLs were posted by 1% of users serving as opinion leaders. Further, Cossard et al [17] identified key users in echo chambers, while Jeon et al [18] evaluated the characteristics of users who broke the echo chamber.

To avoid or break an echo chamber, it is critical to characterize these key users and determine their impact on topic dissemination and opinion evolution, which could facilitate the communication within and between pro- and antivaccine groups, and thereby eliminate vaccine hesitancy. Toward this end, in this study, we developed a framework to evaluate and compare the degree of the effect of different forms of echo chambers on users' interactive behavior using quantitative measurements. We further explored the hidden mechanisms of an echo chamber's formation and its strengthening or disintegration by detecting key users who occupy critical network positions, analyzing the relationship between their behavior pattern, network location, and function both inside and outside of echo chambers. Although this framework was designed based on online debates of COVID-19 vaccine hesitancy as the background to offer insights for public health administrators, it could also be applied and expanded to other controversial theme discussions to serve as a reference for public opinion managers.

Prior Work

Echo Chamber of Vaccine Hesitancy

Most studies in this field have concentrated on the presence, form, and degree of echo chambers, whereas limited research has aimed to develop efficient strategies to address the echo chamber effect. Schmidt et al [6]analyzed vaccine-related posts on Facebook from 2010 to 2017, claiming the existence of highly polarized pro- and antivaccine groups by calculating each user's attitude-polarization score based on their "like" and comment behavior. Mønsted and Lehmann [5] obtained similar results from an analysis of tweets posted on Twitter from 2013 to 2016, using the assortativity coefficient derived from network structures. Rathje et al [19] adopted the same index to examine the degree of the echo chamber effect during the COVID-19 epidemic. Apart from attitude-based self-isolation, Del Vicario et al [20] found highly controversial topics by measuring the distance between how a certain topic is presented in tweets and the related users' emotional response. Cossard et al [17] further compared the echo chamber effect on users' interactive behaviors (retweeting, mentioning) on Twitter during measles outbreaks, and identified key users occupying a central location in interactive networks to tighten the structures of echo chambers. To mitigate the negative effect of echo chambers, Jeon et al [18] performed a user experiment using a game-based methodology to determine the characteristics of users who broke the echo chamber. The breakers were consistently aware of being trapped in echo chambers and tended to maintain diverse perspectives when consuming information.

User Roles in Echo Chambers

Social capital, as a set of resources embedded in relationships, results from holding certain locations in a social structure [21]. Social capital theory suggests that a more central location in a social network, with cohesive social ties fostering trust and cooperation, leads to more bonding relationships. By contrast, structural hole theory emphasizes that social capital results from a bridging position, which can bring the ego diverse and nonredundant information [22,23], as well as control of information flow [24] so as to enhance innovation performance [25]. The idea is grounded in weak tie theory [26]. Weak ties represent loose connections in the network, making it easier to

XSL•FO
RenderX

include a large number of talents with different views, information, and resources [27].

Burt [28] explained that whether social capital performs a greater function of bonding than bridging depends on the context. An "opinion leader" is a term used to broadly refer to any individual or entity with high influence in a network, and should not be predetermined but rather explored in different contexts [29]. Opinion leaders occupy the center of the information network within their local communities, and can influence others by drawing their attention to certain topics or opinions and inspiring reactions to the messages they post [30-32]. Opinion leaders have been found to be responsible for promoting an echo chamber [33]. Through an online-search experiment, Bar-Gill and Gandal [34] found that opinion leaders raised the potential for a topic echo chamber, promoting communities to focus on homogeneous topics. Guo et al [29] analyzed the impact of opinion leaders of different genders, partisanship, and stakeholder categories on political homophily in Twitter communities. However, Dubois and Blank [35] and Dubois et al [36] drew conclusions from survey data that the contribution of opinion leaders to a political echo chamber was overvalued without considering the interests of information receivers and the diversity of information sources. Based on thorough qualitative interviews, Bergström and Jervelycke Belfrage [37] also argued that opinion leaders brought attention to news others would have missed.

The lack of connection among communities forms structural holes in social structures [38]. Individuals filling the holes, acting as intermediaries between different communities, are regarded as "structural hole spanners" [15,39]. By simulating opinion update rules of ordinary agents and structural hole agents, Gong et al [40] proved that structural hole–based approaches could alleviate the echo chamber effect and reduce opinion polarity in social networks. Using social network analysis, Swarnakar et al [41] emphasized that structural hole spanners acted as brokers and bridge-makers for collaboration of heterogeneous patterns on climate change.

Research about opinion leaders' impact on echo chambers has resulted in contradictory conclusions with respect to different social issues. Limited research has focused on the impact of structural hole spanners on echo chambers. Rather, research in this field has mainly focused on opinion-based echo chambers under specific topics, while ignoring users' topic selection prior to opinion contagion. Despite these advancements, a gap remains in the literature: if both bonding and bridging arguments are valid depending on the context, under which conditions should they be complementary or otherwise?

Research Questions

Online echo chambers have been studied in the context of users' interactions (eg, posting, retweeting, commenting, mentioning), focusing on rumor spread and management [42-44], political debates [29,45], and news consumption [46,47]; however, related studies on vaccine hesitancy are rare, especially during public health emergencies. To assess whether users on Sina Weibo, the most popular microblogging platform in China (with a structure similar to Twitter), exhibited an echo chamber effect when discussing COVID-19 vaccines and vaccinations, and to further understand the formation mechanism or to design strategies to break it, we sought to identify the key users, how their behavior patterns relate to their online positions, and how they cooperate or compete to promote (prevent) the formation or strengthening (breaking) of the echo chamber. Toward this end, we established the following research questions (RQs):

RQ1: Is there an echo chamber effect in topic selection and opinion contagion of users on Weibo when discussing COVID-19 vaccines and vaccinations? Does it differ between users' retweeting and commenting behaviors?

RQ2: Do users with different behavior patterns on Weibo tend to be regarded as opinion leaders or structural hole spanners?

RQ3a: Do online opinion leaders and structural hole spanners tend to act as echoers or bridgers in topic dissemination?

RQ3b: Do online opinion leaders and structural hole spanners tend to act as echoers or bridgers in attitude interaction?

RQ4a: Do these key users acting as echoers in topic dissemination tend to play the same role in attitude interaction?

RQ4b**:** Do these key users acting as bridgers in topic dissemination tend to play the same role in attitude interaction?

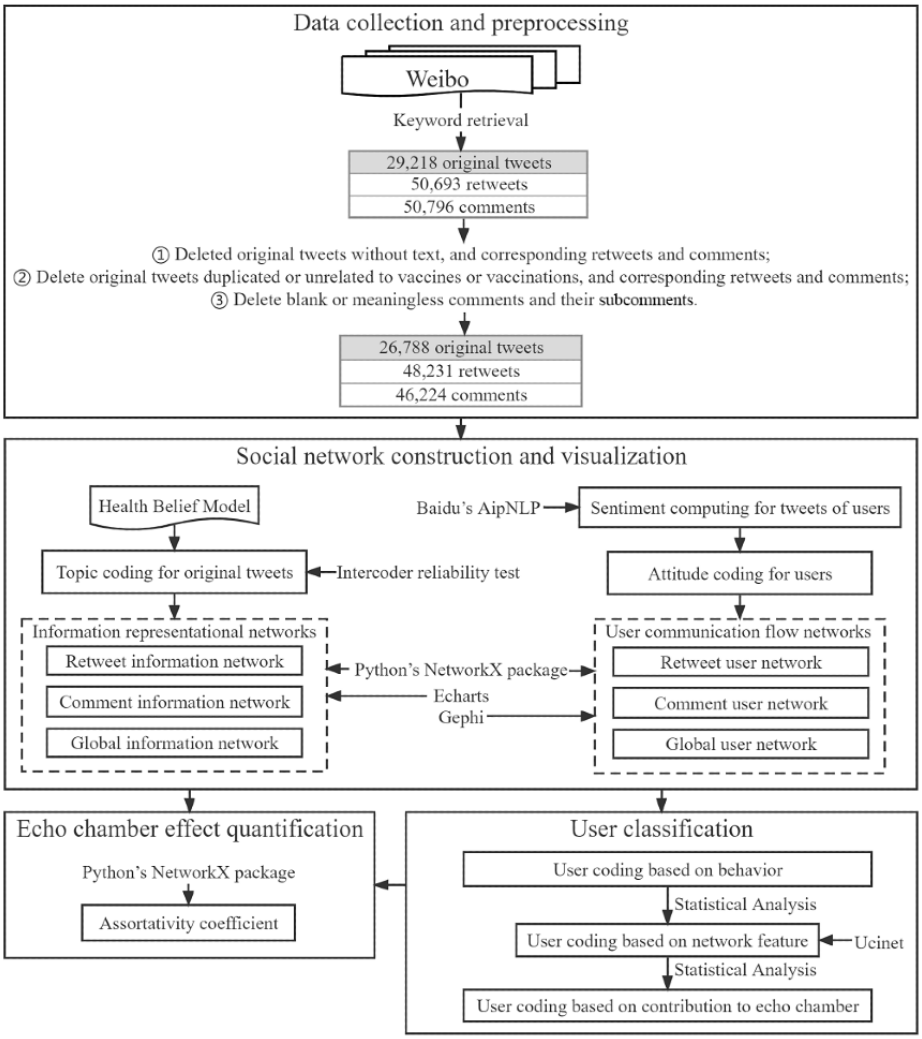
Methods

Design and Definitions

Figure 1 outlines the research framework. Note that although Weibo posts were analyzed in this study, we use the terms "tweet" and "retweet" throughout the manuscript to refer to activities on the platform, equivalent to activities on Twitter, for the sake of convenience. An original "tweet" refers to posts created by a registered Weibo user. A "retweet" refers to users' forwarding behavior on Weibo. "Comments" refer to replies to an original post on Weibo.

XSL•FO
RenderX

Figure 1. Research framework.



Ethical Considerations

Our research did not require ethical board approval because it did not involve human or animal trials. The research data were derived from open access data available on social media, mainly through voluntary contributions from users. Our data and analysis of data were conducted in an unbiased and transparent manner, and the data were used only for scientific research without any ethical violations. To be specific, we anonymized key identifiable information, including the nickname field provided by each user and the ID number assigned to each user by the platform when they registered their unique account. We represented these two fields as nonrepeating consecutive integers incremented from 1 to uniquely identify each user, thus hiding the users' personal information, which had no influence on the study results.

Data Collection and Preprocessing

From January 23, 2020, to February 11, 2021, there were numerous messages posted about the outbreak and cessation of the COVID-19 epidemic, as well as the initial exploration of vaccine development and vaccination on Weibo [48]. As a medical innovation, the vaccine was widely debated in its early diffusion stage [49]. We first used the Sina Weibo application programming interface to collect original tweets containing keywords ("COVID-19 vaccine [新冠疫苗]" or "COVID-19 vaccination [新冠疫苗接种]"). Considering that the interactive data (ie, retweet, comment, and like) of an original tweet could become stable approximately 1 week after it was posted [50,51], we crawled the following-week interactive data for each original tweet, involving likes, retweets, and comments, and the information of posters, retweeting, and commenting users. There were initially 29,218 original tweets, 50,693 retweets, and 50,796 comments.

For data preprocessing, we deleted the original tweets without any text (eg, only pictures, videos, or audio) or those that were duplicated or contained the above keywords but did not include any meaningful content. In addition, blank or meaningless comments and their subcomments were also eliminated. After excluding the corresponding retweets and comments as well as the user information, there were 26,788 original tweets, 48,231 retweets, and 46,224 comments from 77,625 users retained for analysis.

XSL•FO
RenderX

Social Network Construction and Visualization

Interactive Network Design

To answer RQ1-4, we constructed interactive networks. Information representational and user communication flow networks are commonly used as the basis to measure polarization [20,43,44].

Information Representational Network Construction

First, we marked the topic for each original tweet. To cover all aspects of the vaccine, we performed this process based on the Health Belief Model, which indicates that the perceived susceptibility, perceived severity, perceived benefits, perceived barriers, cues to action, and self-efficacy have impacts on individuals' motivation to carry out preventive health behaviors [52,53]. We invited two experienced researchers to label the topics for 10% of the original tweets, and the result passed intercoder reliability tests [54] (κ=0.967). After repeating the review and eliminating disagreements, the topic-coding scheme was developed (we did not consider the construct of self-efficacy owing to its low prevalence in the data set), which is shown in Table 1. This scheme was used to label the remaining original tweets.

Retweeting or commenting on an original tweet indicates that the users are interested in the tweet's topic [20,44]. Based on interactive data, we next established representational networks of information, which were undirected and weighted. In a global information network, each node represents an original tweet; if a user retweets or comments on original tweet *i* and original tweet *j*, an edge exists between *i* and *j*. The edge's weight represents the number of common users who participate in discussion on the two original tweets. The retweet/comment information network only contained retweet/comment relationships. We then used Python's NetworkX package to construct these three networks [55] and calculated their detailed topological attributes. Finally, the chord diagram visualization of Echarts [56] was used to visualize the degree of homophily based on topics in the networks.

Table 1. Topic-coding scheme based on the Health Belief Model.

Construct	Topics
Perceived susceptibility	Risk of getting COVID-19 infection
Perceived severity	Severity of getting COVID-19 infection or refusing COVID-19 vaccination
Perceived benefits	Effectiveness of COVID-19 vaccination
Perceived barriers	Adverse effects of COVID-19 vaccination; cost of COVID-19 vaccination; fake (eg, counterfeit) vaccines, fraudulent information; safety (eg novelty), infectivity of vaccines, and standardization of vaccination process; conspiracy theory
Cues to action	Means to get vaccination; dos and don'ts of vaccination; domestic vaccine development, production, and vaccination; foreign vaccine development, production, and vaccination; personal vaccination experience

User Communication Flow Network Construction

Many studies adopted the sentiment expressed in tweets created/retweeted/commented by users to represent their attitudes toward vaccines [5,20,57]. Considering the Chinese context of Weibo [58], we used Baidu's AipNLP [59] to calculate the sentimental positive probability ($0 \le \alpha \le 1$) of each original tweet, retweet, and comment. If $0 \le \alpha \le 0.5$, the text was regarded as negative; if $0 < \alpha \le 1$, the text was regard as positive; and otherwise, it was regard as neutral. We counted the most frequently expressed sentiment type of the user, which represented their attitude toward vaccines. Next, we established communication flow networks of users, which were directed and weighted. In the global user network, each node represents a user; if user *i* retweets or comments on tweets (including original tweets, retweets, and comments) of user *j*, there is an edge from user *i* to *j*. The edge's weight represents the number of interactions between the two users. The retweet/comment user network only contained retweet/comment relationships. We then used Python's NetworkX package to construct these three networks [55] and calculated their detailed topological attributes. Finally, Gephi was used to visualize the degree of homophily based on users' attitudes, and the Fruchterman Reingold layout algorithm was used to visualize the connectivity in user networks [60].

Echo Chamber Effect Quantification

To answer RQ1, we used Python's NetworkX package to calculate each network's assortativity coefficient r ($-1 \le r \le 1$) based on the nodes' attributes (topic in information networks, attitude in user networks) and their interaction, which measures the network's homophily [5,44]. An $r>0$ indicates that the node generally tends to connect with other nodes with similar properties, and the network is referred to as an assortative network. A larger r value indicates more prominent assortativity. If $r \le 0$, assortativity does not hold [61].

User Classification

Method of Classification

To answer RQ2, we characterized online users' behavior patterns and detected opinion leaders and structural hole spanners based on their network locations. To answer RQ3-4, we defined two types of mediators to represent the above key users' contributions to echo chambers. After coding users from these three perspectives, statistical tests were used to examine the relationships.

User Coding Based on Behavior

Villodre and Criado [62] classified users based on their contrasting behaviors during the dissemination of crisis information. Based on a modification of their rules, we classified all users into 8 categories, as shown in Table 2. We then

XSL•FO
RenderX

analyzed stakeholders for each category by matching keywords in each user's personal authentication, introduction, and tags. Referring to the identity-keyword list from An and Ou [63], after manually marking 10% of all users (intercoder reliability of two coders, κ=0.991), we modified and expanded the list, and finally determined 11 categories, as shown in Table 3. The remaining users were automatically coded using the new list.

Table 2. User behavior taxonomy.

User category	Criterion	Behavior description
Influential		
Speaker	Number of retweets received was three times higher (low speakers), 10 times higher (medium speakers), or 100 times higher (high speakers) than that of tweets they had posted	Users create widely shared content. They show less content-sharing behavior
Networker	Number of tweets≥total mean; number of retweets received≥total mean; number of retweets received/number of retweets sent≥0.5	Users show equilibrium between creating content, sharing content, and being retransmitted
Broadcaster		
Monologist	Number of tweets≥total mean; number of retweets received/own tweets≤0.3	Users create original content that is not widely shared
Retweeter	Number of tweets≥total mean; number of retweets sent/own tweets≥0.5	Users mostly share others' content
Replicator	Number of comments sent/own tweets≥0.6	Users mostly comment on others' content
Isolator	Number of retweets sent=0; number of retweets received=0; number of comments sent=0; number of comments received=0	Users never share/comment on others' content and they create some content that is never shared/commented by others
Automatic	Send same comments multiple times under one tweet; personal information is blank	Users seem to act with automatization
Common user	None of the above	Not applicable

Table 3. Stakeholder types and related keywords.

Stakeholder types	Keywords (partial)
Government	government, police, court, judicial bureau, judicial office, procuratorate, commission for discipline inspection, political and legal committee
Hospital	hospital
Traditional media	newspaper, radio, TV station, news, magazine, broadcast, daily, timely, weekly, monthly, morning post, evening post, channel
We-media	We-media, author, writer, reporter, editor, blogger, commentator, critic
Platform account	Sina Weibo, Weibo medical and health operation, Weibo secretary, Weibo administrator, Weibo rumor rebuttal, Weibo politics
Social organization	association, public welfare
Medical company	vaccine manufacturer ("SINOVAC BIOTECH CO., LTD". ["科兴"], "CanSino Biologics Inc." ["康希诺"], "Hualan," "Zhifei," "Kangtai"), medical enterprise
Common company	company, enterprise
Educational institution	middle school, high school, campus, technical school
Medical personnel	doctor, nurse
Common personnel	None of the above

User Coding Based on Network Features

To measure the extent of each user being regarded as an opinion leader, we adopted in-degree centrality [29], which represents the volume of network ties directed toward a user [64], and the local clustering coefficient, which quantifies the degree to which the user's neighbors aggregate with each other to form a clique (complete graph) [65]. Burt [38] proposed four metrics to describe structural hole spanners: effective size, efficiency, constraint, and hierarchy, the third of which is the most important. The effective size of a node measures the nonredundant connections of a node. Efficiency is the effective size divided by the number of the node's neighbors. Yang et al [25] and Tan et al [66] chose "constraint" (between 0 and 1),

XSL•FO
RenderX

which measures the extent to which the node's contacts are redundant. When the constraint is closer to 0, there are fewer connections between the node's contacts. Hierarchy measures the extent to which the aggregate constraint on the node is concentrated in a single contact. A hierarchy value closer to 0 indicates that the constraint is the same for the node's relationship with each neighbor, whereas a value closer to 1 indicates that all constraints are concentrated in a single contact. The spanner tends to have higher values of effective size, efficiency, and hierarchy, and lower values of constraint [67]. We used Ucinet [68] to compute the above indices for each node in the global user network.

User Coding Based on Contribution to the Echo Chamber

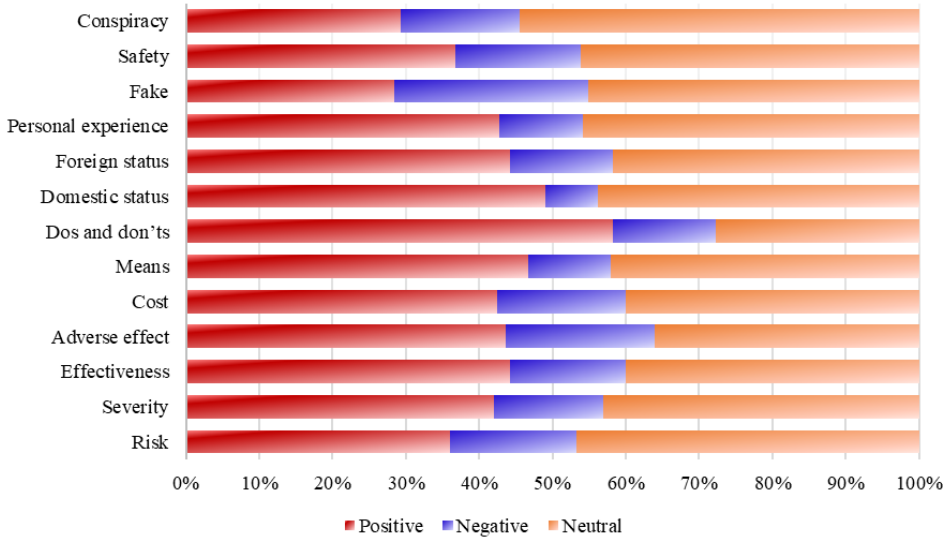
To uncover the mechanisms of intra- and intergroup communication among holders of different interests and viewpoints, we conceptualized two types of social mediators. One was the "echoer," who only initiated interactions with peers whose interests and viewpoints were highly homogenous, thereby contributing to the formation and even consolidation of echo chambers [45]. The other was the "bridger," who tended to initiate intergroup dialog across areas of interest and heterogeneous viewpoints, aiming to break down echo chamber barriers [45]. To be specific, for a topic-based echo chamber, if a user only created, retweeted, or commented on tweets of the same topic, the user was considered to be an echoer; otherwise, they were considered to be a bridger. For an attitude-based echo chamber, if a user only created, retweeted, or commented on tweets from users who had the same attitude, the user was classified as an echoer; otherwise, they were classified as a bridger.

Results

Descriptive Statistics

Tweets about domestic, foreign status, and conspiracy accounted for 24.46% (n=29,653), 20.40% (n=24,734), and 16.46% (n=19,955) of total tweets (N=121,243), respectively. Overall, 42.51% (51,544/121,243) of tweets expressed a positive attitude toward vaccines and 13.12% (15,907/121,243) of tweets held a negative attitude. Figure 2 shows that discussions about domestic status were the least controversial, whereas discussions related to counterfeit vaccines and fraudulent information were the most divisive.

Figure 2. The distribution of attitudes expressed in tweets (original tweets, retweets, comments) about different topics.



Echo Chamber Effect in Networks

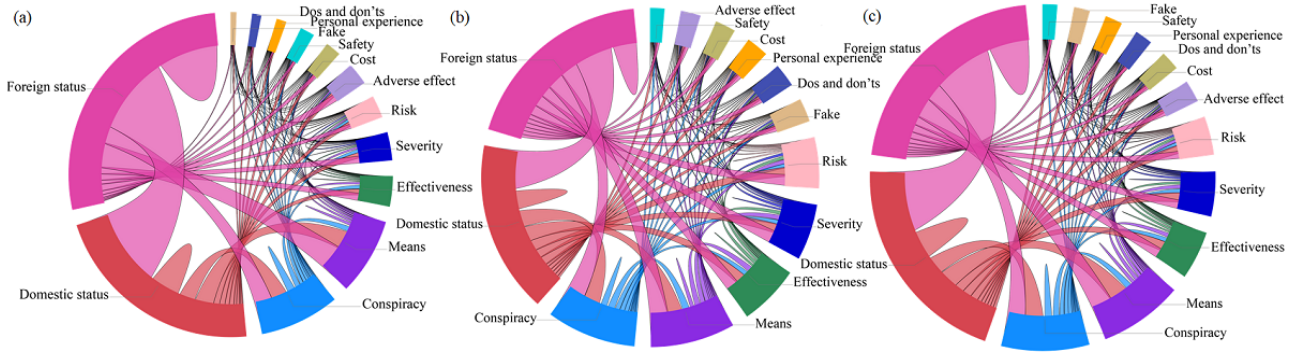
Retweet, comment, and global information networks were all sparse, with a density of 0.003, 0.003, and 0.0002, respectively. In Figure 3, the outer ring of 13 different colors represents a collection of 13 different topics of original tweets, the arc length represents the total connection volume for all of the original tweets belonging to this topic, and the inner colored connecting bands indicate the flow direction and magnitude of the data relationship. The top four topics that interacted most frequently with others were "Foreign status," "Domestic status," "Conspiracy," and "Means." "Foreign status" was often retweeted by users with topics such as "Domestic status," "Conspiracy," and "Means" at the same time, along with "Effectiveness, "Severity," and "Risk." The assortativity coefficients of retweet, comment, and global information networks were 0.060, 0.022, and 0.048, respectively, indicating low topic-based homogeneity and that the retweet information network displayed more obvious homogeneity compared with the comment information network.

Retweet, comment, and global user networks were also sparse, with densities lower than those of information networks. Compared with those of the retweet user network (0.003, 0.0003, 0.011), the comment user network had a higher clustering coefficient, transitivity, and reciprocity (0.007, 0.055, 0.025), indicating that the network built on comment relationships was more cohesive and stable, where users were closely connected and relatively stable [69], while retweeting was mostly used for a one-way flow of information [70]. As shown in Figure 4, in the three user networks, clusters brought together people who were confident about vaccines and people with uncertainty [71].

XSL•FO
RenderX

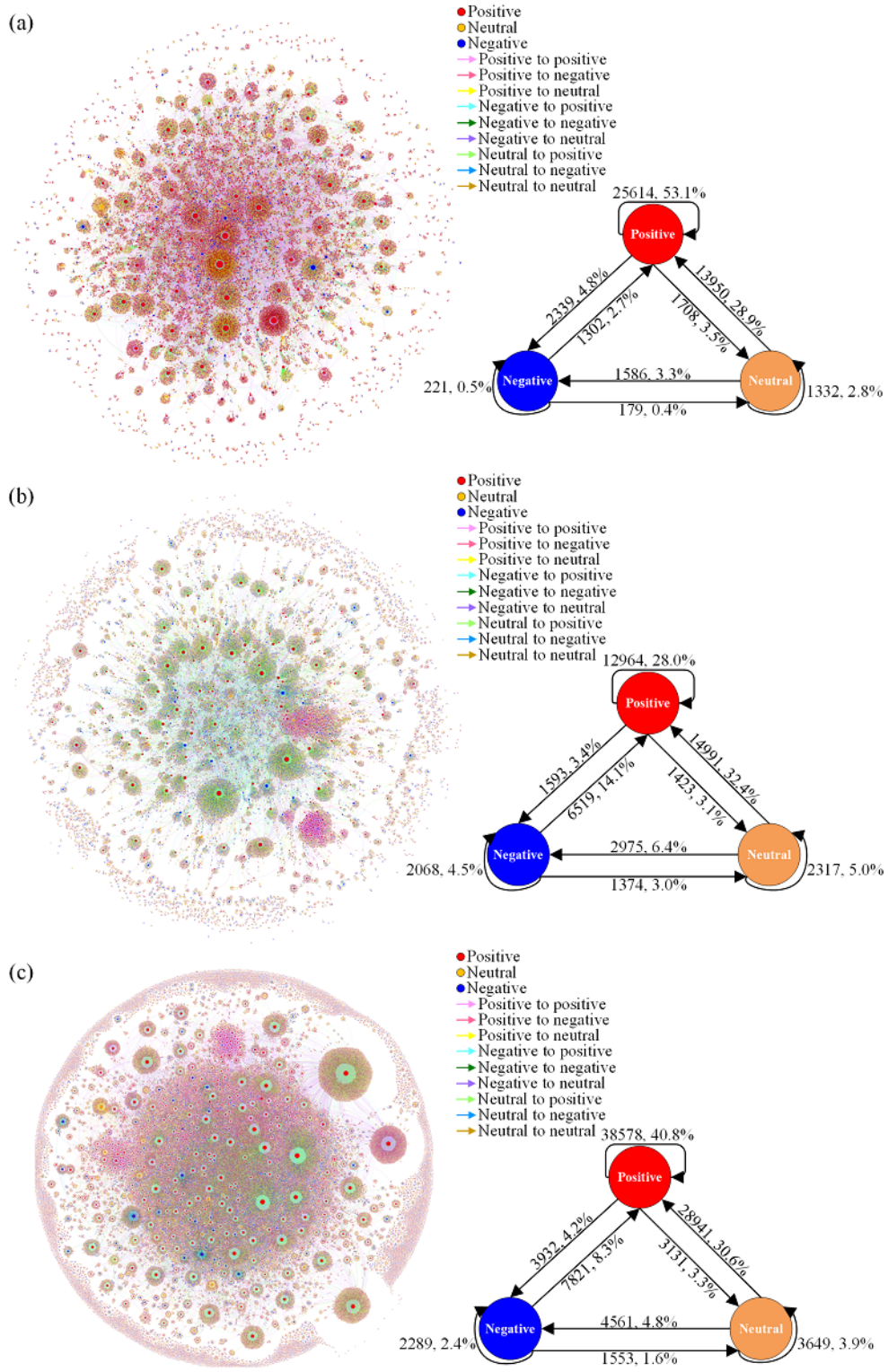
The more common edges were found between users holding a positive attitude and between users with a neutral attitude to users with a positive attitude. Users with a clear attitude hardly retweeted posts from users without a determined attitude. Compared with the retweet user network, the tendency of users with a negative/neutral attitude to comment on posts of other users with the same attitude was more obvious, whereas this tendency was less obvious for users with a positive attitude. The assortativity coefficients of the retweet, comment, and global user networks were 0.031, 0.042, and 0.055, respectively, indicating low attitude-based homogeneity and that the comment user network displayed more obvious homogeneity compared with the retweet user network.

Figure 3. Chord diagram representation of the retweet information network (a), comment information network (b), and global information network (c) colored by topic.



XSL•FO
RenderX

Figure 4. Communication flow network of users in the (a) retweet user network (b) comment user network, and (c) global user network. The size and color of each node represent its in-degree and user's attitude (red="positive", blue="negative", orange="neutral"), respectively. The color of the edge is explained in the corresponding legends in the figure.



Relationships Between User Behavior, Network Position, and Role in the Echo Chamber

As shown in Figure 5, most users were coded as "common user," sending 86.9% of retweets and 93.9% of comments. Only 1.0% of users were speakers, but receiving 81.3% of retweets and 62.8% of comments. Most of the original tweets were created by isolators. Retweeters not only often created tweets but also frequently retweeted others' tweets.

As shown in Figure 6, with respect to the weighted degree, most of the speakers' weighted in-degree centrality was much higher than their weighted out-degree centrality. A different situation was found for retweeters. Speakers had a relatively higher average clustering coefficient (high speakers: 0.012, medium

XSL•FO
RenderX

speakers: 0.013, low speakers: 0.007), whereas the average clustering coefficient of retweeters (0.0002) was the lowest among users (apart from isolators). Compared with that of retweeters, replicators had a higher average clustering coefficient (0.004), with in-degrees and out-degrees of similar size.

Figure 7 and Figure 8 show that influentials had more obvious structural hole properties than broadcasters. Among influentials, speakers had a higher effective size, efficiency, and lower constraint than networkers. High speakers performed in the same manner but with a much greater effect. However, some networkers had higher hierarchy than speakers, which indicated that a networker's constraint was more concentrated on this actor and was more important. Among broadcasters, half of the monologists' constraint values were lower than 0.500 and half of them had hierarchy values lower than 0.092. Although most of the replicators' constraint values were lower than 0.333, they largely showed hierarchy values lower than 0.278. Compared with that of replicators, retweeters' constraint was more concentrated in a single contact.

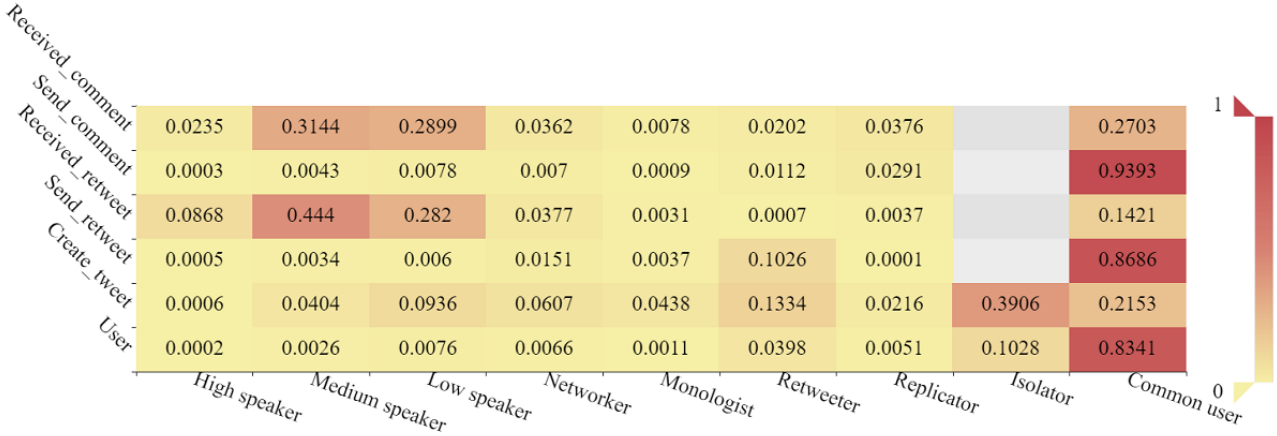
Given the massive network size, we considered the top 5% of users in weighted in-degree centrality and local clustering coefficient as opinion leaders (n=386, 0.5% of all users), and the other users in the bottom 5% in constraint were considered as structural hole spanners (n=3123, 4.0% of all users). These two types of users were considered key users. As shown in Figure 9, opinion leaders were responsible for 38.4% of all of the information flow, while structural hole spanners were responsible for 50.2% of the information flow. Compared with the former, the latter tended to create, retweet, and comment on more tweets.

As shown in Figure 10, common users, speakers, replicators, and networkers accounted for 44.0%, 36.3%, 10.6%, and 4.1% of opinion leaders, respectively. Common users, speakers, networkers, and retweeters accounted for 59.3%, 19.7%, 8.7%, and 8.7% of structural hole spanners, respectively. The χ^2 tests showed a significant difference in the distribution of categories of users between opinion leaders and structural hole spanners (χ^2_7=184.650, P<.001). Posthoc testing further showed that speakers, replicators, and monologists tended to be opinion leaders, whereas common users, retweeters, and networkers tended to be structural hole spanners.

Isolators did not become opinion leaders or structural hole spanners, whereas 89.2% of isolaters were topic-based echoers and all of them were attitude-based echoers. The results of χ^2 tests showed that the proportion of structural hole spanners acting as topic-based bridgers (74.2%) was significantly higher than that of opinion leaders (64.2%) (χ^2_1=17.148, P<.001). The opposite result (χ^2_1=13.193, P<.001) was found when considering attitude-based bridgers (structural hole spanners: 88.1%; opinion leaders: 94.3%). Hence, compared with being echoers, both opinion leaders and structural hole spanners tended to act as bridgers. Structural hole spanners were more likely to become bridgers than opinion leaders in topic-based echo chambers, whereas structural hole spanners were less likely to become bridgers than opinion leaders in attitude-based echo chambers.

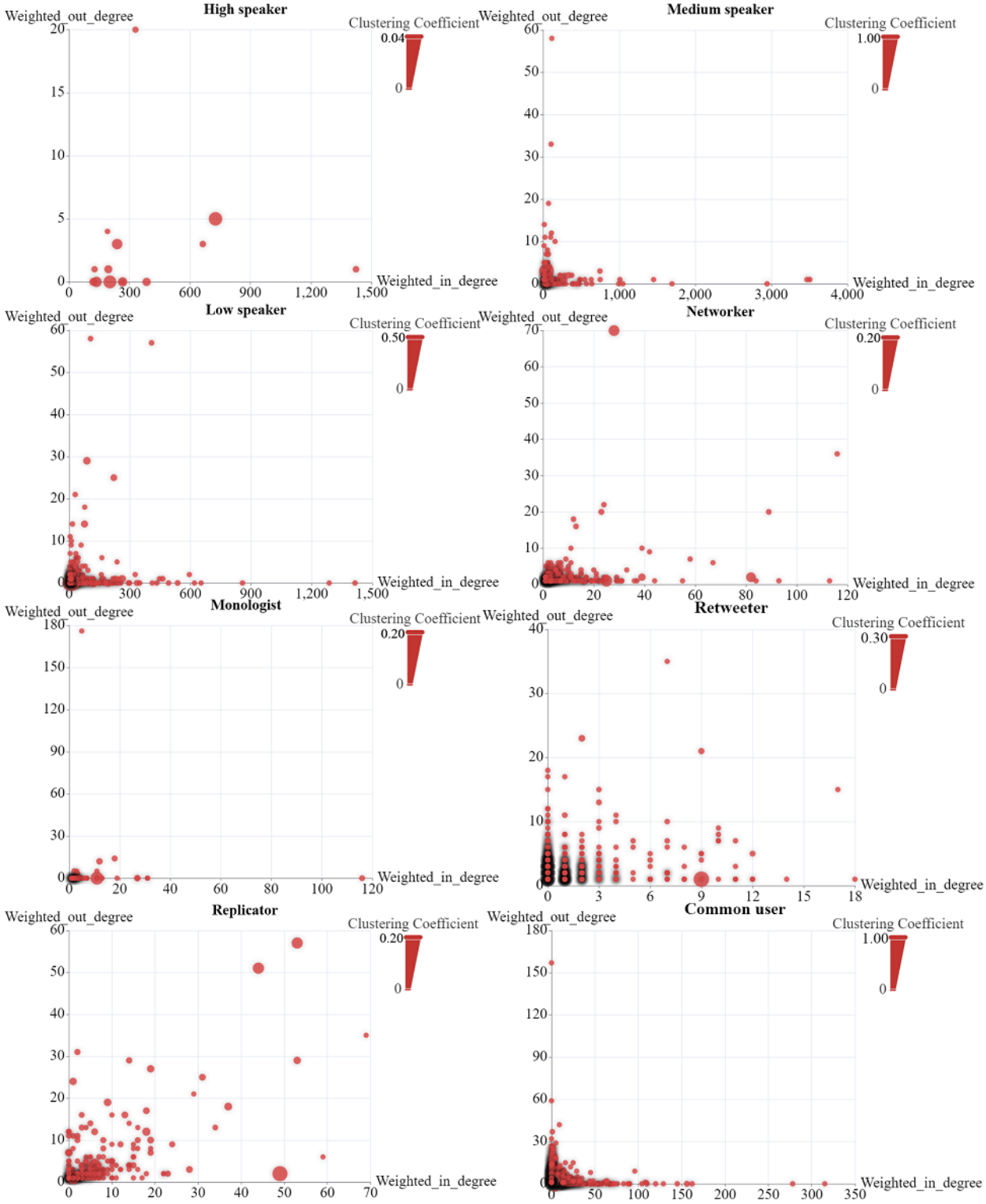
To address RQ4, the support and the confidence of the rule were calculated. As shown in Table 4, RQ4a was declined, whereas RQ4b was supported, with 62.8% of all key users acting as both topic-based and attitude-based bridgers. Approximately 85.9% of topic-based bridgers also acted as attitude-based bridgers. Specifically, 60.6% of opinion leaders (32.9% government accounts, 30.3% We-media, 19.2% traditional media) and 63.0% of structural hole spanners (39.4% common personnel, 28.4% We-media, 17.3% traditional media) acted as both topic-based and attitude-based bridgers.

Figure 5. Percentage of user categories based on their behavior. No automatics were detected in the data set.



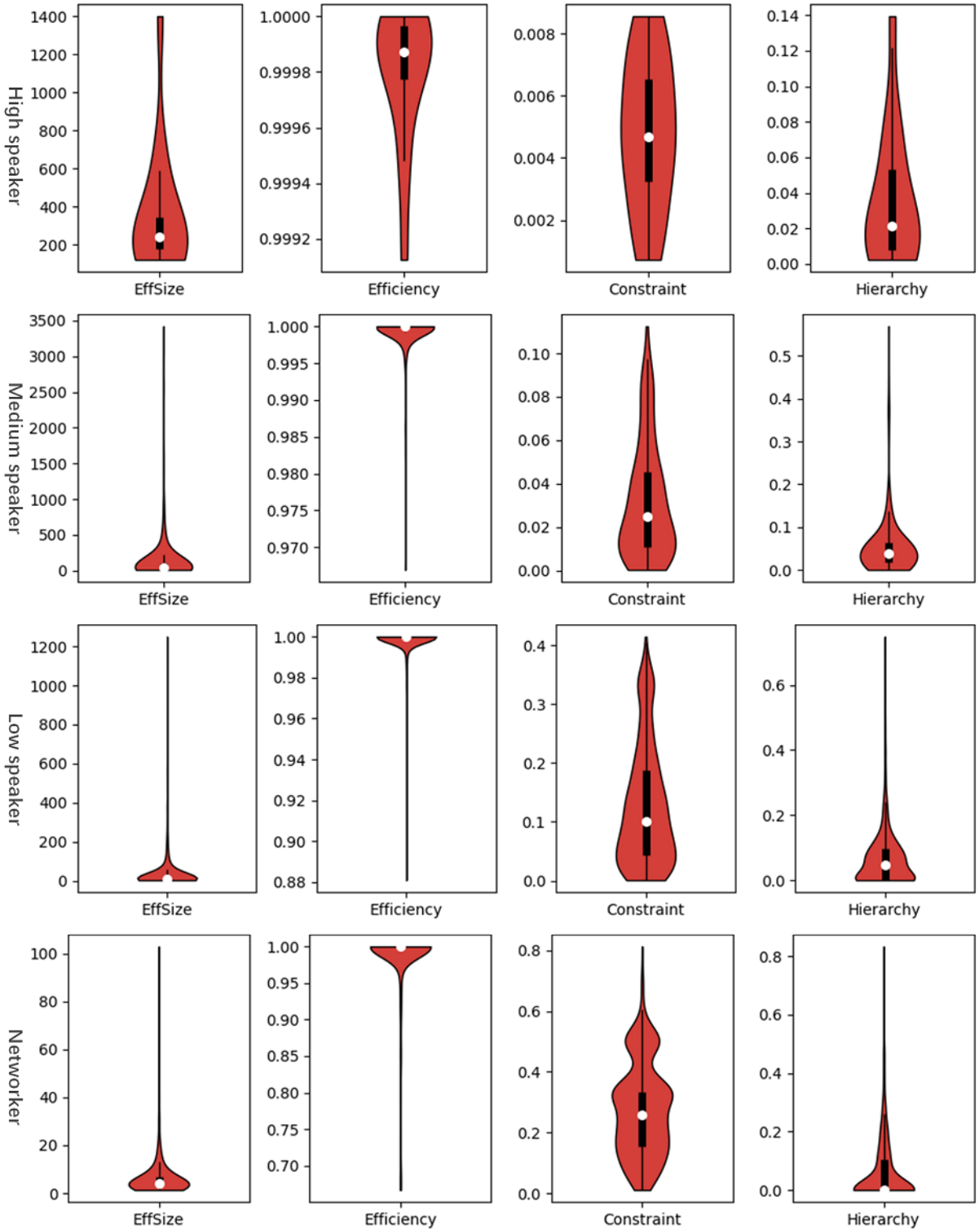
XSL•FO
RenderX

Figure 6. The distribution of users' weighted in-degree centrality, weighted out-degree centrality, and local clustering coefficient (the size of the circle). The depth of the shadow represents the number of users with corresponding centrality and clustering coefficients.



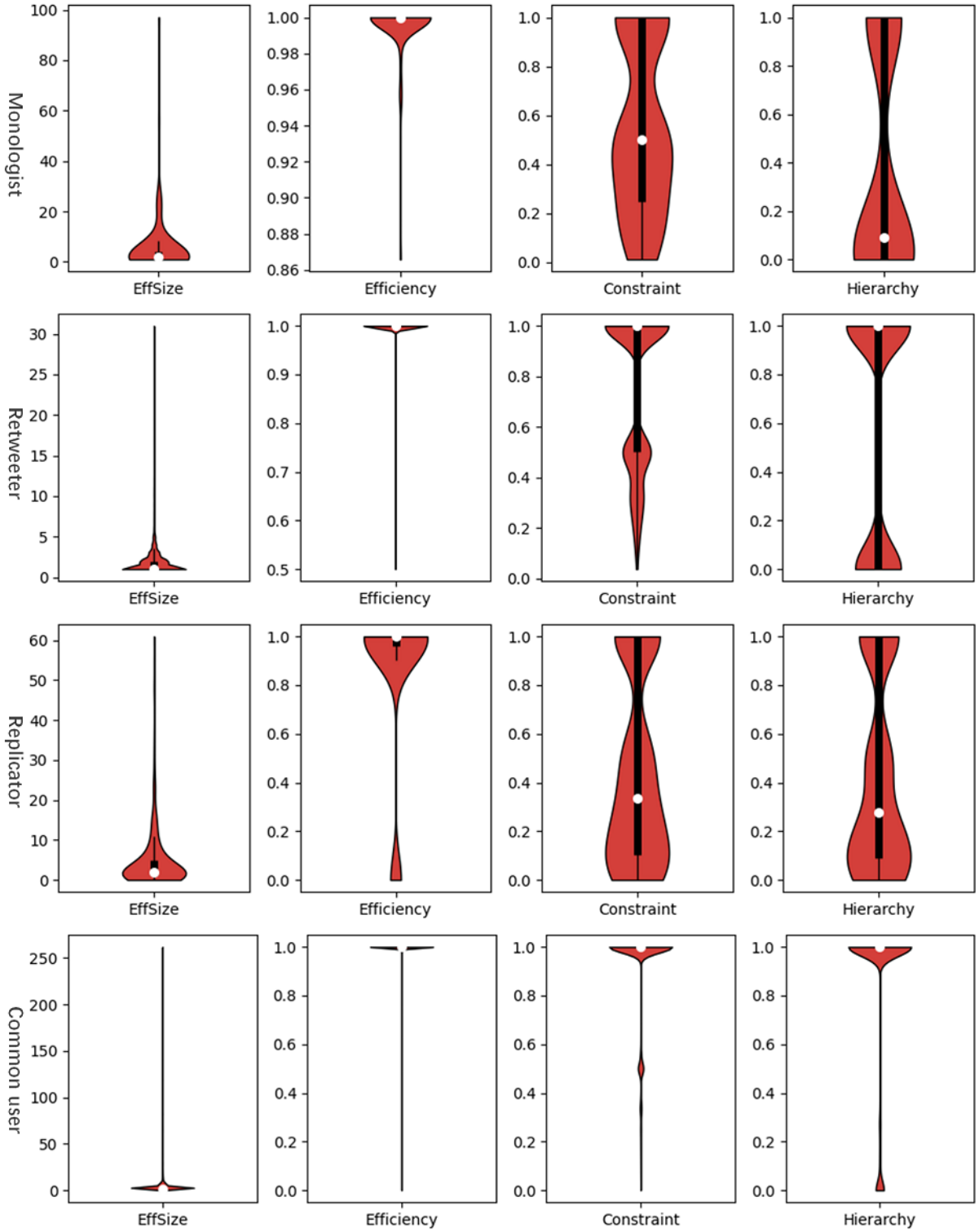
XSL•FO
RenderX

Figure 7. Distribution of users' structural hole indices (speakers and networks). The white dot, and upper and lower lines of the thick black line represent the index's median, third quantile, and first quantile, respectively. The width of the red shadow represents the percentage of specific-category users whose index took on that value.



XSL•FO
RenderX

Figure 8. Distribution of users' structural hole indices (monologists, retweeters, replicators, and common users). The white dot, and upper and lower lines of the thick black line represent the index's median, third quantile, and first quantile, respectively. The width of the red shadow represents the percentage of specific-category users whose index took on that value.



XSL•FO
RenderX

Figure 9. Percentages of tweets from key users.

Figure 10. Composition of user categories in opinion leaders/structural hole spanners and their role in the topic-based echo chamber (left) and attitude-based echo chamber (right).

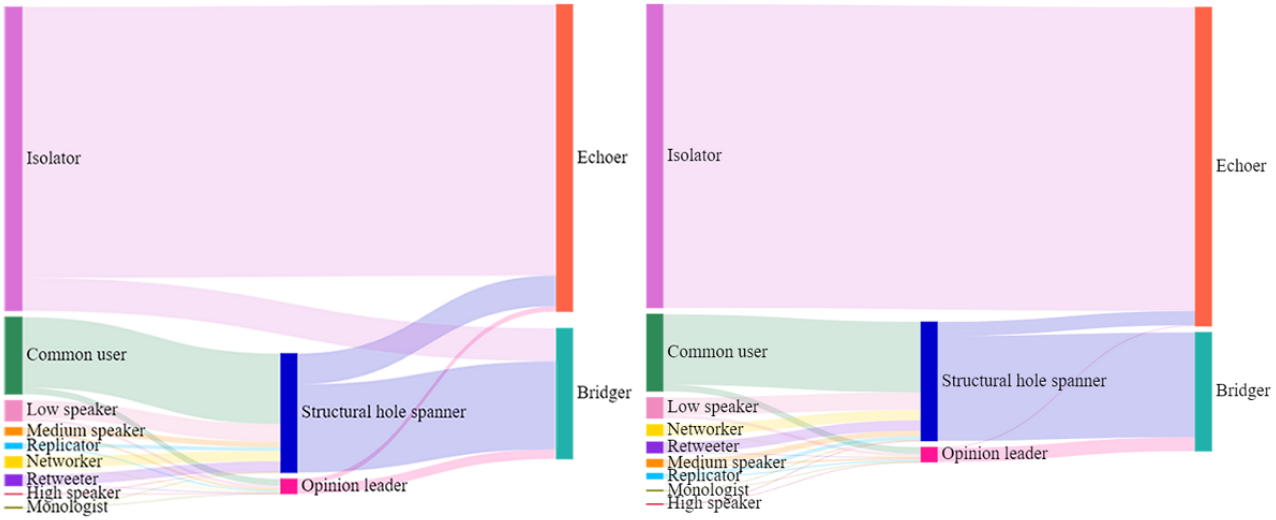


Table 4. Support and confidence of research question 4 (RQ4).

User category	Number of users	Number of users as topic-based echoers/bridgers	Number of users as both topic-based and attitude-based echoers/bridgers	Support[a]	Confidence[b]
RQ4a[c]					
Opinion leader	386	138	8	0.021	0.058
Structural hole spanner	3123	807	23	0.007	0.029
Total	3509	945	31	0.009	0.033
RQ4b[d]					
Opinion leader	386	248	234	0.606	0.944
Structural hole spanner	3123	2316	1968	0.630	0.850
Total	3509	2564	2202	0.628	0.859

[a]Support equals the number of users as both topic-based and attitude-based echoers (RQ4a)/bridgers (RQ4b) divided by the total number of users.

[b]Confidence equals the number of users as both topic-based and attitude-based echoers (RQ4a)/bridgers (RQ4b) divided by the number of users as topic-based echoers/bridgers.

[c]RQ4a: Do key users acting as echoers in topic dissemination tend to play the same role in attitude interaction?

[d]RQ4b**:** Do key users acting as bridgers in topic dissemination tend to play the same role in attitude interaction?

XSL•FO
RenderX

Discussion

Echo Chamber Effect in Online Vaccine Communication

Users showed an overall low echo chamber effect in vaccine-related topic selection and they tended to comment on more diverse topics than retweeting them. Discussions about the status of vaccine development, and vaccination at home and abroad, mostly mixed with conspiracy, largely caught users' attention [72]. The risk of contracting COVID-19 and the serious consequences of refusing to be vaccinated were cocommented with claims of vaccine effectiveness.

In contrast to the findings of Mønsted and Lehmann [5] and Schmidt et al [6], users showed a low echo chamber effect in attitude interaction. Because the COVID-19 vaccine represents a medical innovation directly related to the safety of human life, a rational public, threatened by the public health emergency, was less bound by herd mentality [73]. As the dominant opinion, a positive attitude appealed to following of a neutral crowd, which helped to weaken the echo chamber. These findings are inconsistent with those of Rathje et al [19], possibly because national cultural backgrounds influence the cognition, decision-making, and interactive behavior of people belonging to different parties in the United States and United Kingdom. In addition, in contrast to the findings of Tsai et al [45], we found that the overall homophily was more obvious in commenting than in retweeting. Specifically, users approving vaccines showed a more significant tendency to interact with like-minded neighbors by retweeting than by commenting [43], while users against vaccines or with a neutral attitude acted more significantly by commenting than by retweeting, which suggested that the commenting mechanism might serve as an "anti-spiral of silence" to compete with a "silence spiral" in retweeting to form the global opinion climate [74]. Retweeting amplifies the visibility of individuals' opinions [75], influenced by selective psychology, and opinions contrary to mainstream opinions are silenced. While the commenting network was more modularized and cohesive, users were under greater pressure from within their own communities.

Users' Behavior Patterns Contributing to Their Network Positions

The most common behaviors were helpful in spreading information (high percentages of common users and retweeters), while few users frequently participated in two-way dialogs (low percentage of replicators) [62]. Speakers were relatively scarce, but they created content provoking responses of others, which contributed to their popularity in the network, so as to be regarded as information centers within their communities. Networkers who demonstrated a balance between creating content, sharing content, and being retransmitted were more likely to fill structural holes to link otherwise less-connected communities. The commenting mechanism offered more chances to create cohesive communities, and hence nominate replicators in each cluster as opinion leaders, while the opposite situation was found for the retweeting mechanism and retweeters.

Consistent with Yang et al [14], opinion leaders and structural hole spanners tended to have a stronger influence than ordinary users. These two jointly played an important role in making the information propagate over a wider scale; the former affected their entire communities of the network, while the latter connecting to different communities affected the entire network [76]. Specifically, spanners initiated interactions proactively [77].

Users With Different Network Positions Function in Echo Chamber Formation and Disintegration

Tan et al [66] found that degree centrality and structural holes were complementary at enhancing an organization's innovation performance in low-density networks. Similarly, we found that both opinion leaders and structural hole spanners played a positive role in breaking the echo chamber for topic dissemination and attitude contagion about COVID-19 vaccines. Opinion leaders insulated others against rather than exacerbated the echo chamber [35,36], which contradicts with the findings of Cossard et al [17]. As gatekeepers, because of social pressure and social support based in part on interpersonal trust [78], they were responsible for filtering, curating, and disseminating information they deemed relevant to their social circle to prevent their followers from being trapped in echo chambers. Structural hole spanners diffused information from one group to another, negotiated and synthesized different topics and standpoints, and promoted cooperation in diverse knowledge and ideological fields [79]. An et al [80] found that the same topic could breed multiple emotions and stakeholders with high topic influence that might not necessarily have high sentiment influence, which, to some extent, explained why users as topic-based echoers might not necessarily act as attitude-based echoers, while users as topic-based bridgers also tended to act as attitude-based bridgers in this study. Aware of the negative impact of echo chambers on crisis management and vaccine promotion, despite different cultural backgrounds, government and We-media positively promoted heterogeneity, and the traditional media's agenda-setting power was also evident in both topic and opinion spread [29]. Moreover, Wagner and Reifegerste [81] declared that although isolators were rarely found in their interviews, since participants reported communicating about pandemic-related media coverage "with basically everyone," some participants might turn into isolators during the trajectory of the pandemic. Our findings certainly confirmed this prediction. We found many isolators, which meant that they did not contribute to increasing the scale of information dissemination, which is distinct from the phenomenon noted in disaster-information diffusion [62]. However, the isolators were potential echoers by creating homogenized information and constantly reiterating a single point of view. Although no other users interacted with them, the words of isolators might invisibly reinforce the thoughts of others who saw or read their tweets.

Theoretical Contributions

The main theoretical contributions of this study are as follows. First, echo chambers in vaccine debates during a crisis differ from those related to general social issues. This study not only examined the echo chamber effect in different information-dissemination dimensions (topic, attitude) and

XSL•FO
RenderX

based on different interactive mechanisms (retweeting, commenting), but also dug out the reasons for a low echo chamber effect from the perspective of the relationship of users' network location and their function in preventing or breaking echo chambers. This offers a powerful complement to existing research focusing on echo chambers' form, degree, formation, and depolarization.

Second, we focused on two types of key users, namely opinion leaders and structural hole spanners, and characterized their behavioral patterns, which could be a supplement for feature engineering of these key users' detection or prediction. In addition, referring to the bonding and bridging relationship of social capital, this study proposes two new types of social mediators, namely echoers and bridgers, to quantify key users' impact on echo chambers, thereby enriching the application scope of social capital theories. Hence, users could be classified based on their behavior, network location, impact on echo chambers, and stakeholder theory [63], offering insights for the construction of user portraits.

Third, previous studies about online key users either focused only on their antecedents (factors contributing to individuals occupying a central location/filling a structural hole [77,82,83]) or only on outcome variables (such as the impact of their locations on knowledge management and innovation performance [84,85], information diffusion [14,76], and emotion contagion [86]). This work linked key users' antecedents and outcomes at the same time, which could be used to explore hidden behavioral paradigms.

Fourth, we analyzed the relationship of users' roles in topic-based and attitude-based echo chambers, providing a new research perspective for the dissemination pattern of topic and sentiment.

Finally, most previous studies excluded users who did not interact with others in the data preprocessing step, ignoring their large-scale presence and potential influence on public opinion evolution. This study is thus the first to explore the impact of such users on echo chambers, which could offer a reference for further research about isolators.

Practical Implications

First, although a low echo chamber effect existed in users' selection of topics about vaccines, users tended to focus on some specific topics, namely the status of vaccine development, vaccination at home and abroad, and conspiracies. Health medical and public opinion managers should be aware of the emergence of echo chambers centered on these topics, which might damage international cooperation for vaccinations and epidemic control [87].

Second, users with neutral attitudes toward vaccines were easily influenced by others with determined standpoints. The interaction between opposing viewpoints remained limited. Managers should invite online opinion leaders and structural hole spanners who act as bridgers to offer multiple aspects of vaccine knowledge to correct opponents' misunderstanding and improve their health literacy. At the same time, although provaccine sentiment, as the mainstream opinion, was largely spread and echoed in retweeting, managers should monitor the evolution of other opinions in commenting to prevent the wrong view from turning defeat into victory.

Third, echo chambers have been a major concern of the government, traditional media, and We-media. To obtain better effectiveness, these stakeholders should try to become opinion leaders or structural hole spanners according to their aims by adjusting their own usage behavior on social media. Our results showed that, compared with opinion leaders, structural hole spanners performed better in diffusing diversified topics, whereas opinion leaders performed better in bridging heterogeneous views.

Finally, online isolators should not be ignored. Although these users were reluctant to interact with others and did not receive any feedback from others, they showed interest in creating messages. They were also immersed in personal echo chambers. Managers should take specific measures to break these isolators' echo chambers.

Limitations

First, we simply divided users into two categories, namely echoers and bridgers, according to the rule as to whether the user spread more than one topic or interacted with cross-cutting neighbors, rather than quantifying the extent to which they acted as echoers/bridgers using continuous values. Further exploration is therefore warranted. Second, we did not manually find bot accounts in our data set, which was part of the strategy of Villodre and Criado [62] in their study on Twitter data. To date, no tool has been developed for robot account identification for Weibo. In future research, it will be important to develop automated detection algorithms for larger-scale data [88]. Bot accounts could be classified based on their behaviors such as posting repeatedly to appeal for attention or posting maliciously to damage credibility [89], which might have different impacts on echo chambers. Finally, our data were limited to the early stage of vaccine promotion, and we did not consider the impact of subsequent virus variants on public perceptions of vaccines. Updated data should be supplemented in follow-up studies. Moreover, this research should be extended to other social media platforms (eg, Zhihu), users with higher information literacy [90], and in discussions about different controversial social issues to evaluate the consistency or differences from our results.

Conclusions

By adopting network analysis, this study evaluated and compared the echo chamber effect in users' topic selection and attitude interaction based on different social media mechanisms (retweeting, commenting) in the vaccine debate during the public health emergency of COVID-19. We further used statistical and visual analyses to characterize behavioral patterns of key users (opinion leaders, structural hole spanners), and explored their function in avoiding/breaking or preventing/strengthening topic-based and attitude-based echo chambers. These findings could provide meaningful inspiration for health medical and public opinion managers to break online echo chambers and eliminate vaccine hesitancy.

XSL•FO
RenderX

Acknowledgments

Funding for this study was provided by the National Natural Science Foundation of China (71661167007, 71420107026) and by the National Key Research and Development Program of China (2018YFC0806904-03).

Conflicts of Interest

None declared.

References

1. Kim JH, Marks F, Clemens JD. Looking beyond COVID-19 vaccine phase 3 trials. Nat Med 2021 Feb 19;27(2):205-211. [doi: 10.1038/s41591-021-01230-y] [Medline: 33469205]
2. Chou WS, Budenz A. Considering emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. Health Commun 2020 Dec;35(14):1718-1722. [doi: 10.1080/10410236.2020.1838096] [Medline: 33124475]
3. Soares P, Rocha JV, Moniz M, Gama A, Laires PA, Pedro AR, et al. Factors associated with COVID-19 vaccine hesitancy. Vaccines 2021 Mar 22;9(3):300 [FREE Full text] [doi: 10.3390/vaccines9030300] [Medline: 33810131]
4. Jain G, Sreenivas A, Gupta S, Tiwari A. The dynamics of online opinion formation: polarization around the vaccine development for COVID-19. In: Qureshi I, Bhatt B, Gupta S, Tiwari AA, editors. Causes and symptoms of socio-cultural polarization: role of information and communication technologies. Singapore: Springer Singapore; 2022:51-72.
5. Mønsted B, Lehmann S. Characterizing polarization in online vaccine discourse-A large-scale study. PLoS One 2022 Feb 9;17(2):e0263746 [FREE Full text] [doi: 10.1371/journal.pone.0263746] [Medline: 35139121]
6. 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]
7. Van Poucke M. COVID-19 vaccine hesitancy and shaming on TikTok: A multimodal appraisal analysis. Advance Preprints. Preprint posted online March 30, 2022 URL: https://advance.sagepub.com/articles/preprint/COVID-19_vaccine_hesitancy_and_shaming_on_TikTok_A_multimodal_appraisal_analysis/19387028
8. Van Raemdonck N. The echo chamber of anti-vaccination conspiracies: Mechanisms of radicalization on Facebook and Reddit. Institute for Policy, Advocacy and Governance Knowledge Series, Forthcoming. 2019. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3510196 [accessed 2022-11-12]
9. Wang D, Qian Y. Echo chamber effect in rumor rebuttal discussions about COVID-19 in China: social media content and network analysis study. J Med Internet Res 2021 Mar 25;23(3):e27009 [FREE Full text] [doi: 10.2196/27009] [Medline: 33690145]
10. Sugiono S. Polarization as the impact of strengthening of anti-vaccine groups in social media (echo chamber perspective). J Penelitian Komunikasi Dan Opini Publik 2021 Dec;25(2):166-182. [doi: 10.33299/jpkop.25.2.4194]
11. Biasio LR, Bonaccorsi G, Lorini C, Pecorelli S. Assessing COVID-19 vaccine literacy: a preliminary online survey. Hum Vaccin Immunother 2021 May 04;17(5):1304-1312 [FREE Full text] [doi: 10.1080/21645515.2020.1829315] [Medline: 33118868]
12. Xu S, Coman IA, Yamamoto M, Najera CJ. Exposure effects or confirmation bias? Examining reciprocal dynamics of misinformation, misperceptions, and attitudes toward COVID-19 vaccines. Health Commun 2022 Apr 12:1-11. [doi: 10.1080/10410236.2022.2059802] [Medline: 35414311]
13. Meyer S, Violette R, Aggarwal R, Simeoni M, MacDougall H, Waite N. Vaccine hesitancy and Web 2.0: Exploring how attitudes and beliefs about influenza vaccination are exchanged in online threaded user comments. Vaccine 2019 Mar 22;37(13):1769-1774. [doi: 10.1016/j.vaccine.2019.02.028] [Medline: 30826142]
14. Yang Y, Tang J, Leung C, Sun Y, Chen Q, Li J, et al. RAIN: Social Role-Aware Information Diffusion. 2015 Feb 09 Presented at: AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence; January 25-30, 2015; Austin, TX. [doi: 10.1609/aaai.v29i1.9164]
15. Lou T, Tang J. Mining structural hole spanners through information diffusion in social networks. 2013 Presented at: WWW '13: Proceedings of the 22nd international conference on World Wide Web; May 13-17, 2013; Rio de Janeiro, Brazil. [doi: 10.1145/2488388.2488461]
16. Wu S, Hofman J, Mason W, Watts D. Who says what to whom on twitter. 2011 Presented at: 20th International Conference on World Wide Web; March 28-April 1, 2011; Hyderabad, India. [doi: 10.1145/1963405.1963504]
17. Cossard A, De Francisci Morales G, Kalimeri K, Mejova Y, Paolotti D, Starnini M. Falling into the echo chamber: the Italian vaccination debate on Twitter. 2020 May 26 Presented at: Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020); June 8-11, 2020; Atlanta, Georgia. [doi: 10.1609/icwsm.v14i1.7285]
18. Jeon Y, Kim B, Xiong A, Lee D, Han K. ChamberBreaker: mitigating the echo chamber effect and supporting information hygiene through a gamified inoculation system. Proc ACM Hum Comput Interact 2021 Oct 13;5(CSCW2):472. [doi: 10.1145/3479859]
19. Rathje S, He J, Roozenbeek J, VanBavel J, vanderLinden S. Social media behavior is associated with vaccine hesitancy. PNAS Nexus 2022;1(4):1-11. [doi: 10.1093/pnasnexus/pgac207]

XSL•FO
RenderX

20. Del VM, Gaito S, Quattrociocchi W, Zignani M, Zollo F. News consumption during the Italian referendum: a cross-platform analysis on facebook and twitter. 2017 Presented at: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA); October 19-21, 2017; Tokyo, Japan. [doi: 10.1109/dsaa.2017.33]
21. Bourdieu P, Wacquant L. An invitation to reflexive sociology. Chicago: University of Chicago Press; 1992.
22. Arya B, Lin Z. Understanding collaboration outcomes from an extended resource-based view perspective: the roles of organizational characteristics, partner attributes, and network structures. J Manag 2016 Jun 30;33(5):697-723. [doi: 10.1177/0149206307305561]
23. Beckman CM, Haunschild PR, Phillips DJ. Friends or strangers? Firm-specific uncertainty, market uncertainty, and network partner selection. Organ Sci 2004 Jun;15(3):259-275. [doi: 10.1287/orsc.1040.0065]
24. Gargiulo M, Benassi M. Trapped in your own net? Network cohesion, structural holes, and the adaptation of social capital. Organ Sci 2000 Apr;11(2):183-196. [doi: 10.1287/orsc.11.2.183.12514]
25. Yang H, Lin Z, Lin Y. A multilevel framework of firm boundaries: firm characteristics, dyadic differences, and network attributes. Strat Mgmt J 2010 Mar;31(3):237-261. [doi: 10.1002/smj.815]
26. Granovetter MS. The strength of weak ties. Am J Sociol 1973 May;78(6):1360-1380. [doi: 10.1086/225469]
27. Sheer VC, Rice RE. Mobile instant messaging use and social capital: direct and indirect associations with employee outcomes. Inf Manag 2017 Jan;54(1):90-102. [doi: 10.1016/j.im.2016.04.001]
28. Burt RS. The network structure of social capital. Res Organ Behav 2000;22:345-423. [doi: 10.1016/S0191-3085(00)22009-1]
29. Guo L, A. Rohde J, Wu HD. Who is responsible for Twitter's echo chamber problem? Evidence from 2016 U.S. election networks. Inf Commun Soc 2018 Jul 20;23(2):234-251. [doi: 10.1080/1369118x.2018.1499793]
30. Bruns A. HOW long is a tweet? Mapping dynamic conversations on Twitter using Gawk and Gephi. Inf Commun Soc 2012 Dec;15(9):1323-1351. [doi: 10.1080/1369118x.2011.635214]
31. Gruzd A, Wellman B, Takhteyev Y. Imagining Twitter as an imagined community. Am Behav Sci 2011 Jul 25;55(10):1294-1318. [doi: 10.1177/0002764211409378]
32. Xu WW, Sang Y, Blasiola S, Park HW. Predicting opinion leaders in Twitter activism networks. Am Behav Sci 2014 Mar 13;58(10):1278-1293. [doi: 10.1177/0002764214527091]
33. Guess A, Nyhan B, Lyons B, Reifler J. Avoiding the echo chamber about echo chambers. Knight Foundation. 2018. URL: https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf [accessed 2022-11-12]
34. Bar-Gill S, Gandal N. Online exploration, content choice, and echo chambers: an experiment. Vox EU. 2017 Apr 10. URL: https://cepr.org/voxeu/columns/online-exploration-content-choice-and-echo-chambers-experiment [accessed 2022-11-12]
35. Dubois E, Blank G. The echo chamber is overstated: the moderating effect of political interest and diverse media. Inf Commun Soc 2018 Jan 29;21(5):729-745. [doi: 10.1080/1369118x.2018.1428656]
36. Dubois E, Minaeian S, Paquet-Labelle A, Beaudry S. Who to trust on social media: how opinion leaders and seekers avoid disinformation and echo chambers. Soc Media Soc 2020 Apr 20;6(2):205630512091399. [doi: 10.1177/2056305120913993]
37. Bergström A, Jervelycke Belfrage M. News in Social Media. Digit Journal 2018 Jan 12;6(5):583-598. [doi: 10.1080/21670811.2018.1423625]
38. Burt RS. The contingent value of social capital. Admin Sci Quart 1997 Jun;42(2):339. [doi: 10.2307/2393923]
39. He L, Lu C, Ma J, Cao J, Shen L, Yu P. Joint Community and Structural Hole Spanner Detection via Harmonic Modularity. 2016 Presented at: KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939807]
40. Gong C, Du Y, Li X, Chen X, Li X, Wang Y, et al. Structural hole-based approach to control public opinion in a social network. Engineer Appl Artif Intell 2020 Aug;93:103690. [doi: 10.1016/j.engappai.2020.103690]
41. Swarnakar P, Mukherjee S, Kumar A, Lahsen M. Collaboration networks, structural holes, and homophily: dynamics of non-state actors at the United Nations Climate Change Conference. SSRN preprints. Preprint posted online April 11, 2022 URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4080749
42. Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W. Science vs conspiracy: collective narratives in the age of misinformation. PLoS One 2015 Feb 23;10(2):e0118093 [FREE Full text] [doi: 10.1371/journal.pone.0118093] [Medline: 25706981]
43. Wang D, Zhou Y, Qian Y, Liu Y. The echo chamber effect of rumor rebuttal behavior of users in the early stage of COVID-19 epidemic in China. Comput Human Behav 2022 Mar;128:107088 [FREE Full text] [doi: 10.1016/j.chb.2021.107088] [Medline: 34744299]
44. Wang X, Song Y. Viral misinformation and echo chambers: the diffusion of rumors about genetically modified organisms on social media. Internet Res 2020 Jun 22;30(5):1547-1564. [doi: 10.1108/intr-11-2019-0491]
45. Tsai WS, Tao W, Chuan C, Hong C. Echo chambers and social mediators in public advocacy issue networks. Public Relat Rev 2020 Mar;46(1):101882. [doi: 10.1016/j.pubrev.2020.101882]
46. Schmidt AL, Zollo F, Del Vicario M, Bessi A, Scala A, Caldarelli G, et al. Anatomy of news consumption on Facebook. Proc Natl Acad Sci U S A 2017 Mar 21;114(12):3035-3039 [FREE Full text] [doi: 10.1073/pnas.1617052114] [Medline: 28265082]

XSL•FO
RenderX

47. Zhang X, Ho JCF. Exploring the fragmentation of the representation of data-driven journalism in the Twittersphere: a network analytics approach. Soc Sci Comput Rev 2020 Feb 21;40(1):42-60. [doi: 10.1177/0894439320905522]
48. Xie R, Chu SKW, Chiu DKW, Wang Y. Exploring public response to COVID-19 on Weibo with LDA topic modeling and sentiment analysis. Data Inf Manag 2021 Jan 01;5(1):86-99 [FREE Full text] [doi: 10.2478/dim-2020-0023] [Medline: 35402850]
49. Mo PK, Luo S, Wang S, Zhao J, Zhang G, Li L, et al. Intention to receive the COVID-19 vaccination in China: application of the diffusion of innovations theory and the moderating role of openness to experience. Vaccines 2021 Feb 05;9(2):129 [FREE Full text] [doi: 10.3390/vaccines9020129] [Medline: 33562894]
50. Wang C, Liu Y, Xiao Z, Zhou A, Zhang K. Analyzing internet topics by visualizing microblog retweeting. J Vis Lang Comput 2015 Jun;28:122-133. [doi: 10.1016/j.jvlc.2014.11.007]
51. Wang S, Li Z, Wang Y, Zhang Q. Machine learning methods to predict social media disaster rumor refuters. Int J Environ Res Public Health 2019 Apr 24;16(8):1452 [FREE Full text] [doi: 10.3390/ijerph16081452] [Medline: 31022894]
52. Alkazemi MF, Guidry JPD, Almutairi E, Messner M. #Arabhealth on Instagram: examining public health messages to Arabian Gulf state audiences. Health Commun 2022 Jan 02;37(1):39-47. [doi: 10.1080/10410236.2020.1816283] [Medline: 32873096]
53. Champion V, Skinner C. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health behavior and health education: theory, research, and practice. San Francisco, CA: Jossey-Bass; 2008:45-65.
54. Krippendorff K. Computing Krippendorff's Alpha-Reliability. University of Pennsylvania Repositories. 2011. URL: https://repository.upenn.edu/asc_papers/43 [accessed 2022-11-12]
55. Hagberg A, Schult D, Swartxploring NS, dynamics, function. Exploring network structure, dynamics, and function. 2008 Presented at: 7th Python in Science Conference (SciPy2008); August 19-24, 2008; Pasadena, CA.
56. Li L, Liu Y, Zhu H, Ying S, Luo Q, Luo H, et al. A bibliometric and visual analysis of global geo-ontology research. Comput Geosci 2017 Feb;99:1-8. [doi: 10.1016/j.cageo.2016.10.006]
57. Sufi FK, Razzak I, Khalil I. Tracking anti-vax social movement using AI based social media monitoring. IEEE Trans Technol Soc 2022:online ahead of print. [doi: 10.1109/tts.2022.3192757]
58. Li Z, Zhang Q, Du X, Ma Y, Wang S. Social media rumor refutation effectiveness: Evaluation, modelling and enhancement. Inf Process Manag 2021 Jan;58(1):102420. [doi: 10.1016/j.ipm.2020.102420]
59. Hong D, Gao Z, Luo J, Liu J, Xu M, Suo Z, et al. Emotion analysis of teaching evaluation system based on AI techno-logy towards Chinese texts. MATEC Web Conf 2021 Feb 15;336:05007. [doi: 10.1051/matecconf/202133605007]
60. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract Exper 1991 Nov;21(11):1129-1164. [doi: 10.1002/spe.4380211102]
61. Newman MEJ. Assortative mixing in networks. Phys Rev Lett 2002 Nov 11;89(20):208701. [doi: 10.1103/PhysRevLett.89.208701] [Medline: 12443515]
62. Villodre J, Criado JI. User roles for emergency management in social media: Understanding actors' behavior during the 2018 Majorca Island flash floods. Gov Inf Q 2020 Oct;37(4):101521 [FREE Full text] [doi: 10.1016/j.giq.2020.101521] [Medline: 32904927]
63. An L, Ou M. Social Network Sentiment Map of the Stakeholders in Public Health Emergencies. Libr Inf Serv China 2017;61(20):30. [doi: 10.13266/j.issn.0252-3116.2017.20.013]
64. Wasserman S, Faust K. Social network analysis: methods and applications. Cambridge, UK: Cambridge University Press; 1994.
65. Dubois E, Gaffney D. The multiple facets of influence. Am Behav Sci 2014 Apr 08;58(10):1260-1277. [doi: 10.1177/0002764214527088]
66. Tan J, Zhang H, Wang L. Network closure or structural hole? The conditioning effects of network–level social capital on innovation performance. Entrepreneur Theory Pract 2015 Sep 01;39(5):1189-1212. [doi: 10.1111/etap.12102]
67. Lin Z, Zhang Y, Gong Q, Chen Y, Oksanen A, Ding AY. Structural hole theory in social network analysis: a review. IEEE Trans Comput Soc Syst 2022 Jun;9(3):724-739. [doi: 10.1109/tcss.2021.3070321]
68. Liu J. Whole Net Analysis: A Practical Guide to UCINET Software. Shanghai: Gezhi Publishing House; 2019.
69. Weng L, Menczer F, Ahn Y. Virality prediction and community structure in social networks. Sci Rep 2013 Aug 28;3(1):2522. [doi: 10.1038/srep02522] [Medline: 23982106]
70. Davidsen J, Ebel H, Bornholdt S. Emergence of a small world from local interactions: modeling acquaintance networks. Phys Rev Lett 2002 Mar 25;88(12):128701. [doi: 10.1103/PhysRevLett.88.128701] [Medline: 11909506]
71. Himelboim I, Xiao X, Lee DKL, Wang MY, Borah P. A social networks approach to understanding vaccine conversations on Twitter: network clusters, sentiment, and certainty in HPV social networks. Health Commun 2020 May;35(5):607-615. [doi: 10.1080/10410236.2019.1573446] [Medline: 31199698]
72. Bieber F, Prelec T, Jovic D, Nechev Z. The suspicious virus: conspiracies and COVID19 in the Balkans. European Fund for the Balkans. URL: https://www.balkanfund.org/biepag-publications/the-suspicious-virus-conspiracies-and-covid19-in-the-balkans [accessed 2022-11-12]
73. Boyd D, Golder S, Lotan G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. 2010 Presented at: HICSS-43 IEEE; January 6, 2010; Kauai, HI p. 1530-1605. [doi: 10.1109/hicss.2010.412]

XSL•FO
RenderX

74. Zhu Y, Huang Z, Wang Z, Luo L, Wu S. Influence and extension of the spiral of silence in social networks: a data-driven approach. In: Kaya M, Kawash J, Khoury S, Day MY, editors. Social network based big data analysis and applications. Cham: Springer International Publishing; 2018:143-164.
75. Liu Y, Jin X, Shen H, Cheng X. Do rumors diffuse differently from non-rumors? A systematically empirical analysis in Sina Weibo for rumor identification. In: Advances in Knowledge Discovery and Data Mining. PAKDD 2017. Lecture Notes in Computer Science, vol 10234. Cham: Springer; 2017.
76. Zhang Q, Li X, Fan Y, Du Y. An SEI3R information propagation control algorithm with structural hole and high influential infected nodes in social networks. Eng Appl Artif Intell 2022 Feb;108:104573. [doi: 10.1016/j.engappai.2021.104573]
77. Fu E, Zhuang Y, Zhang J, Zhang J, Chen Y. Understanding the user interactions on GitHub: a social network perspective. 2021 Presented at: 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD); May 5-7, 2021; Dalian, China. [doi: 10.1109/cscwd49262.2021.9437744]
78. Turcotte J, York C, Irving J, Scholl RM, Pingree RJ. News recommendations from social media opinion leaders: effects on media trust and information seeking. J Comput-Mediat Comm 2015 Jun 01;20(5):520-535. [doi: 10.1111/jcc4.12127]
79. Katz E, Lazarsfeld P. Personal influence: The part played by people in the flow of mass communications. Milton Park, UK: Routledge; 2017.
80. An L, Zhou W, Ou M, Li G, Yu C, Wang X. Measuring and profiling the topical influence and sentiment contagion of public event stakeholders. Int J Inf Manag 2021 Jun;58:102327. [doi: 10.1016/j.ijinfomgt.2021.102327]
81. Wagner A, Reifegerste D. "The Part Played by People" in times of COVID-19: interpersonal communication about media coverage in a pandemic crisis. Health Commun 2021 Oct 13:online ahead of print. [doi: 10.1080/10410236.2021.1989786] [Medline: 34645317]
82. He X, Chen Y. Understanding structural hole spanners in location-based social networks: a data-driven study. 2021 Presented at: UbiComp '21: Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers; September 21-26, 2021; virtual. [doi: 10.1145/3460418.3480398]
83. Khadangi E, Bagheri A. Presenting novel application-based centrality measures for finding important users based on their activities and social behavior. Comput Hum Behav 2017 Aug;73:64-79. [doi: 10.1016/j.chb.2017.03.014]
84. Liao Y, Phan PH. Internal capabilities, external structural holes network positions, and knowledge creation. J Technol Transf 2015 Apr 30;41(5):1148-1167. [doi: 10.1007/s10961-015-9415-x]
85. Wu H, Gu X, Zhao Y, Liu W. Research on the relationship between structural hole location, knowledge management and cooperative innovation performance in artificial intelligence. Knowl Manag Res Pract 2020 Sep 01:1-10. [doi: 10.1080/14778238.2020.1813642]
86. Yang Y, Jia J, Wu B, Tang J. Social role-aware emotion contagion in image social networks. 2016 Presented at: Thirtieth AAAI conference on artificial intelligence; February 12-17, 2016; Phoenix, Arizona. [doi: 10.1609/aaai.v30i1.10003]
87. Ullah I, Khan K, Tahir M, Ahmed A, Harapan H. Myths and conspiracy theories on vaccines and COVID-19: potential effect on global vaccine refusals. Vacunas 2021 May;22(2):93-97 [FREE Full text] [doi: 10.1016/j.vacun.2021.01.001] [Medline: 33727904]
88. Xu Y, Zhou Y, Chen K. Observation on spammers in Sina Weibo. 2013 Presented at: 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013); March 22-23, 2013; Hangzhou, China. [doi: 10.2991/iccsee.2013.284]
89. Jin D, Teng J. Study of Bot detection on Sina-Weibo based on machine learning. 2017 Presented at: 2017 International Conference on Service Systems and Service Management; June 16-18, 2017; Dalian, China. [doi: 10.1109/icsssm.2017.7996292]
90. Yang Z. The new stage of public engagement with science in the digital media environment: citizen science communicators in the discussion of GMOs on Zhihu. New Genet Soc 2022 Apr 26;41(2):116-135. [doi: 10.1080/14636778.2022.2063826]

Abbreviations

RQ: research question

XSL•FO
RenderX

Edited by A Mavragani; submitted 01.07.22; peer-reviewed by J Hu, F Sufi; comments to author 27.10.22; revised version received 31.10.22; accepted 09.11.22; published 18.11.22.

Please cite as:
Wang D, Zhou Y, Ma F
Opinion Leaders and Structural Hole Spanners Influencing Echo Chambers in Discussions About COVID-19 Vaccines on Social Media in China: Network Analysis
J Med Internet Res 2022;24(11):e40701
URL: https://www.jmir.org/2022/11/e40701
doi:10.2196/40701
PMID:36367965

©Dandan Wang, Yadong Zhou, Feicheng Ma. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Digital Tools Designed to Obtain the History of Present Illness From Patients: Scoping Review

Carl T Berdahl[1], MD, MS; Andrew J Henreid[1], BA; Joshua M Pevnick[1], MD, MSHS; Kai Zheng[2], PhD; Teryl K Nuckols[1], MD, MSHS

[1]Cedars-Sinai Medical Center, Los Angeles, CA, United States

[2]University of California Irvine Donald Bren School of Information and Computer Sciences, Irvine, CA, United States

Corresponding Author:
Carl T Berdahl, MD, MS
Cedars-Sinai Medical Center
8700 Beverly Blvd
Los Angeles, CA, 90048
United States
Phone: 1 310 423 3091
Email: Carl.Berdahl@csmc.edu

Abstract

Background: Many medical conditions, perhaps 80% of them, can be diagnosed by taking a thorough history of present illness (HPI). However, in the clinical setting, situational factors such as interruptions and time pressure may cause interactions with patients to be brief and fragmented. One solution for improving clinicians' ability to collect a thorough HPI and maximize efficiency and quality of care could be to use a digital tool to obtain the HPI before face-to-face evaluation by a clinician.

Objective: Our objective was to identify and characterize digital tools that have been designed to obtain the HPI directly from patients or caregivers and present this information to clinicians before a face-to-face encounter. We also sought to describe outcomes reported in testing of these tools, especially those related to usability, efficiency, and quality of care.

Methods: We conducted a scoping review using predefined search terms in the following databases: MEDLINE, CINAHL, PsycINFO, Web of Science, Embase, IEEE Xplore Digital Library, ACM Digital Library, and ProQuest Dissertations & Theses Global. Two reviewers screened titles and abstracts for relevance, performed full-text reviews of articles meeting the inclusion criteria, and used a pile-sorting procedure to identify distinguishing characteristics of the tools. Information describing the tools was primarily obtained from identified peer-reviewed sources; in addition, supplementary information was obtained from tool websites and through direct communications with tool creators.

Results: We identified 18 tools meeting the inclusion criteria. Of these 18 tools, 14 (78%) used primarily closed-ended and multiple-choice questions, 1 (6%) used free-text input, and 3 (17%) used conversational (chatbot) style. More than half (10/18, 56%) of the tools were tailored to specific patient subpopulations; the remaining (8/18, 44%) tools did not specify a target subpopulation. Of the 18 tools, 7 (39%) included multilingual support, and 12 (67%) had the capability to transfer data directly into the electronic health record. Studies of the tools reported on various outcome measures related to usability, efficiency, and quality of care.

Conclusions: The HPI tools we identified (N=18) varied greatly in their purpose and functionality. There was no consensus on how patient-generated information should be collected or presented to clinicians. Existing tools have undergone inconsistent levels of testing, with a wide variety of different outcome measures used in evaluation, including some related to usability, efficiency, and quality of care. There is substantial interest in using digital tools to obtain the HPI from patients, but the outcomes measured have been inconsistent. Future research should focus on whether using HPI tools can lead to improved patient experience and health outcomes, although surrogate end points could instead be used so long as patient safety is monitored.

(J Med Internet Res 2022;24(11):e36074) doi:10.2196/36074

KEYWORDS

anamnesis; informatics; emergency medicine; human-computer interaction; medical history taking; mobile phone

XSL•FO
RenderX

Introduction

Background and Significance

Many medical conditions, perhaps 80% of them, can be diagnosed by taking a thorough history of present illness (HPI) [1]. However, in the clinical setting, situational factors such as interruptions and time pressure may cause interactions with patients to be brief and fragmented [2]. One solution for improving clinicians' ability to collect a thorough HPI and maximize efficiency and quality of care could be to use a digital tool to obtain the HPI before face-to-face evaluation by a clinician.

The concept of using a computer to aid in history taking or diagnosis is not new. In fact, some clinicians entered data into computers as early as the 1940s and used software to generate differential diagnoses [3]. In the 1980s, a small minority of clinicians began asking patients to interact with computers directly to enter their own histories, but the process was cumbersome in many cases because patients had to answer dozens or even hundreds of questions [4]. In the early 2000s, investigators tried new methods to decrease the number of required questions, but such tools did not become popular—perhaps because they were not well integrated into emerging electronic health record (EHR) systems [5].

In the contemporary digital age, software developers and research groups in the health sector are developing tools to engage patients in diagnosis and management of their health problems [6-8]. Patients are becoming accustomed to collecting health-related information on their own devices and submitting it to their clinicians. Moreover, starting in 2011, the US federal government began to encourage clinicians and health care systems to collect such information, called patient-generated health data (PGHD), and use it in a meaningful manner. Examples of commonly submitted PGHD include blood pressure measurements, blood glucose measurements, and patient-reported outcome measures for chronic conditions [9,10]. Patient-generated HPI is a less ubiquitous form of PGHD but leveraging it could improve the efficiency and quality of patient care if it is done thoughtfully.

Objectives

In this scoping review, our objective was to identify and characterize patient-facing digital tools that obtain the HPI and present it to clinicians before an in-person encounter. We also sought to describe outcomes reported in studies of these tools, especially those related to usability, efficiency, and quality of care.

Methods

Search Strategy

In consultation with a medical librarian, we developed search terms designed to identify HPI tools of interest from peer-reviewed sources. We then searched the following databases: MEDLINE, CINAHL, PsycINFO, Web of Science, Embase, IEEE Xplore Digital Library, ACM Digital Library, and ProQuest Dissertations & Theses Global. The search was performed in November 2019, and it included all original research and commentary articles that were available in English (for more details of the search strategy, refer to Multimedia Appendix 1).

Article Selection

Titles and abstracts that resulted from the literature search were imported into *DistillerSR* (Evidence Partners) to facilitate screening. Two independent reviewers (CTB and AJH) evaluated all titles and abstracts for relevance based on the inclusion criteria: (1) patient-facing digital tools that obtain the HPI directly from patients or caregivers and (2) present this information to clinicians before a face-to-face encounter. Tools were excluded if they were administered by the clinician rather than the patient, designed to track symptoms over time rather than make a new diagnosis, designed to screen for only one diagnosis (eg, COVID-19 infection), or if there was no mention of outcome measures in any literature describing the tool. Next, in a full-text–review stage, both reviewers reviewed the full text to determine whether the article met the inclusion and exclusion criteria. If there was any disagreement about whether the tool should be included, it was resolved by consensus.

Data Extraction and Synthesis

To obtain information characterizing each tool, a member of the research team (AJH) reviewed the original source material that identified the tool, any other cited references in the original source, and relevant websites of the tool identified through web searches. We developed narrative descriptions of the tools and maintained this information in a data spreadsheet along with their associated references.

To develop a taxonomy of the tools identified, 2 reviewers (CTB and AJH) used a qualitative pile-sorting method [11]. We started by writing the names of tools on small pieces of paper and arranging them into groups that were qualitatively similar. Next, we discussed what qualities the tools shared and which ones made them different from one another. Once we had compiled a list of these defining characteristics, we used them to categorize the tools in our sample.

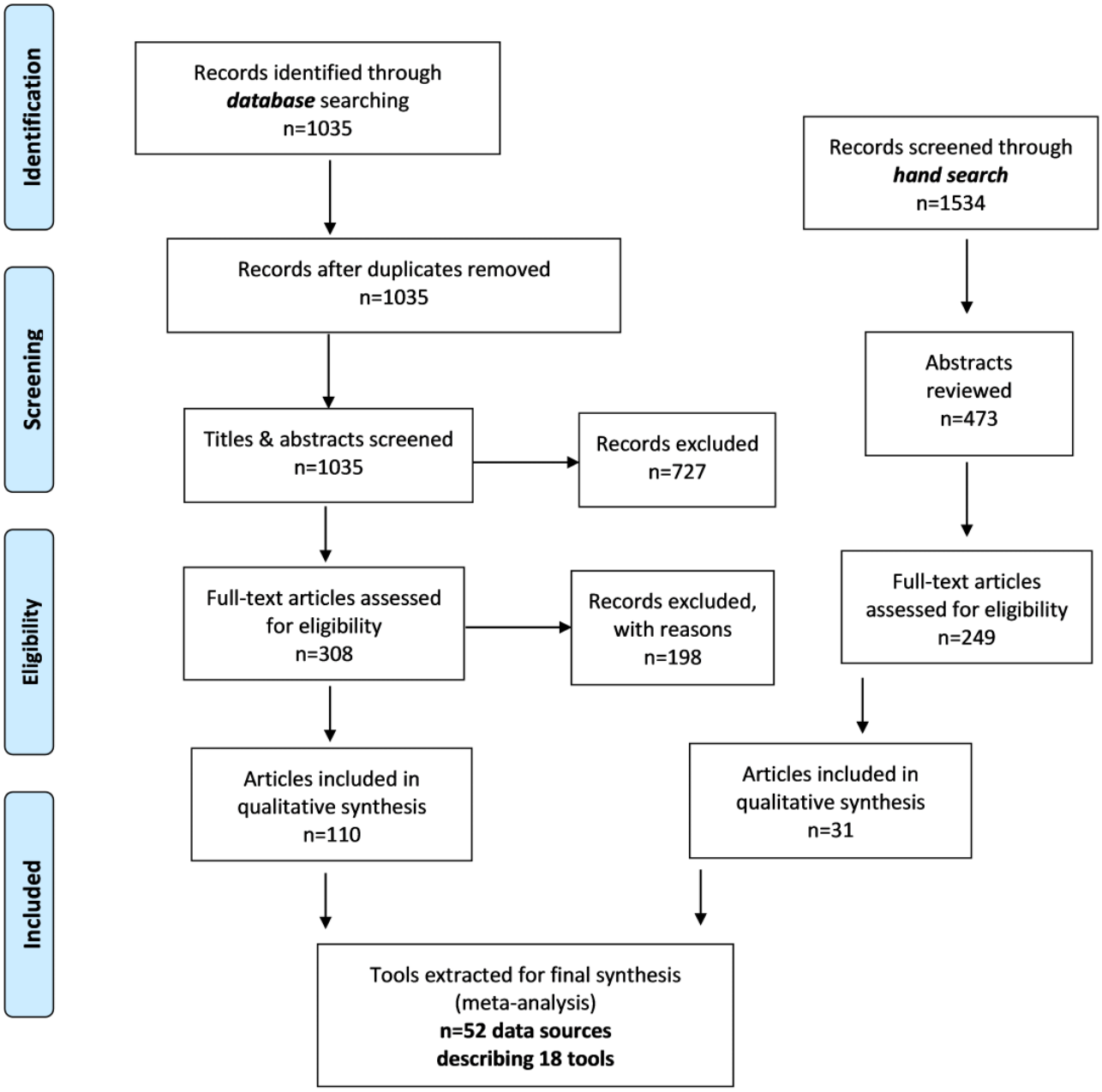
Subsequently, we held a discussion among the research team to review characteristics of the various tools. Informed by this discussion and our newly developed taxonomy, a member of the research team (AJH) reviewed all available materials once again and performed targeted data extraction, including tool name, name of vendor or developer, availability of multilingual support, year of initial development or mention, intended patient user population (eg, pediatric, chest pain, or pulmonary), modality of query delivery (eg, narrative vs structured), decision support capability (patients, clinicians, or both), integration with clinical information systems such as EHRs (yes or no), and outcome measures used in evaluation. A second member of the research team (CTB) reviewed all available materials to verify the accuracy of extracted data. After these steps were completed, we cowrote brief narratives to describe each tool. Finally, we contacted the developer or vendor for each tool to verify the information we had collected.

This study followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for

XSL•FO
RenderX

Scoping Reviews) guidelines for scoping reviews [12,13]. The PRISMA-ScR flow diagram is presented in Figure 1.

Figure 1. Flowchart demonstrating article inclusion and exclusion for our scoping review according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [12,13].



Results

Overview

Our literature search identified 2569 publications of potential interest. After duplicates were removed and titles and abstracts were screened, 557 articles underwent full-text assessment for eligibility. A total of 141 articles met inclusion criteria and were included in our qualitative synthesis. We encountered a total of 18 unique tools to include in our review, which were described by 52 data sources that included outcome measures. For a comparison of individual tools and their characteristics, refer to Table 1, We also compiled a list of synonyms for the process of obtaining the HPI from patients or caregivers using a digital tool, which may be a helpful reference for future investigators (Multimedia Appendix 2).

XSL•FO
RenderX

Table 1. Data extraction table for tools used to obtain the history of present illness.

Name	Developer	Year	Outcome measures	Interaction	Lingual support	Data entry	Patient subpopulation	Delivery	Decision support
Instant Medical History [14-16]	Primetime Medical Software	1985	Completion rate, completion time, and patient usability	Multiple choice	English	Patient and clinician	All	Import to EHR[a]	Clinician
HELP System [17,18]	LDS Hospital	1986	Diagnostic agreement (physician vs tool)	Multiple-choice, open-ended free text	English	Patient	Pulmonary	Import to EHR	Clinician
AIDA [19,20]	Erasmus University	1987	Patient usability, completion time, and complaint agreement and diagnostic agreement (physician vs tool)	Multiple-choice, open-ended free text	Dutch	Patient	Respiration, circulation, gastrointestinal, genitourinary, nervous system, skin, and general disorders	Text-based report	Patient and clinician
ParentLink [21,22]	Blackboard	1989	Completion rate of critical history elements (physician vs tool) and completion time	Multiple-choice, open-ended free text	English	Patient	Pediatrics, emergency, allergy, and trauma	Import to EHR	Clinician
CIDI-Auto [23-26]	Canberra Hospital (computerized version of the World Health Organization's Composite International Diagnostic Interview)	1997	Patient acceptability and diagnostic agreement (physician vs tool)	Multiple-choice, open-ended free text	Multiple languages	Patient and clinician	Psychiatry	Text-based report	Clinician
MEDoctor [27,28]	MEDoctor Systems, Inc	1999	Diagnostic agreement (vignettes vs tool)	Multiple choice	English	Patient	All	Text-based report	Patient
Clinical Expert Operating System (CLEOS) [29,30]	Karolinska Institutet	2008	Patient satisfaction, percentage agreement of symptoms (physician vs tool), and accuracy in excluding acute coronary syndrome	Multiple choice	English, German, and Swedish	Patient	Cardiology	Import to EHR	Clinician
Mediktor [31-36]	Teckel Medical	2011	Diagnostic agreement (physician vs tool)	Conversational, multiple choice	Many languages (>180)	Patient	All	Text-based report	Patient
DocResponse [37-40]	DocResponse	2012	Diagnostic agreement (vignettes vs tool)	Multiple choice	English	Patient and clinician	All	Import to her	Patient and clinician
Digivey [41]	Creoso, in collaboration with researchers at Johns Hopkins University	2013	Patient usability, time to completion, and data entry error rate	Multiple choice	English	Patient	Emergency	Import to her	Clinician
PatientTouch [42]	PatientSafe Solutions Inc	2014	Patient usability and satisfaction	Multiple choice	English, and Spanish	Patient	Emergency	Import to EHR	Clinician
OurNotes [43-45]	OpenNotes	2015	Patient experience and clinician workload (qualitative study)	Open-ended free text	English	Patient and clinician	All	Import to EHR	None

XSL•FO
RenderX

Name	Developer	Year	Outcome measures	Interaction	Lingual support	Data entry	Patient subpopulation	Delivery	Decision support
FirstHx [46,47]	FirstHx Corp	2016	Time to completion, patient usability, and number of questions asked (physician vs tool)	Multiple choice	Multiple languages (10)	Patient	All	Import to EHR	Clinician
Automated Evaluation of Gastrointestinal Symptoms (AEGIS) [48,49]	My Total Health	2016	Rating of tool note quality (comparing physician note vs tool note) and agreement of alarm features (physician vs tool)	Multiple choice	English	Patient	Gastroenterology	Text-based report	Clinician
Digital Communication Assistance Tool (DCAT) [50-53]	aidminutes GmbH	2017	Completion time, patient and physician usability, percentage agreement in symptoms reported, and repeat clinic visits	Multiple choice	Multiple languages (21)	Patient	All	Import to EHR	Clinician
Quro [54-56]	Medius Health	2017	Diagnostic agreement (vignette vs tool)	Conversational, open-ended free text	Multiple languages	Patient	All	Text-based report	Patient
Mandy [57]	Precision Driven Health	2017	Diagnostic agreement (vignette vs tool)	Conversational, open-ended free text	English	Patient	All	Import to EHR	Clinician
Diagnosis and Anamnesis Automated Medical History–Taking Device (DIAANA AMHTD) [58]	Logic-Based Medicine Sàrl, in collaboration with Lausanne University Hospital	2019	Diagnostic agreement (physician vs tool)	Multiple choice	German	Patient	Musculoskeletal	Import to EHR	Clinician

[a]EHR: electronic health record.

Narrative Descriptions of Tools

As the tools we identified differed in many ways (eg, stated purpose, intended setting of use, and outcome measures), we developed a narrative description of each tool (Textbox 1).

XSL•FO
RenderX

Textbox 1. Narrative descriptions of the tools used to obtain the history of present illness from patients.

Digital tools used to obtain the history of present illness and their descriptions

- Instant Medical History
  - General description: Instant Medical History is a tool that was developed by Primetime Medical Software in 1985 to obtain comprehensive information about the history of present illness while also saving physician time and making documentation more complete. The tool has evolved over the last several decades, and it is still in use today.
  - Design: patients are invited to select a chief complaint through a web-based portal from home or in a medical office waiting room. They are then presented with a multiple-choice–question set about their symptom severity, duration, timing, context, modifying factors, and associated signs of illness. This information is next submitted to the electronic health record through an application programming interface for review before the patient visit and additionally for inclusion in the physician's note, if desired [14].
  - Outcomes measured: the company reports that the tool may save up to 6 minutes per clinical encounter [15,16].
  - Extent of use: the tool is currently being used in 7 countries by 44,500 physicians. The vendor estimates that it will be used in 80 million visits in 2020 (email communication with Matthew Ferrante, Primetime Medical Software, July 21, 2020).
- HELP System
  - General description: the HELP System, programmed on the Microsoft Query driver, was described in a 1987 publication by Haug et al [17] titled "A Decision-Driven System to Collect the Patient History." Informaticists at the University of Utah described a computer-administered history-taking system with decision-driven questions designed to create a differential diagnosis for hospital inpatients with pulmonary disease.
  - Design: the system used a cognitive model of question selection along with a Bayesian scoring algorithm that led to targeted question selection using modular diagnostic frames within a program called QUERY. The program contained yes-or-no questions for up to 182 symptoms; however, using the decision-driven system, patients in the study were asked to answer a mean of 51 (SD 31) questions. The program's response report was a list of top 5 differential diagnoses, with accompanying likelihood ranging from 0 to 1.
  - Outcomes measured: when compared with documented discharge diagnoses in a sample of 27 study participants, the tool's list of 5 differential diagnoses included the principal discharge diagnosis for 85% of the patients.
  - Extent of use: the tool was also integrated into the hospital's HELP hospital information system, which pioneered clinical decision support. A subsequent set of tests with an updated diagnostic system and a modified approach to questioning was tested later and reported in the American Association for Medical Systems and Informatics proceedings, including several refinements to the data collection process. Although the tool is no longer in use within the hospital setting, it formed the basis for a subsequent diagnostic application, Iliad, that has been used in medical education [18].
- AIDA
  - General description: AIDA is a software package developed by the department of medical informatics at Erasmus University, Rotterdam, The Netherlands. Its capability to automate medical history taking was described in 1987 by Quaak et al [19] in a special issue of *Computer Methods and Programs in Biomedicine*. The tool's stated purpose was to elicit a comprehensive history and aid physicians in arriving at an accurate diagnosis.
  - Design: patients were asked to read questions on a screen and press keys corresponding with their answers. The system contained >400 questions relating to 179 different items. Regarding acute complaints, the patient was asked system by system about whether symptoms existed (using a 7-point scale from *never* to *always*). If the patient indicated the presence of a symptom, the system asked more questions about frequency, severity, intensity, duration, onset, and location. The final report was displayed to physicians and patients in a narrative format that was designed to mirror how physicians wrote notes.
  - Outcomes measured: the investigators studied the tool's performance compared with the gold standard of a physician interview and found agreement to be 25%. They ultimately found that the tool led to a higher number of *diagnostic hypotheses* and higher diagnostic certainty compared with physician interview alone. Of note, patients required an average of 66 minutes to complete the computerized interview, but their reports of the experience were favorable (92% rated the tool *useful*) [20].
  - Extent of use: the tool was never further developed into a product used in routine care.
- ParentLink
  - General description: ParentLink is a tool designed to obtain information from parents to describe their children's symptoms in the pediatric emergency department setting. ParentLink, the company, was founded in 1989 and acquired by Blackboard in 2014, and the earliest clinical publication was authored by Porter [21] of Boston Children's Hospital in 1999.
  - Design: parents accompanying children with nontraumatic complaints were invited to submit data describing the history of present illness. Parents were asked to independently answer questions on an electronic terminal regarding structured question pathways for fever, respiratory symptoms, or gastrointestinal symptoms. If the reason for the visit was not listed, the parent was invited to select *other*, and they were then shown a textbox for entering open-ended free text. For structured question pathways, parents were invited to answer prompts such as their children's activity level, fluid intake, and urine output.
  -

XSL•FO
RenderX

Outcomes measured: parental data entry took a mean time of 5 minutes, and data were not shown to treating physicians. The investigators measured the validity of content entered by parents and found it to be *comparable* with information documented by physicians with improved sensitivity of parental documentation for hydration status. In a related study from 2002 characterizing the free-response pathway, parents entered between 1 and 142 words to describe the reason for the visit (when presented with a character limit of 2048, which is approximately 350-400 words). Most parents described the chief complaint and elements of the history of present illness, and some asked specific questions and added information about the past medical history. When the parents' text input and physicians' histories in the electronic health record were compared, 23% (7/30) of the parents' entries noted details or observations that were not documented by the physicians [21].

- Extent of use: subsequent studies of ParentLink have included pediatric patients with other complaints, including head trauma, ear pain, and dysuria [22].

- CIDI-Auto
  - General description: CIDI-Auto is a computerized version of the World Health Organization's Composite International Diagnostic Interview (CIDI). The tool's intended purpose is to allow patients to privately answer a series of questions that lead to the automatic generation of a list of psychiatric differential diagnoses. The computerized version was first described and evaluated by investigators at Canberra Hospital in Canberra, Australia, in 1997 [23].
  - Design: patients were asked to sit at a computer workstation during an acute psychiatric hospital admission and answer yes-or-no questions about psychiatric symptoms. The core module of the instrument (CIDI-Core) contained 20 major questions and 59 subquestions, which took approximately 75 minutes to administer [24]. Each patient's responses were organized into a report of diagnosis and symptoms that was given to a physician. The report consisted of a summary of active International Classification of Diseases, Tenth Revision, diagnoses (active in the last 30 days); lifetime diagnoses, which were active >1 month ago; and symptoms in major diagnostic areas.
  - Outcomes measured: psychiatric physicians (the *gold standard*) agreed with 50% of the CIDI-Auto current diagnoses and indicated that only 22% of the CIDI-Auto reports provided useful new diagnoses, although 63% helped to clarify diagnoses, and 58% could save clinicians some time. They endorsed the CIDI-Auto as a possible aid to indirect or remote diagnosis where histories would be taken by nonexpert staff. With regard to patients, 94% liked the computerized interview, 83% understood the questions without difficulty, and 60% felt more comfortable with the computerized interview than with a physician. Education and previous computer experience promoted positive attitudes and satisfaction with the computerized interview [25].
  - Extent of use: the CIDI-Auto has now evolved into the World Health Organization World Mental Health-CIDI Instrument, which is now administered by computer in diverse settings across the world [26].

- MEDoctor
  - General description: MEDoctor is a commercially produced symptom checker designed to produce a list of differential diagnoses for users with acute or chronic symptoms [27]. MEDoctor Systems Inc, the developer of the tool, was founded in 1999, and the company's renewed mission statement since 2017 defines its goal as providing patients access to "actionable medical information...before seeing the clinician" so that patients can make cost-effective decisions about their health.
  - Design: the MEDoctor tool uses a *rule out* basis that includes the value of negative symptom answers (*no input*) to rule out disease probabilities. The engine navigates through >4200 symptoms using Bayesian statistics to produce each interview item based upon numerous factors accumulated, such as sex, onset, and yes-or-no responses. The patient interface consists of drop-down menus and yes-or-no questions to characterize their symptoms. At the end of the process, patients are presented with a list of the top 3 differential diagnoses, and they are offered an opportunity to view a text-based report that displays all completed responses and can be sent to a clinician [28].
  - Outcomes measured: diagnostic agreement between the tool and vignettes is measured.
  - Extent of use: The diagnostic tool has been used worldwide for >5 years. As of July 2020, according to the chief executive officer, it has been completed 36,860 times by users in the United States, the United Kingdom, South Africa, and the Philippines (email communication with Charles Kelly, MEDoctor Systems Inc, July 21, 2020).

- Clinical Expert Operating System
  - General description: Clinical Expert Operating System (CLEOS) is a tool created in 2008 by Zakim [29] that is currently being tested in a clinical trial at Karolinska Institutet in Stockholm, Sweden. The tool is designed to facilitate thorough history taking from patients, and it also includes capabilities for decision support of diagnosis, management, and risk stratification for clinicians.
  - Design: the history-taking program is based on the principles of pathophysiology formalized as software algorithms representing medical knowledge as 450 decision trees. A gating mechanism plus feed-forward and feedback loops enable the tool to perform detailed explorations of any aspects of the history that are significant while also avoiding issues of medical redundancy. The data obtained via CLEOS can be formatted into a narrative summary of the most pertinent history findings similar to that of a physician's note.
  -

XSL•FO
RenderX

Outcomes measured: in a study of 45 patients published in 2008 in which patients underwent usual care and completed a CLEOS interview, the tool detected 3.5 additional problems per patient, some of which were deemed to be clinically significant, such as unrecognized transient ischemic attacks [29]. Currently, the tool is being used in a clinical trial at Karolinska University Hospital in Stockholm, Sweden, for evaluation of patients with chest pain. After initial triage by a physician in the emergency department, patients are invited to answer questions within CLEOS on a tablet computer. Data about pain in the context of the differential diagnoses for chest pain are contained within 29 decision graphs presenting questions that are a mixture of yes-or-no, multiple-choice, and image-based question types. The arrangement of questions for these patients is based on principles of cardiovascular pathophysiology and includes ratings of symptom severity, nature, pain location and radiation, associated cardiac or vagal symptoms, and precipitators of pain. The primary outcome measure is the successful exclusion of acute coronary syndrome at 7 days (using physician diagnosis as the comparator). Secondary outcomes include the ability to calculate risk scores for acute coronary syndrome; exclusion of acute coronary syndrome for 30 days and 1 year; direct costs and resource use; and patient experience regarding feasibility, acceptance, comprehensibility, and technical aspects such as usability [30].

- Extent of use: in the clinical trial, CLEOS is being used at a single hospital: Danderyd University Hospital in Stockholm, Sweden.

- Mediktor
  - General description: Mediktor is a symptom-checker tool developed in 2011 by Teckel Medical and StartUp Health, an IBM affiliate. The developers describe Mediktor as “an interactive tool that can analyze users’ symptoms and evaluate their state of health” [31]. Mediktor is available for consumers to use for free on its website as well as major smartphone and tablet systems in the form of a downloadable mobile app [32,33].
  - Design: the tool presents users with either multiple-choice questions or conversational-style prompts. After a series of questions and responses, users are provided with an assessment of triage urgency, a summary sheet denoting inputted responses and possible diagnoses, and an opportunity to pay for a telemedicine visit with a licensed clinician. Accessible to patients in >180 supported languages, the tool is powered by artificial intelligence and natural language processing.
  - Outcomes measured: according to an academic study conducted in Spain, Mediktor’s diagnostic accuracy was tested against the gold standard of physician diagnosis, and its primary list of diagnoses matched in 91% of the cases [36].
  - Extent of use: the tool has become available through Amazon’s Alexa as a skill that can be enabled for free [34] and through Telegram as a chatbot messaging feature [35]. According to the developer, the tool is currently being used at 3 clinical sites in Europe and the United States with 1.2 million user evaluations in 2019 (email communication with Fabiana Rojas, Mediktor, July 21, 2020).

- DocResponse
  - General description: DocResponse is a clinical workflow, patient intake, and documentation tool launched in 2012 that engages users electronically to schedule visits and enter previsit data about past medical history and acute complaints before their face-to-face encounter with a clinician. The tool was developed by a team of technology experts and physicians from various specialties [37]. The tool’s stated goal is to reduce data entry by front desk reception personnel, medical assistants, and clinicians.
  - Design: patients are invited to enter data on any hospital-provided or personal smart device. The user can complete consent forms; enter data describing past medical, family, surgical, and social histories; complete the review of systems; and use an *assessment tool* that populates the history of present illness in the medical record before the patient’s arrival. Clinicians are then provided with clinical decision support, including a preliminary differential diagnosis and treatment recommendation. At the end of the visit, the tool can also generate relevant education materials for the patient to review [38].
  - Outcomes measured: in a 2015 study, DocResponse was found to be the symptom checker most likely to arrive at the correct principal diagnosis out of a sample of 23 similar tools (although its diagnostic accuracy was 18 out of 36 [50% CI 33%-67%]) [39,40].
  - Extent of use: the tool is currently being used at >170 clinical sites in many care settings such as urgent care, primary care, orthopedics, pediatrics, gastroenterology, multispecialty, and federally qualified health centers throughout the United States. Per vendor report, the tool was used for >225,000 encounters in 2019 (email communication with Tarek Fahl, MD, DocResponse, July 16, 2020).

- Digivey
  - General description: Digivey is a self-administered computer-assisted interview tool (delivered on Digivey survey software) that was designed to improve diagnostic accuracy and patient safety.
  - Design: A 2013 academic publication from Newman-Toker’s research group at Johns Hopkins University School of Medicine described the design of this tool [41]. Digivey delivered adaptive questionnaires with approximately 40 items to emergency department patients to elicit clinical history information about their individual symptom presentations. Participants used one of three electronic devices: mobile kiosk, touch-screen monitor, or laptop computer.
  - Outcomes measured: all 3 electronic device interfaces were deemed to be usable, there were low rates of user error, and administration required approximately 6 minutes [41].
  - Extent of use: Newman-Toker has continued to use the system extensively in research studies, including for data collection from patients in an ongoing 5-site multicenter clinical trial that screened >3000 patient encounters using the tool. The Johns Hopkins department of neurology is working to deploy the system in several clinical areas, including via mobile platforms where patients will be able to consistently self-administer their medical data (email communication with David Newman-Toker, MD, July 7, 2020).

XSL•FO
RenderX

- PatientTouch
  - General description: PatientTouch is a tool created by PatientSafe Solutions Inc to elicit patients to use an electronic questionnaire to describe their symptoms. In a pilot study published in 2014, investigators at Los Angeles County+USC Medical Center described the tool's use among adult medically stable patients presenting to the emergency department.
  - Design: patients were asked to use a handheld touch-screen tablet and complete an electronic questionnaire in either English or Spanish before contact with their physician. First, eligible patients were prompted to select one of six chief complaints: low back pain, upper extremity injury, lower extremity injury, abdominal pain, headache, or motor vehicle collision. Next, they were guided through chief complaint–specific algorithms based in questions provided in multiple-choice or point-and-click format.
  - Outcomes measured: users were asked to rate their experience, usability, and satisfaction with the technology. Patients reported feeling that the device would help them better communicate with their physician and improve their overall quality of care [42].
  - Extent of use: the tool is not presently being used at any clinical sites, according to its developer.
- OurNotes
  - General description: OurNotes is a tool developed in 2015 by the OpenNotes movement in which the patient is invited to contribute information to their own ambulatory visit notes [43]. The developers' stated goal is to enhance patient-clinician communication, support patients' engagement in their care, and save clinician time.
  - Design: a pilot study publication described the enrollment of primary care clinicians and patients who were already registered to use an institutional messaging portal. A few days before a scheduled visit, the patient was asked to submit answers to two open-ended free-text questions: "How have you been since your last visit?" "What are the most important things you would like to discuss at your visit?" Before the visit, the patient's answers to these inquiries were routed to the clinician who could then incorporate them into the visit note. The patient could view the note by logging in to the patient portal after the visit [44].
  - Outcomes measured: patient experience and clinician workload were assessed in a qualitative study, with patients *supporting the idea* and clinicians thinking it was possible that their workload could decrease if patients helped to produce their own visit notes [44].
  - Extent of use: 4 academic medical centers participated in the pilot study, with 160 primary care clinicians and 2500 patients participating between 2018 and 2020. Evaluation of the pilot is underway, and pilot sites are considering expansions of the program [45].
- FirstHx
  - General description: FirstHx is a patient intake tool founded in 2016 and developed by physicians in Toronto, Ontario, Canada, that collects information describing a patient's symptoms and generates focused medical histories in advance of an acute care visit. The developer's stated goals include a dedication to improving physician-patient communication, reducing documentation time, mitigating medical error, and improving the quality of care [46].
  - Design: it is built specifically for use in emergency departments, urgent care, and telemedicine. The tool is designed to use a line of questioning similar to that of physicians and generate a history of present illness report covering >240 presenting complaints in up to 10 supported patient languages.
  - Outcomes measured: the company's website reports that patients can complete the 3- to 6-minute digital intake using either their personal smartphone, provided tablet device, or through a dedicated kiosk [47].
  - Extent of use: the developers state that it is available in the Epic App Store. It has undergone pilot testing at >10 sites, and its estimated use will be 600,000 visits per year (email communication with Mark Benaroia, MD, FirstHx, January 27, 2022).
- Automated Evaluation of Gastrointestinal Symptoms
  - General description: Automated Evaluation of Gastrointestinal Symptoms (AEGIS) is a tool developed in 2016 by researchers in Los Angeles, California, United States, and Ann Arbor, Michigan, United States, to automatically obtain reports of symptoms from patients in the gastroenterology clinic and transform them into a coherent history of present illness.
  - Design: patients are invited to answer questions through a web-based portal to characterize symptoms as delineated by the Patient-Reported Outcomes Measurement Information System framework. If a patient reports several symptoms, the AEGIS system prompts the user to select the most bothersome symptom. An algorithm generates a physician-facing report that is designed to look like a physician-generated history of present illness.
  - Outcomes measured: in 1 peer-reviewed publication, patients underwent both computer-generated history taking and usual care, and blinded ratings compared the quality of both sets of documentation. Computer-generated histories were found to be more complete, more useful, better organized, more succinct, and more comprehensible [48]. Another peer-reviewed publication focused on the AEGIS system's ability to detect alarm features and found its performance to be superior to physician detection (alarm features detected in 53% vs 27%, respectively) [49].
  - Extent of use: AEGIS has not been used in routine patient care outside of the aforementioned studies (email communication with Christopher Almario, MD, Cedars-Sinai Medical Center, July 19, 2020).
- DCAT

XSL•FO
RenderX

- General description: DCAT, created in 2017 by German research and technology experts at *aidminutes GmbH*, is an anamnesis tool that facilitates communication between patients and health care providers in primary care settings, with particular attention to refugee care sites in Germany that experienced an influx of Syrian patients. Aidminutes, which is affiliated with the department of general practice at the University Medical Center Göttingen in Göttingen, Germany, refers to the tool as a digital communication assistance tool that aims to improve diagnostic accuracy by improving medical history taking [50].
- Design: once patients arrive at the outpatient clinic waiting room, they are given tablet devices that allow them to enter information describing their symptoms. The tool also facilitates entry of data describing past medical history, current medications, allergies, and any psychological comorbid conditions. To facilitate use by patients with limited literacy, the tool is designed to be visually intuitive, and it also includes audio prompts. After patients enter their data, responses are translated into the clinician's preferred native language and presented as a data synopsis, including alerts regarding *red flags* in the history that have been discovered [51].
- Outcomes measured: use of the tool has been described in peer-reviewed publications, including a recent study demonstrating good usability and acceptance by patients speaking Levantine Arabic, Modern Standard Arabic, Egyptian Arabic, Farsi, Sorani Kurdish, and Turkish [52]. In 1 study, patients successfully completed their assessments in an average time of 13 minutes [53].
- Extent of use: the tool was used in approximately 10,000 multilingual visits in urgent care and family medicine clinics during 2021 (email communication with Frank Muller and Andreas Lippke, May 3, 2022).

- Quro
  - General description: Quro is a *chatbot health assistant* created in 2017 by Medius Health (Sydney, New South Wales, Australia) that uses machine learning artificial intelligence to deliver health assessments to users [54]. Patients can access the tool through a desktop web browser or smartphone.
  - Design: users are invited to answer a set of free-response and multiple-choice prompts generated by the back-end sequential question prediction algorithm using a large-scale clinical knowledge graph to mimic the taking of a medical history. Each new question is predicted based on previous user-chat context. After completion, the patient is provided a list of differential diagnoses with interpretations, nearby health services, recommendations about the urgency of their condition, and a detailed report that displays the answers to their individual set of questions. The tool's website advertises a built-in medical dictionary of >7 million disease and illness presentation patterns, scored against content sourced from "trusted sources like medical journals" to produce its personalized clinical assessments [55].
  - Outcomes measured: in an article published by Quro's developers, the tool's triage accuracy was assessed using 30 case-based scenarios (10 for emergency care, 10 for general practitioner care, and 10 for self-care) and found to be accurate in 83% (25/30) of the cases [56].
  - Extent of use: Quro is marketed to, and used by, several health and wellness service providers to engage with, and onboard, patients remotely (email communication with Shameek Ghosh, founder and chief technology officer, September 1, 2020).

- Mandy
  - General description: Mandy is a primary care conversational-style dialogue system developed in 2017 by a public-private research partnership of computer scientists at the University of Auckland in Auckland, New Zealand, funded by Precision Driven Health and Orion Health, aimed at improving health outcomes through data science. The tool is designed to assist health care staff by automating the patient intake process.
  - Design: patients interact with the tool by answering the conversational-style prompts with open-ended free-text responses. An analytic engine uses natural language processing to interpret the patient's text, queries a symptom-to-cause mapper for reasoning about potential diagnostic causes, and then generates further interview questions. Once the system has obtained sufficient information from the patient, it reports a differential diagnosis for the clinician to consider.
  - Outcomes measured: in a proof-of-concept paper, the developers reported on the application's *question accuracy* (ability to generate key follow-up questions after an initial chief complaint) and diagnosis *prediction accuracy* (ability to generate relevant differential diagnoses after obtaining responses to its questions) by using gold-standard cases from a medical textbook. Out of 6 cases, the tool generated appropriate questions in 5 cases and had case-by-case prediction accuracy ranging from 14% to 100% [57].

- Diagnosis and Anamnesis Automated Medical History–Taking Device
  - General description: Diagnosis and Anamnesis Automated Medical History–Taking Device (DIAANA AMHTD) was developed by Logic-Based Medicine Sàrl in collaboration with Adrien Schwitzguebel of Lausanne University Hospital in Lausanne, Switzerland, to improve diagnostic accuracy through helping physicians to generate more comprehensive differential diagnoses. To date, the tool exclusively addresses musculoskeletal complaints.
  - Design: in a pilot study published in 2019, the developer and coinvestigators tested DIAANA AMHTD at a teaching hospital in Geneva, Switzerland. Patients were eligible if they were waiting to be seen at an ambulatory clinic for evaluation of musculoskeletal symptoms. Patients in the experimental group were asked to complete a digital form on a touch pad before the resident physician's evaluation, including questions about specific symptoms and risk factors. Through the completion of an adaptive questionnaire that draws from a data set of 269 questions, DIAANA AMHTD then generated a comprehensive anamnesis summary and a list of top differential diagnoses. The list of differential diagnoses (selected from 126 possibilities) was then presented to medical residents for consideration before the face-to-face evaluation.
  - Outcomes measured: residents who used the tool were found to be more likely to have included the final diagnosis from the list of initial differential diagnoses than those residents who did not use the tool (75% vs 59%, respectively) [58].

XSL•FO
RenderX

- Extent of use: the tool is currently being used regularly by a single physician with 250 patient encounters in 2019 and has also been implemented by a Swiss telemedicine system called Soignez-Moi in Biel, Switzerland (Email communication with Adrien Schwitzguebel, MD, July 2020).

Descriptions of Tool Characteristics in the Taxonomy

As a result of our pile-sorting procedure, we developed an HPI tool taxonomy that includes the following categories: interaction modality, lingual support, patient versus caregiver data entry, patient subpopulation (by age, chief complaint, or body system), modality of result delivery, and decision support target (patient or clinician). The resulting taxonomy is reported in Textbox 2.

Textbox 2. Taxonomy of characteristics describing the tools used for obtaining the history of present illness in the study sample.

Query style

- Open-ended free text
- Multiple choice
- Conversational style (*chatbot*)

Language capability

- Single language only
- Multiple languages
- Discordant language support

Tasked to perform data entry

- Patient
- Parent or caregiver

Patient subpopulation

- All patients
- Limited by patient age, care setting, or body system

Output format

- Text-based report
- Data imported to electronic health record

Decision support

- Patient facing
- Clinician facing

Interaction Modality

Among the 18 tools reviewed, 1 (6%) used exclusively open-ended free-text interaction (OurNotes [43]), whereas 3 (17%) used a conversational style of interaction (ie, a *chatbot* style) simulating a human text-message interaction (Mediktor [31], Quro [55], and Mandy [57]). The remaining tools (14/18, 78%) used either an entirely multiple-choice format or used primarily a multiple-choice format with some open-ended free-text components.

Lingual Support

Of the 18 tools, 1 (6%) was developed specifically to address language barriers in Germany, and it was deployed especially to assist Syrian refugees with communication (DCAT [50]), whereas 1 (6%) specifically reported the capability to facilitate clinical communication between patients and clinicians in language-discordant encounters (FirstHx [46]); 5 (28%) other tools reported the ability to capture the HPI in English as well as other languages (CIDI-Auto [25], Clinical Expert Operating System (CLEOS) [30], Mediktor [31], PatientTouch [42], and Quro [55]), 1 (6%) was Dutch only (AIDA [19]), and 1 (6%) was German only (Diagnosis and Anamnesis Automated Medical History–Taking Device [DIAANA AMHTD] [58]). The remaining tools (9/18, 50%) were English only [16,20,21,27,40,41,46,51,57].

Patient Subpopulation

Of the 18 tools, 10 (56%) were tailored to specific patient populations by reason for visit or body system (HELP System: pulmonary [20]; AIDA: multiple specific body systems [19]; ParentLink: pediatric emergency, allergy, and trauma [21,22]; CIDI-Auto: psychiatry [26]; CLEOS: cardiology [29]; Digivey: emergency [41]; PatientTouch: emergency [42]; Automated

XSL•FO
RenderX

Evaluation of Gastrointestinal Symptoms (AEGIS): gastroenterology [51]; and DIAANA AMHTD: musculoskeletal [58]). The remaining tools (8/18, 44%) did not specify a target subpopulation of patients [14,27,31,40,46,49,53,56,57].

Modality of Delivery

Of the 18 tools, 12 (67%) reported the capacity to import patient data directly into the clinician's EHR system [14,20,21,29,40-42,46,49,53,57,58], and the other 6 (33%) provided users a text-based report of patients' responses to prompts or a list of top differential diagnoses [19,26,27,34,51,56]. It is possible that some of the tools with text-based reports have back-end functionality that allows for transmission of information to partnering clinicians, although we are unable to ascertain which tools can do this based on review of tool web sites and articles in our sample.

Decision Support

Of the 18 tools, 5 (28%) offered patient-facing decision support that came either in the form of a triage acuity level or a list of differential diagnoses that was displayed for patients to read [19,27,34,40,56], whereas 15 (83%) offered clinician-facing decision support, which was displayed to clinicians as a list of differential diagnoses with or without links to evidence-based management recommendations [14,19-21,26,29,40-42,49,51,53,57,58].

Outcome Measures

There was a wide range of outcome measures reported among the tools in the sample. Outcome measures were categorized as being related to the domains of tool usability, efficiency of care, and quality of care.

Tool Usability

Patient usability was studied for 33% (6/18) of the tools (Instant Medical History [14], AIDA [19], Digivey [41], PatientTouch [42], FirstHx [49], and DCAT [53]). Other related constructs were measured for several (5/18, 28%) of the other tools, including patient acceptability (CIDI-Auto [23-26]), patient satisfaction (PatientTouch [42] and CLEOS [29]), and patient experience (CLEOS [29] and OurNotes [46]). Completion rate was studied for 11% (2/18) of the tools (Instant Medical History [14] and AIDA [19]), and data entry error rate was studied for 6% (1/18) of the tools (Digivey [41]). Clinician usability was only reported for 6% (1/18) of the tools (DCAT [50,51,53,59]), although a concern about clinician workload was mentioned in the qualitative study of OurNotes [44].

Efficiency of Care

Time to completion was the most commonly reported measure of efficiency, used in studies of 33% (6/18) of the tools (Instant Medical History [14], AIDA [19,20], ParentLink [21,22], Digivey [41], FirstHx [46,47], and DCAT [50,51,53,59]). A study of FirstHx compared the number of questions asked by the tool and the clinician. There was no report of any research related to the impact on length of the visit or time spent by clinicians on direct or indirect care [46,47]. The protocol describing a currently ongoing clinical trial of CLEOS reported a plan to measure direct costs and resource use for patients in the intervention group versus a control group (refer to the next section [Quality of Care] for additional details) [30]. The protocol of another clinical trial reported a plan to measure the rate of repeat visits to a clinic as a proxy measure for improved patient-clinician communication at a refugee clinic (DCAT) [53].

Quality of Care

There were a wide variety of measures that have implications for quality of care. A study of 6% (1/18) of the tools reported agreement of the patient's chief complaint between the tool and a physician (AIDA) [20]. Another study reported agreement in the questions asked (FirstHx) [47], and articles describing 11% (2/18) of the tools reported agreement in the symptoms reported between the tool and those the physician documented (CLEOS [30] and DCAT [53]). A study of AEGIS reported agreement in alarm features reported via the tool or in physician documentation [49], and a study of ParentLink reported completion rate of critical history elements for histories acquired via the tool and a physician [21]. Another study of AEGIS compared the quality of documentation produced by the tool with that produced by treating physicians by having blinded physicians rate the quality of documentation produced by both sources [48].

Diagnostic agreement was the most commonly reported measure related to quality of care. Studies of 22% (4/18) of the tools reported on agreement between tool-produced diagnoses and vignettes with prespecified diagnoses (MEdoctor [28], DocResponse [39,40], Quro [56], and Mandy [57]). Studies of 28% (5/18) of the tools reported agreement between physician-generated diagnoses in health care environments and diagnoses generated by the tools (HELP System [17], AIDA [19], CIDI-Auto [25], Mediktor [36], and DIAANA-AMHTD [58]). The research protocol for an ongoing study of CLEOS was the only article describing measurement of patients' health outcomes. In the study, the primary outcome was reported to be the tool's ability to exclude acute coronary syndrome at 7 days for patients presenting to the emergency department for chest pain, and secondary outcomes were the ability to exclude acute coronary syndrome at 30 days and 1 year [30].

Discussion

Principal Findings

In this scoping review, we identified 18 digital tools that have been used to collect HPI information about patients' symptoms and communicate it to clinicians. These tools varied widely in their stated purposes (eg, to improve patient-clinician communication, to enhance patient engagement, to save clinician time, to improve diagnostic accuracy, and to exclude acute coronary syndrome). They also varied widely in their interaction modality (open-ended free text vs multiple choice vs conversational style), modality of results delivery (text-based report vs EHR integration), and decision support capability (patient facing vs clinician facing). There did not seem to be any consensus on how information should be collected or how it should be presented to patients or clinicians.

For the tools identified in this review, peer-reviewed publications describing empirical evaluation findings were

XSL•FO
RenderX

somewhat limited. However, results from several studies provide preliminary evidence that implementation of such tools is acceptable to patients; for example, a well-designed usability study demonstrated that patients found the precursor to Digivey easy to use [41]; a multi-institutional survey revealed that 2 out of 3 clinicians who had used OurNotes supported continued coauthorship of visit notes with patients [45]; and studies of ParentLink and AEGIS found that both tools improved documentation of critical elements of patients' histories [21,22,49,60]. In the coming years, investigators in Europe will determine, through a clinical trial of CLEOS, whether use of the tool can predict adverse clinical outcomes more accurately than clinicians among a sample of 2000 patients evaluated for acute chest pain. As these tools are studied more intensively, we recommend that academic evaluators should adopt one of several recently published frameworks to design comparative effectiveness studies of multiple tools [61-63]. However, although the rigorous randomized controlled trial of CLEOS is underway, other investigators and tool developers may not be willing to spend the time and resources to follow patients through the many steps in their journeys and measure health outcomes in a controlled fashion. Instead, surrogate outcomes may be considered so long as measures are in place to ensure that patients are not being harmed by HPI-tool implementation. In the emergency department work environment, for example, reducing documentation time for clinicians and improving satisfaction for patients could be adequate primary outcomes so long as there is no evidence of health outcome inferiority, such as increased length of stay for inpatients or increased return emergency department visits for discharged patients. Future investigators should consider prioritizing the following measures for patients: usability, documentation time, accuracy of the history, and satisfaction with the visit. For clinicians, we recommend prioritizing usability, documentation time, accuracy and completeness of the history (including elicitation of red flags for certain chief complaints), and satisfaction with the visit. For the visit overall, measures could include patient-clinician interaction time and emergency department length of stay (measured from the time the clinician signs up for the patient until the disposition decision).

Although academic investigators are carrying out such studies to test HPI tools, vendors of commercial products are already implementing their products in clinical environments across the world; for example, the vendor of Instant Medical History reported that it would be used in an estimated 80 million visits in 2020. To our knowledge, this tool has not been rigorously studied in peer-reviewed literature; however, it has been in use for decades, and its recent, rapid real-world implementation may indeed lead to important advances in the field of informatics.

On the basis of our review of the literature and our experience in the field, there are several barriers to the adoption of digital HPI tools by clinicians and patients. First, selecting 1 tool for testing and implementation is nontrivial because there is a lack of consensus on which features would make care more efficient and higher quality. Second, clinicians may not be interested in adopting new technology before a careful study of workflow is undertaken. Third, integration of a tool into a clinician's EHR system may require substantial time and effort. Fourth, clinicians may fear that storing patient-generated information may increase the risk of malpractice litigation. Fifth, asking patients to use technology may exacerbate already existing health disparities for vulnerable patients. Sixth, and last, there is no clear evidence that adopting digital HPI tools leads to improved health outcomes—although evidence will likely be forthcoming in the next few years.

After evaluating all the tools described in this review, our opinion is that patients can successfully and safely be engaged to compose their own HPI information. We believe that both patients and clinicians would benefit from a tool with an intuitive design that allows patients to create a history through a combination of unstructured and structured prompts and then transforms this history into a narrative that is cohesive and communicates the key elements of the history to clinicians so that the differential diagnosis can be narrowed appropriately. We believe that the ideal tool would begin with a patient-generated narrative and then ask patients to complete a sequence of closed-ended questions to narrow the differential diagnosis. This is the sequence that clinicians are trained to use because patients can offer information that they believe is relevant, and then clinicians can use closed-ended questions to fill in gaps in knowledge and narrow the differential diagnosis [64]. For blocks of text generated by patients, natural language processing can be used to add structure [65] and feed analyses that can determine which closed-ended questions should be administered to obtain a complete history in a reasonable amount of time. This approach is similar to the one used by Quro, and it avoids the need for extremely long question chains that were used in older rule-based tools such as AIDA. This text-based approach could additionally be supplemented by touch-based graphical items so that patients would be able to signal the location of their symptoms [66]. Alternatively, it is possible that rule-based decision trees could be adequate for certain conditions in certain clinical settings. In the emergency department setting in particular, patients can present with any complaint, which makes flexibility essential for any HPI tool.

Limitations

There are several limitations to this study. First, although we developed our search terms in consultation with a medical librarian and among a multidisciplinary research team, it is possible that the failure to include certain search terms or databases could have led us to miss key publications. Second, tools such as those we have reviewed tend to evolve over time, which may have limited our ability to characterize them accurately based on the published literature. Third, and last, our review of tools included several sources of information such as peer-reviewed literature, non–peer-reviewed articles, and websites of for-profit entities. Although we have tried to verify the accuracy of the information we obtained, it is possible that information from some sources is biased.

Conclusions

Many HPI tools with various features are available to aid in obtaining the HPI. Future research should examine which tools improve patients' health outcomes and which design features are pivotal to improving communication, diagnosis, and,

XSL•FO
RenderX

subsequently, patient health outcomes. We recommend that future tools use a combination of narrative text, closed-ended questions, and graphical items so that the histories obtained can successfully communicate the patients' symptoms to clinicians and narrow the differential diagnosis.

Acknowledgments

This research was supported by internal funding from Cedars-Sinai Medical Center. The funding source did not play any role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The authors thank Caroline Marshall, MLS, AHIP, for assisting with development of the search strategy.

Authors' Contributions

CTB conceptualized the study. CTB and AJH conducted the literature search and data analysis and drafted the article. CTB is responsible for the integrity of the work. All authors participated in writing and revising the article. All aspects of the study (including design; collection, analysis, and interpretation of data; writing of the report; and decision to publish) were led by the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 13 KB - jmir_v24i11e36074_app1.docx]

Multimedia Appendix 2
List of synonyms for digital tools designed to obtain the history of present illness from patients.
[DOCX File , 14 KB - jmir_v24i11e36074_app2.docx]

References

1. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med J 1975 May 31;2(5969):486-489 [FREE Full text] [doi: 10.1136/bmj.2.5969.486] [Medline: 1148666]
2. Slade D, Chandler E, McGregor J, Scheeres H, Chandler E, Stein-Parbury J, et al. Communicating in Hospital Emergency Departments. Heidelberg, Germany: Springer; 2015.
3. Brodman K, Erdmann Jr AJ, Lorge I, Wolff HG, Broadbent TH. The Cornell medical index: an adjunct to medical interview. J Am Med Assoc 1949 Jun 11;140(6):530-534. [doi: 10.1001/jama.1949.02900410026007] [Medline: 18144531]
4. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2016 May 20;Suppl 1:S48-S61 [FREE Full text] [doi: 10.15265/IYS-2016-s006] [Medline: 27199197]
5. Ajami S, Bagheri-Tadi T. Barriers for adopting Electronic Health Records (EHRs) by physicians. Acta Inform Med 2013;21(2):129-134 [FREE Full text] [doi: 10.5455/aim.2013.21.129-134] [Medline: 24058254]
6. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan 21;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
7. Fox S, Duggan M. Mobile Health 2012: half of smartphone owners use their devices to get health informationone-fifth of smartphone owners have health apps. Pew Research Center and American Life Project. 2017. URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/webcite [accessed 2022-11-02]
8. Mobile fact sheet. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-12-30]
9. Patient-Generated Health Data. HelathIT.gov. URL: https://www.healthit.gov/topic/scientific-initiatives/patient-generated-health-data [accessed 2021-12-30]
10. Deering M. Issue Brief: Patient-Generated Health DataHealth IT. The Office of the National Coordinator for Health Information Technology. 2013 Dec 20. URL: https://www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf [accessed 2021-12-30]
11. Ryan GW, Bernard HR. Techniques to identify themes. Field Methods 2003 Feb;15(1):85-109. [doi: 10.1177/1525822x02239569]
12. 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]

XSL•FO
RenderX

13. 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]
14. Bachman J. Improving care with an automated patient history. Fam Pract Manag 2007;14(7):39-43 [FREE Full text] [Medline: 17696057]
15. Wenner AR, Ferrante M, Belser D. Instant medical history. Proc Annu Symp Comput Appl Med Care 1994:1036 [FREE Full text] [Medline: 7949874]
16. Instant medical history homepage. Instant Medical History. URL: https://www.interviewsoftware.com/instant-medical-history-imh [accessed 2021-12-30]
17. Haug PJ, Warner HR, Clayton PD, Schmidt CD, Pearl JE, Farney RJ, et al. A decision-driven system to collect the patient history. Comput Biomed Res 1987 Apr;20(2):193-207. [doi: 10.1016/0010-4809(87)90045-0] [Medline: 3595100]
18. Haug P, Rowe K, Rich T, Famey R, Pearl J, Schmidt CD. A comparison of computer-administered histories. In: Proceedings of the American Association of Medical System Informatics. 1988 Presented at: AAMSI '88; 1988; New York, NY, USA.
19. Quaak MJ, van der Voort PJ, van Bemmel JH. AIDA for the automation of the patient history. Comput Methods Programs Biomed 1987;25(3):297-304. [doi: 10.1016/0169-2607(87)90087-3] [Medline: 3427941]
20. Quaak MJ, Westerman R, Schouten JA, Hasman A, Van Bemmel JH. Appraisal of computerized medical histories: comparisons between computerized and conventional records. Comput Biomed Res 1986 Dec;19(6):551-564. [doi: 10.1016/0010-4809(86)90029-7]
21. Porter SC. Data acquisition in emergency medicine: electronic communication using free text. Pediatr Emerg Care 2002 Feb;18(1):15-18. [doi: 10.1097/00006565-200202000-00005] [Medline: 11862131]
22. Porter SC, Forbes P, Manzi S, Kalish LA. Patients providing the answers: narrowing the gap in data quality for emergency care. Qual Saf Health Care 2010 Oct;19(5):e34. [doi: 10.1136/qshc.2009.032540] [Medline: 20511242]
23. Rosenman SJ, Levings CT, Korten AE. Clinical utility and patient acceptance of the computerized Composite International Diagnostic Interview. Psychiatr Serv 1997 Jun;48(6):815-820. [doi: 10.1176/ps.48.6.815] [Medline: 9175192]
24. Robins LW, Wittchen H, Helzer J, Babor T, Burke J, Farmer A, et al. Composite international diagnostic interview. APA PsycTests. 1988. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft02121-000 [accessed 2021-12-30]
25. The World Health Organization World Mental Health Composite International Diagnostic Interview (WHO WMH-CIDI). Harvard University Department of Health Care Policy. 2017. URL: https://www.hcp.med.harvard.edu/wmhcidi/about-the-who-wmh-cidi/ [accessed 2021-12-30]
26. Download the WHO WMH-CIDI Instruments. Harvard University Department of Health Care Policy. 2017. URL: https://www.hcp.med.harvard.edu/wmhcidi/download-the-who-wmh-cidi-instruments/ [accessed 2021-12-30]
27. MEDOCTOR homepage. MEDOCTOR. URL: https://www.medoctor.com/ [accessed 2021-12-30]
28. MEDoctor - Immediate Doctor's Attention. YouTube. 2018 Mar 7. URL: https://youtu.be/SjOvqx-cQI8 [accessed 2021-12-30]
29. Zakim D, Braun N, Fritz P, Alscher MD. Underutilization of information and knowledge in everyday medical practice: evaluation of a computer-based solution. BMC Med Inform Decis Mak 2008 Nov 05;8:50 [FREE Full text] [doi: 10.1186/1472-6947-8-50] [Medline: 18983684]
30. Brandberg H, Kahan T, Spaak J, Sundberg K, Koch S, Adeli A, et al. A prospective cohort study of self-reported computerised medical history taking for acute chest pain: protocol of the CLEOS-Chest Pain Danderyd Study (CLEOS-CPDS). BMJ Open 2020 Jan 21;10(1):e031871 [FREE Full text] [doi: 10.1136/bmjopen-2019-031871] [Medline: 31969363]
31. Mediktor. Teckel Medical. URL: https://www.mediktor.com/en [accessed 2021-12-30]
32. Teckel Medical. Mediktor - Symptom Checker. App Store. URL: https://apps.apple.com/us/app/mediktor-symptom-checker/id725611856 [accessed 2021-12-30]
33. Teckel Medical. Mediktor. Google Play. URL: https://play.google.com/store/apps/details?id=com.teckelmedical.mediktor&hl=en_US [accessed 2021-12-30]
34. Teckel Solutions. Mediktor. Amazon. URL: https://www.amazon.com/Teckel-Solutions-S-L-Mediktor/dp/B075XRN21X [accessed 2021-12-30]
35. Mediktor. Telegram. URL: https://t.me/Mediktor_bot [accessed 2021-12-30]
36. Moreno Barriga E, Pueyo Ferrer I, Sánchez Sánchez M, Martín Baranera M, Masip Utset J. [A new artificial intelligence tool for assessing symptoms in patients seeking emergency department care: the Mediktor application]. Emergencias 2017;29(6):391-396 [FREE Full text] [Medline: 29188913]
37. DocResponse About Us. DocResponse. URL: https://www.docresponse.com/about/ [accessed 2021-12-30]
38. DocResponse - Medical digital check-in made simple. YouTube. 2019 Sep 27. URL: https://www.youtube.com/watch?v=qj4yvf_ETZQ [accessed 2021-12-30]
39. Digital patient check-in. DocResponse. URL: https://www.docresponse.com/digital-check-in/ [accessed 2021-12-30]
40. The athenahealth Marketplace. Athena Health. URL: https://marketplace.athenahealth.com/product/docresponse-inc [accessed 2021-12-30]
41. Herrick DB, Nakhasi A, Nelson B, Rice S, Abbott PA, Saber Tehrani AS, et al. Usability characteristics of self-administered computer-assisted interviewing in the emergency department: factors affecting ease of use, efficiency, and entry error. Appl Clin Inform 2013;4(2):276-292 [FREE Full text] [doi: 10.4338/ACI-2012-09-RA-0034] [Medline: 23874364]

XSL•FO
RenderX

42. Arora S, Goldberg AD, Menchine M. Patient impression and satisfaction of a self-administered, automated medical history-taking device in the Emergency Department. West J Emerg Med 2014 Feb;15(1):35-40 [FREE Full text] [doi: 10.5811/westjem.2013.2.11498] [Medline: 24695871]
43. OpenNotes. OpenNotes. URL: https://www.opennotes.org/ [accessed 2021-12-30]
44. Mafi JN, Gerard M, Chimowitz H, Anselmo M, Delbanco T, Walker J. Patients contributing to their doctors' notes: insights from expert interviews. Ann Intern Med 2018 Feb 20;168(4):302-305 [FREE Full text] [doi: 10.7326/M17-0583] [Medline: 29132154]
45. Kriegel G, Bell S, Delbanco T, Walker J. Covid-19 as innovation accelerator: cogenerating telemedicine visit notes with patients. NEJM Catalyst. 2022 May 12. URL: https://tinyurl.com/5h7y6d72 [accessed 2022-11-02]
46. FirstHx. URL: https://www.firsthx.com/ [accessed 2021-12-30]
47. Benaroia M, Elinson R, Zarnke K. Patient-directed intelligent and interactive computer medical history-gathering systems: a utility and feasibility study in the emergency department. Int J Med Inform 2007 Apr;76(4):283-288. [doi: 10.1016/j.ijmedinf.2006.01.006] [Medline: 16473548]
48. Almario CV, Chey W, Kaung A, Whitman C, Fuller G, Reid M, et al. Computer-generated vs. physician-documented history of present illness (HPI): results of a blinded comparison. Am J Gastroenterol 2015 Jan;110(1):170-179 [FREE Full text] [doi: 10.1038/ajg.2014.356] [Medline: 25461620]
49. Almario CV, Chey WD, Iriana S, Dailey F, Robbins K, Patel AV, et al. Computer versus physician identification of gastrointestinal alarm features. Int J Med Inform 2015 Dec;84(12):1111-1117 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.07.006] [Medline: 26254875]
50. Furaijat G, Kleinert E, Simmenroth A, Müller F. Implementing a digital communication assistance tool to collect the medical history of refugee patients: DICTUM Friedland - an action-oriented mixed methods study protocol. BMC Health Serv Res 2019 Feb 06;19(1):103 [FREE Full text] [doi: 10.1186/s12913-019-3928-1] [Medline: 30728030]
51. Kerby LR. Founders. In: Saving History: How White Evangelicals Tour the Nation’s Capital and Redeem a Christian America. Chapel Hill, NC, USA: North Carolina Scholarship Online; Apr 2020.
52. Kleinert E, Müller F, Kruse S, Furaijat G, Simmenroth A. [Usability and efficiency of a digital communication assistance tool to obtain medical history from non-German-speaking patients]. Gesundheitswesen 2021 Jul;83(7):531-537. [doi: 10.1055/a-1144-2848] [Medline: 32413911]
53. Müller F, Chandra S, Furaijat G, Kruse S, Waligorski A, Simmenroth A, et al. A digital communication assistance tool (DCAT) toobtain medical history from foreign-languagepatients: development and pilot testing in a primaryhealth care center for refugees. Int J Environ Res Public Health 2020 Feb 20;17(4):1368 [FREE Full text] [doi: 10.3390/ijerph17041368] [Medline: 32093286]
54. Unravel the mystery of Health Data. Medius Health. URL: https://mediushealth.org/ [accessed 2021-12-30]
55. Medius Health. Quro Dictionary. URL: https://www.quro.ai/dictionary/ [accessed 2021-12-30]
56. Ghosh S, Bhatia S, Bhatia A. Quro: facilitating user symptom check using a personalised chatbot-oriented dialogue system. Stud Health Technol Inform 2018;252:51-56. [Medline: 30040682]
57. Ni L, Lu C, Liu N, Liu J. MANDY: towards a smart primary care chatbot application. In: Proceedings of the 18th International Symposium on Knowledge and Systems Sciences. Singapore, Singapore: Springer; 2017 Presented at: KSS '17; November 17–19, 2017; Bangkok, Thailand p. 38-52. [doi: 10.1007/978-981-10-6989-5_4]
58. Schwitzguebel AJ, Jeckelmann C, Gavinio R, Levallois C, Benaïm C, Spechbach H. Differential diagnosis assessment in ambulatory care with an automated medical history-taking device: pilot randomized controlled trial. JMIR Med Inform 2019 Nov 04;7(4):e14044 [FREE Full text] [doi: 10.2196/14044] [Medline: 31682590]
59. Kleinert E, Müller F, Kruse S, Furaijat G, Simmenroth A. [Usability and efficiency of a digital communication assistance tool to obtain medical history from non-German-speaking patients]. Gesundheitswesen 2021 Jul;83(7):531-537. [doi: 10.1055/a-1144-2848] [Medline: 32413911]
60. DesRoches CM, Leveille S, Bell SK, Dong ZJ, Elmore JG, Fernandez L, et al. The views and experiences of clinicians sharing medical record notes with patients. JAMA Netw Open 2020 Mar 02;3(3):e201753 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.1753] [Medline: 32219406]
61. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
62. Levine DM, Co Z, Newmark LP, Groisser AR, Holmgren AJ, Haas JS, et al. Design and testing of a mobile health application rating tool. NPJ Digit Med 2020;3:74 [FREE Full text] [doi: 10.1038/s41746-020-0268-9] [Medline: 32509971]
63. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digit Health 2019 Jun;1(2):e52-e54 [FREE Full text] [doi: 10.1016/S2589-7500(19)30013-5] [Medline: 33323229]
64. Bickley L, Szilagyi P, Bates B. Bates' Guide to Physical Examination and History-Taking. 11th edition. Philadelphia, PA, USA: Lippincott Williams & Wilkins; 2013.
65. Nair N, Narayanan S, Achan P, Soman KP. Clinical note section identification using transfer learning. In: Proceedings of 6th International Congress on Information and Communication Technology. Singapore, Singapore: Springer; 2021 Presented at: ICICT '21; February 25–26, 2021; London, UK p. 533-542. [doi: 10.1007/978-981-16-2377-6_50]

XSL•FO
RenderX

66. Calisto F, Ferreira A, Nascimento J, Gonçalves D. Towards touch-based medical image diagnosis annotation. In: Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces. 2017 Presented at: ISS '17; October 17-20, 2017; Brighton, UK. [doi: 10.1145/3132272.3134111]

Abbreviations

AEGIS: Automated Evaluation of Gastrointestinal Symptoms
CLEOS: Clinical Expert Operating System
DIAANA AMHTD: Diagnosis and Anamnesis Automated Medical History–Taking Device
EHR: electronic health record
HPI: history of present illness
PGHD: patient-generated health data
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

Edited by G Eysenbach; submitted 31.12.21; peer-reviewed by A Joseph, SS Amritphale, EM Noack, D Schlossman, FM Calisto; comments to author 23.02.22; revised version received 25.05.22; accepted 19.10.22; published 17.11.22.

Please cite as:
Berdahl CT, Henreid AJ, Pevnick JM, Zheng K, Nuckols TK
Digital Tools Designed to Obtain the History of Present Illness From Patients: Scoping Review
J Med Internet Res 2022;24(11):e36074
URL: https://www.jmir.org/2022/11/e36074
doi:10.2196/36074
PMID:36394945

©Carl T Berdahl, Andrew J Henreid, Joshua M Pevnick, Kai Zheng, Teryl K Nuckols. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Chatbot to Support Young People During the COVID-19 Pandemic in New Zealand: Evaluation of the Real-World Rollout of an Open Trial

Nicola Ludin[1], BSc, MSc, PhD; Chester Holt-Quick[1,2], BSc, MBChB; Sarah Hopkins[1], BSc, MSc, PhD; Karolina Stasiak[1], MA, PhD; Sarah Hetrick[1,3], BA, MA, PGCertHealSc, DPsych; Jim Warren[1,2], BSc, PhD; Tania Cargo[1], MEd, PGDip (Clinical Psychology), PGDip (Infant Mental Health), PhD

[1]Department of Psychological Medicine, The University of Auckland, Auckland, New Zealand

[2]School of Computer Science, The University of Auckland, Auckland, New Zealand

[3]Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Jim Warren, BSc, PhD
School of Computer Science
The University of Auckland
University of Auckland Private Bag 92019
Auckland, 1142
New Zealand
Phone: 64 21 739 632
Email: jim@cs.auckland.ac.nz

Abstract

Background: The number of young people in New Zealand (Aotearoa) who experience mental health challenges is increasing. As those in Aotearoa went into the initial COVID-19 lockdown, an ongoing digital mental health project was adapted and underwent rapid content authoring to create the Aroha chatbot. This dynamic digital support was designed with and for young people to help manage pandemic-related worry.

Objective: Aroha was developed to provide practical evidence-based tools for anxiety management using cognitive behavioral therapy and positive psychology. The chatbot included practical ideas to maintain social and cultural connection, and to stay active and well.

Methods: Stay-at-home orders under Aotearoa's lockdown commenced on March 20, 2020. By leveraging previously developed chatbot technology and broader existing online trial infrastructure, the Aroha chatbot was launched promptly on April 7, 2020. Dissemination of the chatbot for an open trial was via a URL, and feedback on the experience of the lockdown and the experience of Aroha was gathered via online questionnaires and a focus group, and from community members.

Results: In the 2 weeks following the launch of the chatbot, there were 393 registrations, and 238 users logged into the chatbot, of whom 127 were in the target age range (13-24 years). Feedback guided iterative and responsive content authoring to suit the dynamic situation and motivated engineering to dynamically detect and react to a range of conversational intents.

Conclusions: The experience of the implementation of the Aroha chatbot highlights the feasibility of providing timely event-specific digital mental health support and the technology requirements for a flexible and enabling chatbot architectural framework.

(J Med Internet Res 2022;24(11):e38743) doi:10.2196/38743

KEYWORDS

COVID-19; youth; chatbots; adolescent mental health; dialog-based intervention; digital mental health

XSL•FO
RenderX

Introduction

Background

Looking after young people's mental health is a global public health priority [1]. In New Zealand (Aotearoa), the rates of mental health challenges are increasing among rangatahi (young people) [2]. There are a range of supports being implemented nationally at present to transform care for young people and improve mental health. The Aroha chatbot is one such support tool.

The COVID-19 global pandemic has created trying circumstances at the population level around the globe because of the unprecedented changes and ongoing uncertainty about the future [3]. The COVID-19 pandemic has ongoing impacts on young people in terms of the transition from education to vocation [4-6], and in the context of developmental changes and emergence to adulthood [1].

Before the onset of the pandemic, it was estimated that between 10% and 20% of children and young people experience mental health challenges. This represents 60% to 70% of disability-adjusted life years among young people [7]. Despite many young people needing support and most not seeking or receiving any mental health care, public mental health services are organized in a way that can create significant barriers for those who do seek care. This includes limited capacity, even for those with more severe needs, lack of convenience or visibility, and cost. Other barriers include services being unacceptable to young people with fears such as lack of confidentiality, lack of privacy, and stigma. There is also a lack of equity of access to services [8].

There is now a large and growing body of research on the use of digital mental health support for young people. Digital tools have the benefits of being nonjudgemental, private, stigma free, flexible, and accessible, and have far greater reach than traditional forms of treatment. Digital mental health tools have been shown to be acceptable and effective treatments for the common mental health problems of anxiety and depression in youth [9,10].

Conversational Agents

A dialog agent or "chatbot" style interaction for digital mental health has attracted interest since Eliza in the 1960s [11], even though its imitation of a psychotherapist through the simple linguistic token manipulation that was possible at the time was limited. A contemporary example of a conversational agent is Woebot, which delivers cognitive behavioral therapy (CBT). When tested with students with depression, those who used Woebot significantly reduced their symptoms over the study period. There was no reduction in symptoms for those in the information control group, who were offered a self-help book [12]. A systematic review of conversational agents in health care found mental health to be the most common area of application [13].

COVID-19 Pandemic in Aotearoa

On March 21, 2020, at the beginning of the COVID-19 pandemic in Aotearoa, a 4-level COVID-19 alert system was announced [14]. Beginning at 11:59 PM on March 25, 2020, alert level 4 was instituted, putting the country into a nationwide lockdown with strict stay-at-home orders. This lockdown remained in place until 11:59 PM on April 27, 2020, and then was withdrawn in stages to the lowest alert level on June 8, 2020. With the perception that young people would benefit from and enjoy tailored digital support, we saw an opportunity to leverage existing digital infrastructure to guide rangatahi in Aotearoa through the stringent requirements of lockdown.

This period represented an opportunity for the uptake and trial of digital mental health technologies, and we used the Aroha chatbot. There was an apparent requirement for support other than face-to-face support, as community access was largely suspended and demand for support increased during this time [15].

Responsiveness to Tāngata Whenua

Our research is grounded in Te Tiriti O Waitangi (the Treaty of Waitangi). Māori as Tāngata Whenua (the indigenous people of Aotearoa) have their indigenous status supported through government legislation in Te Tiriti O Waitangi, which guarantees partnership, participation, and protection for Māori. These principles are critical as we try to address the inequities that exist for Māori in a range of health outcomes in Aotearoa [16-18].

The HABITs (Health Advances through Behavioural Intervention Technologies) project has been developing an ecosystem of screening and e-therapy tools designed to meet the needs of young people in Aotearoa since 2016. The Aroha chatbot was developed out of the HABITs project, and the project was greatly inspired by the effectiveness and approach of SPARX (Smart, Positive, Active, Realistic, X-factor thoughts), which is an online game [19]. It helps young people who are feeling down (depressed, stressed, anxious, and low). The SPARX program's development included a Māori cocreator, input from Māori CBT experts, cultural guidance from kaumātua (respected elder), and a Māori game development company [20]. The Māori and Pākehā (New Zealander of European descent) co-leadership in the HABITs project is crucial. It has ensured that as a design team, we are transformational and support the indigenization of mental health services for Māori [21].

Working biculturally (Māori and Pākehā) guaranteed not only relevant content, but also a respectful process in the co-creation of the Aroha chatbot. For this chatbot to connect with rangatahi Māori, we took guidance from our Māori advisors and chose Aroha as a name. Aroha means caring and kind; this allowed for Aro (meaning focus) and Hā (essence) to be included in the intent of the application, which was to provide a "caring and kind" e-therapy support tool that "focuses" on a person's "essence." In addition, strategies and activities were designed for and targeted rangatahi Māori and their whānau (family).

Objectives

The objective of this study was to support young people in Aotearoa in managing challenges during the COVID-19 lockdown. We sought to develop chatbot architecture by leveraging existing technology, determine the process and

XSL•FO
RenderX

experience of developing the Aroha chatbot, and evaluate the real-world rollout of an open trial.

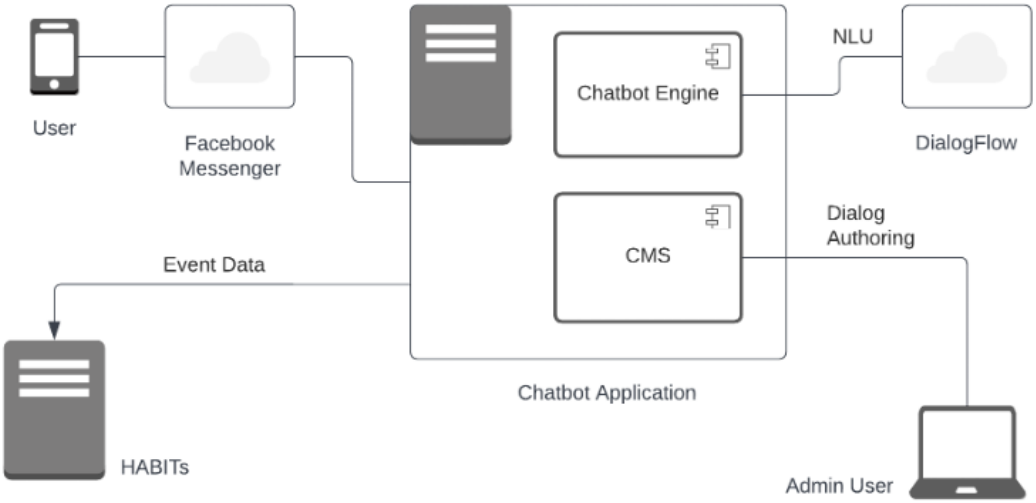
Methods

Initial Architecture

Existing chatbot technology and broader online trial infrastructure already developed for the HABITs project were leveraged to develop the Aroha chatbot. The base chatbot technology for content management and delivery in the HABITs project was developed in partnership with a contractor, RUSH Digital [22]. This technology had already been used for chatbot deployments called "Headstrong" and "Stress-Detox." Each of these were designed to promote resilience with methods grounded in CBT and positive psychology, and field trials have been conducted for each [22,23].

A high-level overview of the chatbot architecture is depicted in Figure 1. The main components are the messaging channel, chatbot engine, natural language understanding module, and dialog content authoring system (content management system [CMS]). The web application framework Django was used for the application technology, in combination with Facebook Messenger as a client.

Figure 1. Simplified hybrid component architecture diagram of the system used for Aroha. CMS: content management system; HABITs: Health Advances through Behavioural Intervention Technologies; NLU: natural language understanding.



The messaging channel selected for Aroha was Facebook Messenger, the sole channel initially supported by the base technology. Facebook Messenger (now Messenger by Meta Platforms) supports sending text with or without quick reply options; a variety of attachment types including simple and carousel images, GIFs, audio, and video; and buttons and webviews (web pages that display within Messenger). For instance, webviews were used in Aroha for a set of games called "Swipe Sports."

The chatbot engine handled the user input/chatbot response conversational exchange. User input in the messaging channel invokes a request to the chatbot server, where chat requests are handled by the chatbot engine, dispatching one or more responses back to the user. In this fashion, the user has the experience of chatting with a persona (in this case Aroha) on Messenger. Aroha appears alongside contacts the user has already chatted with.

The chatbot dialog has a directed graph representation wherein vertices (or "nodes") of various types hold the dialog content and other aspects of the chat logic, while edges (arcs between nodes) specify the available transitions. When a request is handled by the chatbot engine, this directed graph is traversed from the user's current vertex until a vertex is reached that requires user input again. This procedure corresponds to a dialog turn. There is a range of vertex types in the dialog graph. Some vertices when visited invoke sending a message (text, attachment, etc) to the user, while others update a variable or assess a condition (and select an out-edge to traverse) based on user input or the value of a variable. It is common for the chatbot to return 2 or 3 responses to the user, which is an important way to break up longer pieces of dialog. A conversational humanness is conferred by a short thread sleep between dispatched responses as well as a "typing on" response to the user, which shows the chatbot is creating another message and appears as a dialog bubble with 3 moving dots.

The natural language understanding component of the system includes functionality for free-text intent classification. However, there are few locations within the dialog graph in which the chatbot encourages input of free text, with the dialog principally mediated by the provision of quick options for the user to choose from. When input is made by the user, free text is classified to 1 member of a predefined intent set through an application programming interface (API) request to DialogFlow [24]. Forty distinct intents were defined at a level sufficiently granular to discern expressions representing main mood/emotion states in a meaningful way. For example, intents were defined for anger, sadness, loneliness, stress, fear, anxiety, happiness, and excitement. At any vertex within the dialog graph where

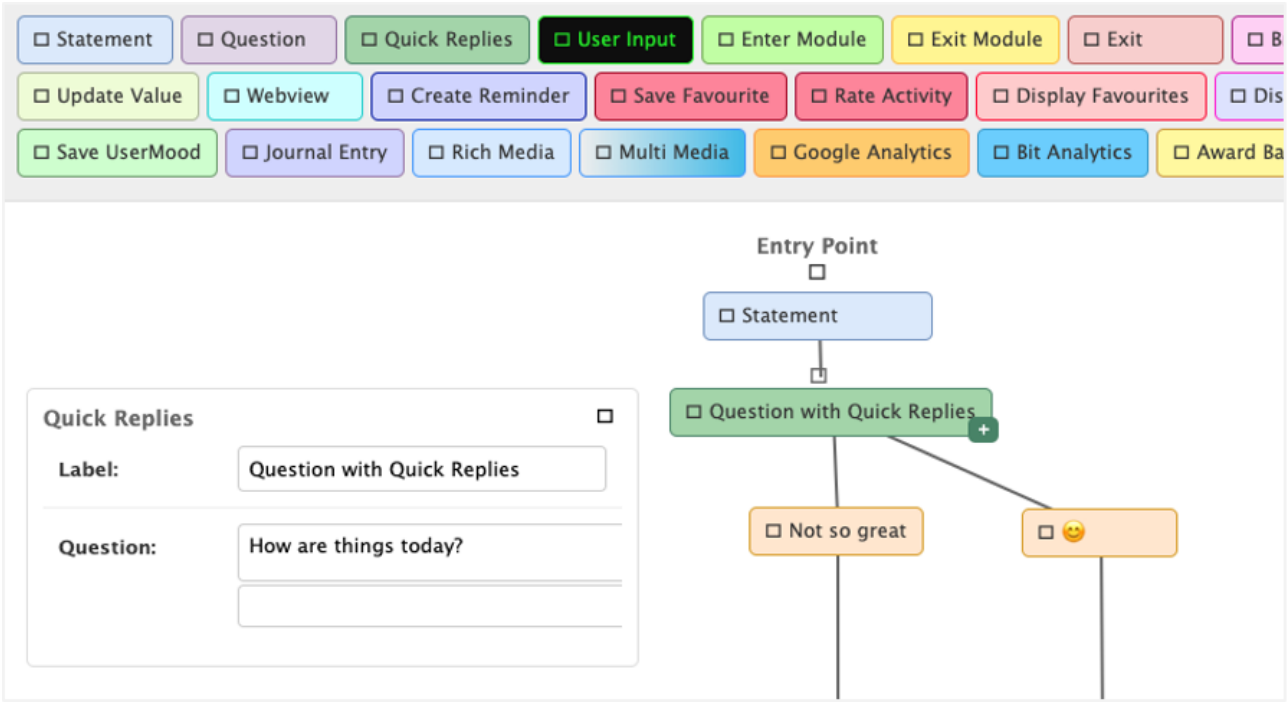
XSL•FO
RenderX

free text is handled, there are outgoing edges corresponding to specified intents and 1 fallback edge for free text that is unmapped. Multiple intents with similar characteristics will sometimes converge on the same child node when that node is sufficiently general to handle a group of intents (eg, fear, stress, and anxiety). For example, if a user typed "I'm sad" or "I'm down," the chatbot responded with "Hey, I am sorry to hear you are feeling that way," and then offered other resources.

Dialog authoring is undertaken using a web browser–based graphical user interface (GUI). The overall dialog has 1 entry point and is organized into dialog modules, such that a user moves through a sequence of modules, some of which are conditionally served. Given that the dialog is large, this modularization is particularly important to support nontechnical authors.

Figure 2 illustrates a portion of the user interface for dialog module authoring. Each dialog module has 1 entry point and 1 or more exit points (yellow "Exit Module" nodes). Authoring the dialog in each module involves defining both the sequence and content of dialog nodes. To define the sequence, the GUI canvas allows nodes to be dragged from the top toolbar into the canvas, and each out-edge must then connect (using a responsive snaplock) to a child node. Certain node types naturally allow for multiple out-edges. For example, a purple "Question" node receives free-text user input, and thus, the out-edges of this node type are associated with specific intents that free text is expected to be classified to (configured via a dropdown option available to this node type when clicked). The green "Quick Replies" node type also has multiple out-edges corresponding to a set of quick options a user may select from (beige colored labels to out-edges). There are also condition checking nodes (pink, "Branch") that are transparent to the user but control the flow of the dialog using variables that can be set throughout the dialog (lime "Update Value" nodes).

Figure 2. The dialog module authoring canvas of the chatbot architecture used for Aroha.



The content of each node depends on the node type. Figure 2 illustrates editing of the content of a "Quick Replies" node. The content of this node type includes both the question presented to the user and a set of 1 or more quick reply options that the user may select from. "Rich Media" (light blue) nodes define an attachment response with an image or audio file, and thus, configuring content for this node type involves uploading the associated file type.

This authoring interface affords the opportunity for users without technical expertise to configure the dialog, and this helps to address the problem of the knowledge engineering bottleneck. Our experience with both Headstrong and Stress-Detox was that users without technical expertise can author the majority of dialogs for chatbots using this authoring system. For Aroha, clinical psychologists authored dialog modules directly.

The chatbot architecture was designed to interoperate with the HABITs information technology platform. To facilitate rapid and concurrent field trials within the HABITs project, a configurable web portal and suite of web services have been developed [25,26]. The portal supports online self-registration of users to trials with tailored project information and informed consent, as well as administration of online assessments at specified times. The web services allow applications to log event data (eg, sessions of use and activities completed) and link these with assessments for subsequent analysis. These features were used to set up the online trial of Aroha.

Aroha Chatbot Development for Initial Rollout

In the context of a strict lockdown (all schools and nonessential workplaces closed, and the public instructed to stay at home and have contact only with those in their "bubble"), the expectation was that young people would be using Aroha at

XSL•FO
RenderX

home, generally on their phones. The target age range for the chatbot users was 13 to 24 years.

Aroha was designed to have a more user-controlled flow than our previous chatbots that had a programmed set of day-by-day activities. After initial rapport building, including brief assessment and empathetic feedback, as well as psychoeducation around managing pandemic-related anxiety, the user was offered the option to select a module to suit them. Table 1 lists the modules in the initial release of Aroha and those added shortly after launch.

Table 1. Aroha chatbot modules.

Variable	Modules
First session/onboarding	• Intro • Onboarding • Brief assessment • Introductory information (modified at re-entry to just a greeting and repeat of the brief assessment)
Activity modules	• Stay connected • Calming activities • Practice gratitude • Spirituality • Distract yourself • Get active • Get expert help • General tips 1. Self-care 2. Have a routine 3. Protect your sleep • Alcohol and drugs
Activity modules added within 3 months after launch	• Money worries • Anger management • Violence • Prime Minister's message
Outro	• Recheck of the brief assessment • Give feedback • Check your hauora (health and well-being) • Other resources • Outro

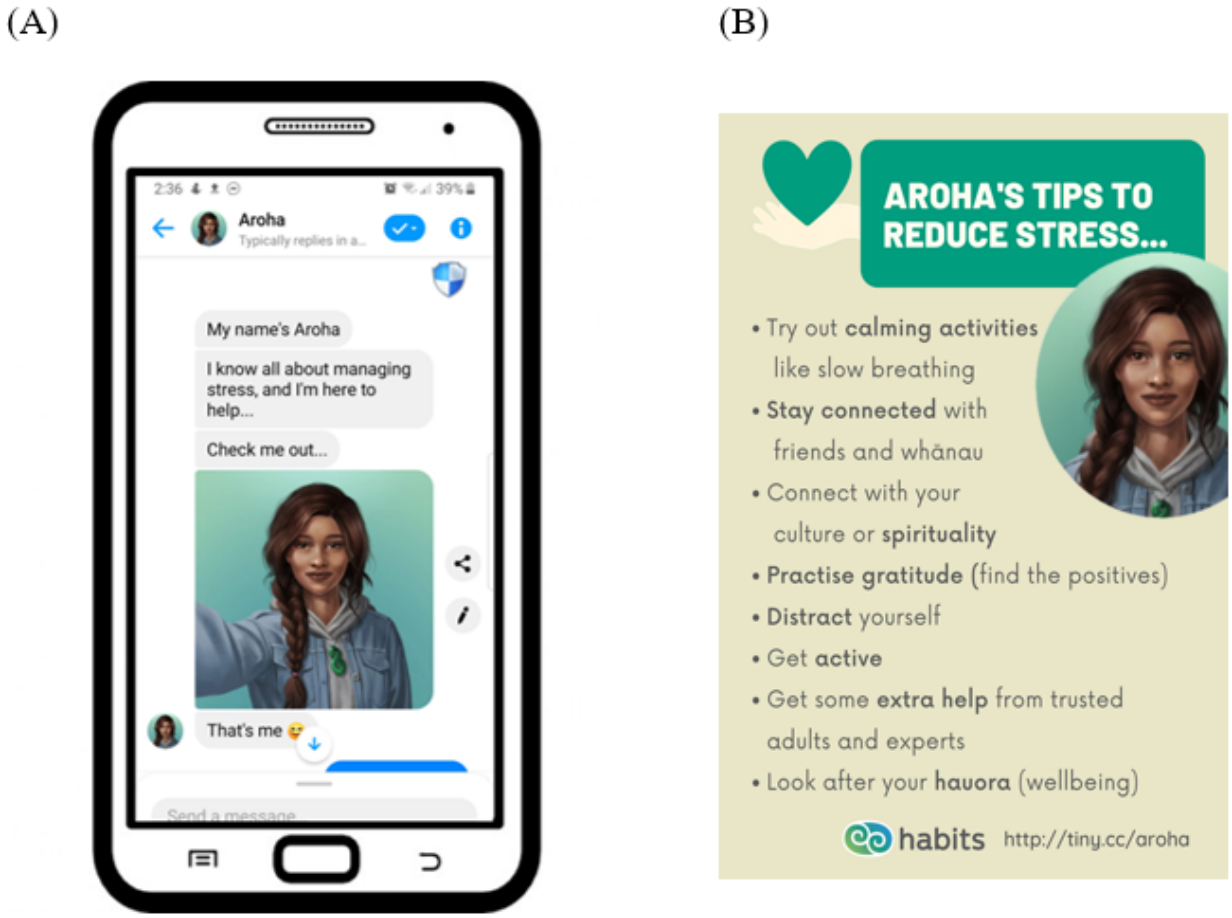
Activity modules within Aroha were short. These modules teach evidence-based mental health strategies for good well-being that are based on best practice. Activities were designed to support young people to maintain general well-being within the context of the COVID-19 pandemic. The flow through the modules is largely driven by the user, and modules can be repeated, although users are encouraged to try another skill they might not have tried previously.

For example, the "practice gratitude" module promotes a grateful mindset and orients thinking toward positive things in life. The users are offered examples of things to be grateful for, such as whānau and technology, and then encouraged to make their own entry. Clinical trials show that practicing gratitude promotes happiness and well-being, and reduces symptoms of depression and anxiety [27]. Users could easily choose to start another module; otherwise, they were offered other resources upon module completion, as well as the option to finish the chat/exit.

Figure 3 illustrates some content from Aroha. All users had Aroha as a guide (unlike our previous chatbot deployments where there was a choice of 4 personas). Aroha offered an introduction and a "faux selfie" (Figure 3A) and reminded the user that the Aroha character is not a real person. The chat consisted of text and images in the message stream. "Poster" style content was frequently used to provide sets of related tips, such as "Aroha's tips to reduce stress" (Figure 3B). Tips on a single image were able to be structurally linked, unlike a series of consecutive text bubbles. Posters were richly styled to reinforce the tone of the message, and users were offered the option to save posters to review later.

XSL•FO
RenderX

Figure 3. Example Aroha content. (A) Chatbot introduction with faux selfie. (B) "Poster" style advice.



Recruitment Procedure and Feedback

A tiny URL (tiny.cc/aroha) was created that redirected to the HABITs portal for Aroha trial enrollment and was used as the basis for disseminating the chatbot. The URL was used in communications released by the authors' institution and interviews given by the authors, and was sent to contacts in the community inviting feedback and encouraging promotion of the service to young people. While these channels were unlikely to reach many young people directly, the intention was to reach professionals who would endorse the tool to young people, including guidance counsellors, mental health workers, teachers, and school senior management.

Early content generation was based on user testing (held online) with school guidance counsellors, clinicians, and rangatahi. We connected to our network of youth advisory groups and community organizations to recruit participants to a focus group that a clinical psychologist (SH) conducted in June 2020. The focus group was held online with 7 participants and had the specific objective of learning about the issues rangatahi faced with respect to the COVID-19 pandemic and lockdowns. The focus group consisted of introductions and karakia (Māori incantation and blessing), and an explanation of the Aroha chatbot, as well as the following guide questions: (1) What have your experiences and challenges been during COVID-19? (2) What do you expect the next 3 months to be like for you? (3) What are some of the challenges you or your friends might face going forward? (4) What do you expect the next week to be like? (5) What would the next 3 months ideally look like for you? (6) Do you think the challenges you've experienced over the last few months (referring back) will continue? The session was closed with karakia, and the youth voice from this focus group was used to inform the ongoing content iteration to the Aroha chatbot.

We conducted early user testing of Aroha using our existing networks and contacts. As we were working under lockdown conditions, feedback was collected from users remotely (using videoconferencing where possible) and in writing. We approached a small number of adolescents to try Aroha and asked them for their initial impressions, what they liked, and what improvements we could make. We also made Aroha available to an undergraduate class and invited students to send their feedback about the chatbot's usability, style of communication, and features.

Given that the primary motivation for the development of Aroha was to ensure relevant and accessible support to young people, only a brief research assessment was incorporated via a single item assessing anxiety about COVID-19 on a scale with scores ranging from 0 to 10 ("no worries" to "totally freaking out"), collected at onboarding and the introduction of subsequent sessions. Users were invited but not required to record the single item at session outro. Those who used Aroha were also invited by automated email to give feedback about the experience of using Aroha 5 days after onboarding with the chatbot.

XSL•FO
RenderX

Ethics Approval

This study was supported by an amendment to the University of Auckland Human Participants Ethics Committee protocol 023234 (a protocol initially used for Stress-Detox, amended to a lower target age, to provide COVID-19 focused content, and to remove the original study's pre-post surveys). All participants provided informed consent.

Results

Initial Uptake

In the 2 weeks following the launch of the chatbot and the open trial, there were 393 registrations, and 238 users logged into the chatbot, of whom 127 were in the target age range (13-24 years). Moreover, 70.9% (90/127) were female and 47.2% (60/127) identified as New Zealand European. On average, target users engaged with Aroha for 11 minutes, and of the 127 users in the target age range, 31 returned for repeat sessions. There were 30 users (out of 81, 37%) who completed the pre-post measure of COVID-19 anxiety, both in the initial session and the outro, and these showed a reduction in self-reported anxiety from a mean of 5.1 (range 0-10; SD 2.6) to 4.3 (SD 2.5). Further uptake and evaluation data collected following the initial launch of the Aroha chatbot will be shared in a subsequent paper.

Feedback

Feedback from users of Aroha who responded to the invitation sent by automated email has been provided in Textbox 1.

Textbox 1. Feedback quotes from young people in the target age range (13-24 years) who used Aroha.

Question 1: "Did you find me useful?"

Responses:

1. "Yup I did, you gave a lot of good tips forma range of topics and it was rlly useful and good"
2. "Yes"
3. "Yes"
4. "Yeah, can I ask something personal lol"
5. "Yes, I did"
6. "Yes. Very good. Thank you"
7. "Yes, I did"

Question 2: "What was the most useful part?"

Responses:

1. "The range of tips from all the different topics"
2. "Ideas"
3. "The easy feeling of it being a conversation"
4. "Getting calm"
5. "Meditation"
6. "Breathing"
7. "Activities to reduce stress"

Question 3: "What could I do better?"

Responses:

1. "Ummm nothing really"
2. "Not sure"
3. "Some symbols don't display on my phone, can make it hard to know what response I'm giving"
4. "Maybe ask what's wrong first"
5. "I don't know how to save favorites and I can't see my progress"
6. "Nothing"
7. "Help with anxiety"

Feedback highlighted that Aroha was accessible and acceptable. Young people gave positive feedback about the aspects of engagement that were specifically used to encourage uptake. User comments were generally encouraging and highlighted the need for this intervention at the time of global uncertainty and high stress in the community (Textbox 2).

XSL•FO
RenderX

Textbox 2. Feedback (illustrative quotes) from young people and community members on the COVID-19 pandemic lockdown and experience of Aroha.

COVID-19 pandemic and lockdown feedback

"The struggle is real"

"Thought I'd miss my friends but I didn't"

"Hard not to take them home and have a tangi [mourning ceremony]"

"Not being able to grieve in a traditional way affected my mental health"

"I love them to death but they were really annoying and especially with my uni work…having to home school my 4 year old sister…."

"All my 8 siblings being at home"

"Guilt of not using my time in lockdown in the most productive way"

"Will New Zealand get attacked too by the corona virus"

Aroha chatbot feedback

"I found it cool to see messenger used"

"It felt like you were actually having a casual conversation with someone else rather than being lectured to"

"the use of emojis and GIFs helped to increase my engagement"

"I really liked how Aroha used language that I use to communicate; she used words like *heaps* and *whānau* which made it feel like she was down-to-earth and easy to engage with and talk to"

"Having it set within messenger makes it feel natural as if you were talking to one of your friends and it normalizes the idea of talking/reaching out to somebody"

"I liked that you could respond with a pre-determined emoji as it took the pressure off trying to think of a response. She was very clear and easy to understand and the replies were also made very easy"

Community feedback

"This chatbot is neat! Aroha breathed positivity; I can imagine this helping many going through a tough time during this pandemic"

"Having the chatbot sit within Messenger (and existing app that users already had on their phones) made it easy to access and made getting help easier"

"Young people liked the language, use of humor, emojis, GIFs and felt that all those features *humanized* Aroha"

"The dialogue was praised for being realistic and the avatar was well liked, and many people felt a sense of connection to it"

"There were concerns about a limited range of responses, and some wished there was more opportunity for the user to express their emotions through free text"

"I'm not sure how you would develop such a software, but it would also be great to engage with a more empathetic figure, someone that can understand emotive language coming through text language"

Feedback also highlighted important improvements that were needed to increase accessibility and engagement. The themes were consistent across respondents (n=20) and were thematically grouped and included the following: (1) emergent needs as the COVID-19 situation evolved (in Aotearoa, the initial fear of contracting COVID-19 quickly subsided due to strict social restriction measures instituted by the government, and more salient concerns evolved, such as unemployment and resulting poverty, as well as experiencing violence and abuse while in lockdown); (2) improved conversationality; and (3) improved personalization of content. It was further deemed important to adapt the dialog content to reflect the changing nature of the pandemic.

The development effort in response to initial feedback included expanding the set of dialog modules to include content for emergent needs. Modules were developed to address financial stress and domestic violence. Having a large amount of informational content requires presenting this content in diverse ways to foster user engagement. In the months following the initial launch, we expanded on the use of posters (image files) within the messenger channel by developing more dynamic webviews (single webpages using JavaScript for dynamics) to deliver content. This included a public health message from the Prime Minister's Office, and there was iteration through each statement of the message (describing personal top tips for managing stress) in sequence on a touch event.

The need for improved conversationality and personalization was addressed by (1) enlarging the set of defined intents, (2) making some intent areas more fine-grained, (3) enlarging the set of out-edges (locally recognized intents) at free-text question nodes, and (4) implementing "priority intents" (intents that are detected by the dialog agent independent of a user's location within the overall dialog). Prior to implementing priority intents, except for risk phrase detection for statements representing potential self-harm expressions, free-text user input that did not match any intent-labeled out-edge at a free-text or quick reply question node would invoke only a general response to the default fallback out-edge. For example, if a user expressed that they were feeling *nervous* at a dialog node with labeled out-edges for sadness, happiness, or default, then the chatbot would respond only with a statement that was generally applicable such as an emoji "shrug" or "ponder." This impairs

XSL•FO
RenderX

the chatbot's pretense of intelligence and conversationality and thus user experience. "Priority intents" have since been defined that allow the chatbot to detect a set of expressions even if they are not locally configured and thus allow the user to digress elsewhere in the dialog. These priority intents include expressions related to self-harm, risk of abuse, low mood, fear, boredom, and quit/exit. This change shifts the system toward being more responsive to the user's agenda as compared to its preprogrammed agenda. User-driven digressions made possible by priority intents are logged as usage data on the HABITs platform to elucidate the frequency of priority intent activation.

Events following the initial launch have underscored the need for easy content modification that can quickly be rolled through to the production chatbot service. For instance, in August 2020, there was a rapid reintroduction of COVID-19 restrictions in Aotearoa, but this time, there was strong regional variation (with the outbreak being focused in Auckland). Similarly, there were many regional changes during 2021, with varying restriction levels of stay-at-home orders from August 17, 2021, to December 2021 [28]. This required adjustment of both the tone of content and asking about user lockdown conditions, such that advice was relevant to local conditions.

Discussion

Principal Findings

Our experience of developing the Aroha chatbot indicates the feasibility of implementing chatbot-based digital mental health support in response to emerging events. The straightforward authoring that is possible through the content management system and the chatbot architecture means that mental health and well-being support can be tailored to any event, not only that of a global pandemic.

Initial user feedback provided guidance to update content and features as the COVID-19 situation in Aotearoa evolved. In the case of Aroha, we initially used techniques drawn from CBT and positive psychology, and integrated in a bicultural context, which have been shown to be efficacious across many groups and via conversational agents [12,29]. Based on user feedback, we identified the need for additional content, particularly in the areas of distress related to problems of living. Feedback also gave a clear signal that users wanted more dynamic conversationality. Initial efforts had been focused on imparting a large amount of dialog content into Aroha to cover the variety of different stressors that users were facing given the pandemic context. A strength of our chatbot architecture is the ability to easily author large amounts of dialog content, including directly by domain experts. Yet, user feedback established that Aroha required better ability for a user to drive the conversation rather than always follow the chatbot-guided dialog.

In a viewpoint article, McGreevey et al identified a range of considerations for implementing conversational agents in health care [30]. Among the leading considerations were patient safety and trust, and transparency. A key element of Aroha is the use of evidence-based approaches with expert authorship of content. A further element is that the agent persona, although drawn as a relatable human, clearly identifies itself as a computer program. Further, detection of self-harm phrases in user input at any time triggers escalation to confirmation, expression of empathy, direction to a help hotline, and shut down of the computer-based dialog.

Another important consideration is health equity. In Aotearoa, it is crucial for the chatbot to at least be bicultural and show positive effects for Māori. To highlight this and to ensure a culturally responsive product, we developed a chatbot with a Māori persona that had a background story about how it was designed. We also ensured bicultural clinician input and authorship, which ensured that there was relevant Māori content available in the chatbot. While anecdotally it appears that these values were reflected in good reach for rangatahi Māori, further analyses and discussions will be presented in a subsequent paper.

In terms of cybersecurity, extensive user input is not encouraged (which, at any rate, we would not be able to use therapeutically in a safe and trustworthy manner with current technology). Further, specific user input is not logged to our research platform database (only the series of activities and intents to understand usage). However, the conversation log is held with Facebook, and while the project information presented in the user consent process says that the interaction with the chatbot is visible to Facebook and is subject to Facebook's privacy policy, most young people would not read the information in full or entirely appreciate such a notice. In terms of research, development, and innovation considerations, we have framed Aroha as a trial with informed consent, and we believe this is appropriate until the effectiveness of the intervention is better understood.

Limitations

This study has some limitations. We have not yet conducted a randomized controlled trial of Aroha [31]. In fact, the initial rollout has emphasized ease of user experience over data collection. However, the intervention is based on evidence-based therapies, and indeed the context of COVID-19–induced stress is sufficiently dynamic that any trial will have limitations in terms of transferability.

We included a research assessment via a single-item anxiety question (0-10 scale, "no worries" to "totally freaking out"). A single-item measure of anxiety limited our ability to assess severity, although single-item measures have shown reasonable sensitivity and specificity in screening for anxiety in the hospital setting [32]. The advantage is that they are quick and simple to implement in a chatbot application, but further validation of this is necessary.

In the months following launch, we received funding for further structured user engagement (youth focus groups) and implementation of a Te Reo Māori version. The natural language functionality is being enhanced by including free-text named entity recognition. As a longer-term direction, we are exploring how deep learning could be used to create a more dynamic and engaging user experience. The ability to create empathetic dialog and achieve large numbers of conversational turns per session as demonstrated with Xiaoice [33] would be beneficial to our system if only to lead users to a larger "dose" of chatbot-based therapy. However, we would not want to use deep learning in a way that compromises the quality of expert-authored content.

XSL•FO
RenderX

In the first instance, our expanding use of intent recognition provides a compromise that preserves the use of expert-authored content in the chatbot response. A further extension of deep learning that is consistent with our approach would be deep learning of dialog policy (eg, machine learning of what module to choose or recommend next, but where the module content is still expert authored [34]). An additional option is to allow machine-learnt dialog for rapport-building chit-chat and possibly to learn more about the users and their needs, but then fall back to more reliable content for therapeutic advice.

Future Directions

One area of further work that we have pursued is the development of a custom chat app as an alternative to Messenger. There are benefits of using Messenger, such as capitalizing on the fact that young people habitually use social media and many already use Messenger [35,36], and that there is no requirement to download an app to use the chatbot. However, development of a standalone app means that a superior experience can be delivered without limitations on user notifications or logging of conversation details with a third party.

Since its initial launch, the Aroha chatbot has evolved from the initial main objective of supporting young people through COVID-19 lockdown and strict stay-at-home orders. Aroha has developed to include general mental health and well-being support. For example, content for Matariki (Māori new year) and Kohinga Māori activities (a collection) were added to the chatbot, and modules for current events like daylight savings and examinations were included. Supporting mental well-being is a notion in line with Te Kāwanatanga o Aotearoa (the New Zealand government's) Kia Kaha, Kia Māia, Kia Ora Aotearoa, and COVID-19 psychosocial and mental well-being plan. This initiative aims to support individuals, whānau, and communities to respond, recover, adapt, and thrive in the context of the COVID-19 pandemic [16]. Following the Kia Kaha plan (released in 2020), a long-term plan, Kia Manawanui: Long-term pathway to mental well-being, was published in 2021 [17]. Kia Manawanui outlines the transformation of the approach to mental well-being. The Aroha chatbot has and continues to be an imperative component of psychosocial recovery for rangatahi.

Conclusions

We have identified elements of a chatbot architecture sufficient for responsiveness to emerging situations, such as a pandemic lockdown, with easy authoring by domain experts and a rapid deployment channel as cornerstones. In our case, the deployment channel included a configurable portal for web-based trial recruitment linked to chatbot interaction through Facebook Messenger. Further, we found user demand for increased responsiveness to a range of inputs from boredom to fear, as well as a need for additional content. This emphasized the need for a flexible and extensible output, and the ability to easily update the service. Our experience is that an architecture with these elements for creating supportive chatbots has wide application for flexible and rapid responsiveness to other events and situations.

Acknowledgments

We extend our thanks to the study participants. This research was supported by Ministry of Business, Innovation and Employment, A Better Start, E Tipu E Rea (UOAX1511); Ministry of Health COVID-19 National Psychosocial Campaign–University of Auckland–Aroha Chatbot (CC20421); and Ministry of Business, Innovation and Employment, COVID-19 Innovation Acceleration Fund (CIAF 1390 PROP – 71297 – CIAF – UOA). The authors have no commercial interest in RUSH Digital but may be designated as contributors in intellectual property relating to novel dialog components or content described in this paper.

Conflicts of Interest

None declared.

References

1. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. The Lancet 2007 Apr;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7]
2. Fleming T, Tiatia-Seath J, Peiris-John R, Sutcliffe K, Archer D, Bavin L, et al. Youth19 Rangatahi Smart Survey, Initial Findings: Hauora Hinengaro / Emotional and Mental Health. Squarespace. New Zealand: The Youth19 Research Group, The University of Auckland, and Victoria University of Wellington URL: https://static1.squarespace.com/static/5bdbb75ccef37259122e59aa/t/5f338e4cfb539d2246e9e5ce/1597214306382/Youth19+Mental+Health+Report.pdf [accessed 2022-10-10]
3. Wang Y, Kala MP, Jafar TH. Factors associated with psychological distress during the coronavirus disease 2019 (COVID-19) pandemic on the predominantly general population: A systematic review and meta-analysis. PLoS One 2020;15(12):e0244630 [FREE Full text] [doi: 10.1371/journal.pone.0244630] [Medline: 33370404]
4. Joint Statement - COVID-19 impact likely to lead to increased rates of suicide and mental illness. Australian Medical Association. 2020. URL: https://www.ama.com.au/media/joint-statement-covid-19-impact-likely-lead-increased-rates-suicide-and-mental-illness [accessed 2022-10-10]
5. Webb S, Kingstone S, Richardson E, Flett J. Rapid Evidence Brief: COVID-19 Youth Recovery Plan 2020–2022. Wellington: Te Hiringa Hauora/Health Promotion Agency; 2020.

XSL•FO
RenderX

6. Lee J. Mental health effects of school closures during COVID-19. Lancet Child Adolesc Health 2020 Jun;4(6):421 [FREE Full text] [doi: 10.1016/S2352-4642(20)30109-7] [Medline: 32302537]
7. Victorian Burden of Disease Study: Mortality and morbidity in 2001. AUSVEG VIC. URL: http://ausvegvic.com.au/pdf/2001%20Victorian%20Burden%20of%20Disease%20Report_summary.pdf [accessed 2022-10-10]
8. Tylee A, Haller DM, Graham T, Churchill R, Sanci LA. Youth-friendly primary-care services: how are we doing and what more needs to be done? Lancet 2007 May 05;369(9572):1565-1573. [doi: 10.1016/S0140-6736(07)60371-7] [Medline: 17482988]
9. Stasiak K, Fleming T, Lucassen MFG, Shepherd MJ, Whittaker R, Merry SN. Computer-Based and Online Therapy for Depression and Anxiety in Children and Adolescents. J Child Adolesc Psychopharmacol 2016 Apr;26(3):235-245. [doi: 10.1089/cap.2015.0029] [Medline: 26465266]
10. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: A systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
11. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 1966 Jan 21;9(1):36-45. [doi: 10.1145/365153.365168]
12. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
13. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
14. About our COVID-19 response. New Zealand Government. URL: https://covid19.govt.nz/covid-19/restrictions/alert-system-overview/ [accessed 2022-10-10]
15. Pavlova A, Witt K, Scarth B, Fleming T, Kingi-Uluave D, Sharma V, et al. The Use of Helplines and Telehealth Support in Aotearoa/New Zealand During COVID-19 Pandemic Control Measures: A Mixed-Methods Study. Front Psychiatry 2021;12:791209 [FREE Full text] [doi: 10.3389/fpsyt.2021.791209] [Medline: 35153859]
16. Kia Kaha, Kia Māia, Kia Ora Aotearoa: COVID-19 Psychosocial and Mental Wellbeing Plan. Ministry of Health Manatū Hauora. URL: https://www.health.govt.nz/system/files/documents/publications/kia-kaha-kia-maia-kia-ora-aotearoa-covid-19-psychosocial-mental-wellbeing-plan-21dec2020-v2.pdf [accessed 2022-10-10]
17. Kia Manawanui Aotearoa – Long-term pathway to mental wellbeing. Ministry of Health Manatū Hauora. 2021. URL: https://www.health.govt.nz/publication/kia-manawanui-aotearoa-long-term-pathway-mental-wellbeing [accessed 2022-10-10]
18. He Korowai Oranga Māori Health Strategy. Ministry of Health Manatū Hauora. URL: https://www.health.govt.nz/our-work/populations/maori-health/he-korowai-oranga [accessed 2022-10-10]
19. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
20. Shepherd M, Fleming T, Lucassen M, Stasiak K, Lambie I, Merry SN. The design and relevance of a computerized gamified depression therapy program for indigenous māori adolescents. JMIR Serious Games 2015 Mar 03;3(1):e1 [FREE Full text] [doi: 10.2196/games.3804] [Medline: 25736225]
21. Durie M. Indigenizing mental health services: New Zealand experience. Transcult Psychiatry 2011 Apr;48(1-2):24-36. [doi: 10.1177/1363461510383182] [Medline: 21511844]
22. Holt-Quick C, Warren J, Stasiak K, Christie G, Hetrick S, Hopkins S, et al. A chatbot architecture for promoting youth resilience. Studies in Health Technology and Informatics 2021;276:99-105. [doi: 10.3233/SHTI210017]
23. Williams R, Hopkins S, Frampton C, Holt-Quick C, Merry SN, Stasiak K. 21-Day Stress Detox: Open Trial of a Universal Well-Being Chatbot for Young Adults. Social Sciences 2021 Oct 30;10(11):416. [doi: 10.3390/socsci10110416]
24. Dialogflow. Google. URL: https://cloud.google.com/dialogflow/docs [accessed 2022-10-10]
25. Warren J, Hopkins S, Leung A, Hetrick S, Merry S. Building a Digital Platform for Behavioral Intervention Technology Research and Deployment. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: 53rd Hawaii International Conference on System Sciences; January 7-10, 2020; Maui, Hawaii, USA. [doi: 10.24251/hicss.2020.414]
26. Warren J, Tempero E, Warren I, Sathianathan A, Hopkins S, Shepherd M, et al. Experience Building IT Infrastructure for Research with Online Youth Mental Health Tools. 2018 Presented at: 25th Australasian Software Engineering Conference (ASWEC); November 26-30, 2018; Adelaide, SA, Australia p. 161-165. [doi: 10.1109/aswec.2018.00029]
27. Wood AM, Froh JJ, Geraghty AWA. Gratitude and well-being: a review and theoretical integration. Clin Psychol Rev 2010 Nov;30(7):890-905. [doi: 10.1016/j.cpr.2010.03.005] [Medline: 20451313]
28. History of the COVID-19 Alert System. New Zealand Government. URL: https://covid19.govt.nz/about-our-covid-19-response/history-of-the-covid-19-alert-system/#timeline-of-key-events [accessed 2022-10-10]
29. Hollon SD, Stewart MO, Strunk D. Enduring effects for cognitive behavior therapy in the treatment of depression and anxiety. Annu Rev Psychol 2006;57:285-315. [doi: 10.1146/annurev.psych.57.102904.190044] [Medline: 16318597]

XSL•FO
RenderX

30. McGreevey JD, Hanson CW, Koppel R. Clinical, Legal, and Ethical Aspects of Artificial Intelligence-Assisted Conversational Agents in Health Care. JAMA 2020 Aug 11;324(6):552-553. [doi: 10.1001/jama.2020.2724] [Medline: 32706386]
31. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
32. Young Q, Nguyen M, Roth S, Broadberry A, Mackay MH. Single-item measures for depression and anxiety: Validation of the Screening Tool for Psychological Distress in an inpatient cardiology setting. Eur J Cardiovasc Nurs 2015 Dec;14(6):544-551. [doi: 10.1177/1474515114548649] [Medline: 25139467]
33. Zhou L, Gao J, Li D, Shum H. The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. Computational Linguistics 2020 Mar;46(1):53-93. [doi: 10.1162/coli_a_00368]
34. Holt-Quick C, Warren J. Establishing a Dialog Agent Policy using Deep Reinforcement Learning in the Psychotherapy Domain. 2021 Presented at: ACSW '21: 2021 Australasian Computer Science Week Multiconference; February 1-5, 2021; Dunedin, New Zealand. [doi: 10.1145/3437378.3437395]
35. Social Media Use in 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ [accessed 2022-10-10]
36. Teens, Social Media and Technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2022-10-10]

Abbreviations

CBT: cognitive behavioral therapy
GUI: graphical user interface
HABITs: Health Advances through Behavioural Intervention Technologies
SPARX: Smart, Positive, Active, Realistic, X-factor thoughts

Edited by T Leung, V Arnold, H Gouda; submitted 14.04.22; peer-reviewed by D Richards, R Pine; comments to author 13.06.22; revised version received 11.07.22; accepted 18.08.22; published 04.11.22.

Please cite as:
Ludin N, Holt-Quick C, Hopkins S, Stasiak K, Hetrick S, Warren J, Cargo T
A Chatbot to Support Young People During the COVID-19 Pandemic in New Zealand: Evaluation of the Real-World Rollout of an Open Trial
J Med Internet Res 2022;24(11):e38743
URL: https://www.jmir.org/2022/11/e38743
doi:10.2196/38743
PMID:36219754

©Nicola Ludin, Chester Holt-Quick, Sarah Hopkins, Karolina Stasiak, Sarah Hetrick, Jim Warren, Tania Cargo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Mental Health Chatbot for Young Adults With Depressive Symptoms During the COVID-19 Pandemic: Single-Blind, Three-Arm Randomized Controlled Trial

Yuhao He[1,2], BE, MEd; Li Yang[1,2], BM, MM, PhD; Xiaokun Zhu[3], BEd, MEd; Bin Wu[4], BE, ME; Shuo Zhang[4], BE, ME; Chunlian Qian[1,2], BS, MEd; Tian Tian[5], BE, ME

[1]Institute of Applied Psychology, College of Education, Tianjin University, Tianjin, China

[2]Laboratory of Suicidology, Tianjin Municipal Education Commission, Tianjin, China

[3]Tianjin Vocational Institute, Tianjin, China

[4]Tianjin Quesoar Intelligent Technology Co, Ltd, Tianjin, China

[5]College of Intelligence and Computing, Tianjin University, Tianjin, China

Corresponding Author:
Li Yang, BM, MM, PhD
Institute of Applied Psychology
College of Education
Tianjin University
135 Yaguan Road, Jinnan District
Tianjin, 300354
China
Phone: 86 13752183496
Email: yangli@tju.edu.cn

Abstract

Background: Depression has a high prevalence among young adults, especially during the COVID-19 pandemic. However, mental health services remain scarce and underutilized worldwide. Mental health chatbots are a novel digital technology to provide fully automated interventions for depressive symptoms.

Objective: The purpose of this study was to test the clinical effectiveness and nonclinical performance of a cognitive behavioral therapy (CBT)–based mental health chatbot (XiaoE) for young adults with depressive symptoms during the COVID-19 pandemic.

Methods: In a single-blind, 3-arm randomized controlled trial, participants manifesting depressive symptoms recruited from a Chinese university were randomly assigned to a mental health chatbot (XiaoE; n=49), an e-book (n=49), or a general chatbot (Xiaoai; n=50) group in a ratio of 1:1:1. Participants received a 1-week intervention. The primary outcome was the reduction of depressive symptoms according to the 9-item Patient Health Questionnaire (PHQ-9) at 1 week later (T1) and 1 month later (T2). Both intention-to-treat and per-protocol analyses were conducted under analysis of covariance models adjusting for baseline data. Controlled multiple imputation and δ-based sensitivity analysis were performed for missing data. The secondary outcomes were the level of working alliance measured using the Working Alliance Questionnaire (WAQ), usability measured using the Usability Metric for User Experience-LITE (UMUX-LITE), and acceptability measured using the Acceptability Scale (AS).

Results: Participants were on average 18.78 years old, and 37.2% (55/148) were female. The mean baseline PHQ-9 score was 10.02 (SD 3.18; range 2-19). Intention-to-treat analysis revealed lower PHQ-9 scores among participants in the XiaoE group compared with participants in the e-book group and Xiaoai group at both T1 ($F_{2,136}$=17.011; P<.001; d=0.51) and T2 ($F_{2,136}$=5.477; P=.005; d=0.31). Better working alliance (WAQ; $F_{2,145}$=3.407; P=.04) and acceptability (AS; $F_{2,145}$=4.322; P=.02) were discovered with XiaoE, while no significant difference among arms was found for usability (UMUX-LITE; $F_{2,145}$=0.968; P=.38).

Conclusions: A CBT-based chatbot is a feasible and engaging digital therapeutic approach that allows easy accessibility and self-guided mental health assistance for young adults with depressive symptoms. A systematic evaluation of nonclinical metrics for a mental health chatbot has been established in this study. In the future, focus on both clinical outcomes and nonclinical metrics is necessary to explore the mechanism by which mental health chatbots work on patients. Further evidence is required to confirm the long-term effectiveness of the mental health chatbot via trails replicated with a longer dose, as well as exploration of its stronger efficacy in comparison with other active controls.

XSL•FO
RenderX

Trial Registration: Chinese Clinical Trial Registry ChiCTR2100052532; http://www.chictr.org.cn/showproj.aspx?proj=135744

(J Med Internet Res 2022;24(11):e40719) doi:10.2196/40719

KEYWORDS

chatbot; conversational agent; depression; mental health; mHealth; digital medicine; randomized controlled trial; evaluation; cognitive behavioral therapy; young adult; youth; health service; mobile health; COVID-19

Introduction

The COVID-19 pandemic has had a huge impact on people's mental health, increasing the rates of depression and anxiety by more than 25% globally in the first year, with people aged 20-24 years being more affected than older people [1]. However, there are still many limitations in traditional face-to-face psychotherapy and mental health services, including expensive treatment, geographical limitations, few experienced therapists, and delayed treatment [2], and stigma is considered as the most significant barrier to providing mental health services [3,4]. Limited accessibility and acceptability were more obvious with the rising risk of mental health problems [5,6] led by quarantine and social isolation during the COVID-19 pandemic [7], especially among adolescents [8]. Mobile health and digital medicine have rapidly become an important area of study [9] in response to the conundrum posed by the escalating demand for mental health assistance [10] and the severe shortage of traditional health care facilities [11].

Driven by digital technologies, such as computers, the internet, mobile devices, mobile software apps, and virtual reality (VR), treatment for mental health problems has undergone an unprecedented transformation [12].

A chatbot, as a novel digital technology for mental health service, is a software program that simulates conversations with users through text or voice depending on artificial intelligence (AI) [13]. The first chatbot, ELIZA, was applied in the field of psychology, and users could input text to simulate a conversation with a Rogerian psychotherapist [14]. A mental health chatbot provides more accessibility than traditional face-to-face counseling and psychotherapy [15,16], through which users can feel accompanied and understood [17,18]. In addition, chatbots were designed to focus on interactive capabilities instead of single psychological education for facilitating the process of psychotherapy [19]. Most mental health chatbots can independently provide service to users, without requiring the participation and guidance of human therapists [20]. However, studies have shown that mental health chatbots have some risks as well, such as "misunderstanding," which may lead to ineffective or even harmful interventions, lack of crisis warning mechanisms, and lack of privacy protection [21]. Chatbots in mental health are nascent [22], and although chatbots have demonstrated feasibility to provide mental health treatment, more high-quality evidence regarding the effectiveness and acceptability of mental health chatbots is needed [23], particularly during the COVID-19 pandemic [24].

According to the latest data of the World Health Organization, there were 3.22 million depressed people worldwide in 2015 [25]. In China, the figure has been reported to be 95 million [26], and the prevalence among college students was 28.4% [27], reaching 34% during the COVID-19 pandemic [28]. However, the use of health services for depressive disorders in China has been rather limited, with the access rate of adequate treatment being less than 0.5% [29]. Cognitive behavioral therapy (CBT) has been continuously developed and is currently recognized as a widespread and effective evidence-based psychotherapy for depression [30,31], serving as one of the crucial theoretical frameworks for chatbot interventions. In recent years, a number of mental health chatbots have emerged, and their effectiveness has been tested through randomized controlled trials, providing interventions for different mental health problems, with Woebot [32], Tess [33,34], Wysa [35], Vivibot [36], and XiaoNan [37] directly targeting depression and anxiety symptoms; Shim [38], SABORI [39], and Bella [40] directly targeting stress, well-being, or quality of life; and MYLO [41-43] and Help4Mood [44] directly targeting general psychological distress such as problem solving and negative cognition.

The technology and format of chatbots for mental health problems have evolved from script bots with only text communication to embodied conversational agents [45] with image and voice, and digital humans [40] and virtual humans [46], which discern and control emotional and facial expressions during interactions with individuals in real time, have also been reported.

However, previous studies focused more on the role of a chatbot as a technical carrier in the intervention, neglecting the verification and innovation of the psychological process and content itself. As a result, there is a gap between the progress of psychology and AI in the field of digital mental health. Chatbots are intended to foster collaboration, integration, and co-development between psychological science and other fields [47]. Thus, a direct comparison between mental health chatbots and general chatbots is essential in a trial. Methodological limitations that existed in previous trials involved an insufficient sample size, a lack of follow-up assessment, failure to comprehensively investigate the long-term effectiveness of the intervention, and ignorance of the sensitivity analysis to ensure robustness of the conclusion.

As an alternative and useful precursor to clinical effectiveness, nonclinical metrics are just as important as clinical outcomes and may contribute to further exploring the mechanism by which the mental health chatbots work [48]. Fitzpatrick et al [32] also noted that therapeutic process factors of mental health chatbots may facilitate or undermine the treatment. From the technical perspective, there is currently no standard method in use to evaluate mental health chatbots. As a result, we attempted to establish a systematic evaluation of nonclinical metrics for mental health chatbots covering adherence, engagement,

XSL•FO
RenderX

working alliance, usability, acceptability, and thematic analysis of users' feedback. Working alliance (also known as "therapeutic alliance") represents the cooperative and emotionally connected relationship between the client and the therapist, and is considered a common factor in psychotherapy outcomes [49] and a metric to assess the computer-patient relationship as well [50,51]. Three recent studies [52-55] by Dosovitsky et al, Beatty et al, and Darcy et al had emphasized the viability and significance of the relationship and working alliance in digital treatment, and several randomized controlled trials of mental health chatbots had employed the Working Alliance Inventory (WAI) as a measurement method of working alliance, with all of these demonstrating good measure effects [37,56,57]. Important issues to be addressed for chatbots in the future could be extracted from the perceptions and opinions of patients [58], and thematic analysis with a topic model is a qualitative research method to accurately capture and concisely present key information in texts [59].

A randomized controlled trial including 148 Chinese college students was conducted in this study to evaluate the performance and efficacy of a mental health chatbot (XiaoE) for depression. We expected that, compared with an e-book and a general chatbot, the mental health chatbot would be more effective in reducing depressive symptoms after 1-week treatment and that this effect would persist for 1 month after the intervention (primary hypothesis). Additionally, we hypothesized that the mental health chatbot would make it easier to build relationships with users, enhance engagement, and improve user experience during the therapeutic process (secondary hypothesis).

Methods

Study Design and Participants

The study was a single-blind, 3-arm randomized controlled trial performed at a university in Tianjin, China. College students were recruited from social media outlets, online platforms, and university communities or were referred here by their counselor in the counseling center. All potential participants were screened by counseling psychologists for eligibility against the following inclusion criteria: (1) age 17-34 years; (2) average score of the depression subscale in the College Students Mental Health Screening Scale (CSMHSS) [60] within 2 to 3; and (3) ability to read Chinese. Participants were excluded if they (1) reported a score of ≥3 for any item in the suicide subscale in the CSMHSS; (2) reported a standard score of >3 in the suicide subscale or hallucination/delusion subscale in the CSMHSS; or (3) were taking a psychiatric medication. The CSMHSS is the mainstream tool for mental health screening of college students in China. The screening scale includes 22 dimensions that involve the main mental health problems of college students and is divided into 3 levels of screening that indicate 3 levels of mental health risk. The CSMHSS is a relevant tool for the inclusion criteria because it can not only measure the degree of depressive symptoms but also screen out individuals with high mental health risk for exclusion. Moreover, the CSMHSS is easier to implement in a university with the help of corresponding assessment platforms, given the large number of recruits. Before the enrollment, the participants were required to carefully read and sign the written informed consent form to confirm their acceptance of the study. Participants were provided with access to artificial psychological counseling services if they had any risk of suicide, self-injury, or severe psychological distress during or after the trial, to avoid further damage. At the end of the trial, participants in control conditions were offered access to XiaoE. The trial was prospectively registered with the ChiCTR registry on October 30, 2021 (number: ChiCTR2100052532). Final data were collected on December 16, 2021. Participants received a compensation of RMB 70 (approximately US $10) for their participation in this trial.

Randomization and Masking

Randomization with stratification by gender was performed via computer programs independently developed by the technical development team of XiaoE. Participants who were randomly assigned (1:1:1) to receive the mental health chatbot intervention, e-book intervention, or general chatbot intervention would automatically enter the corresponding intervention process when they checked into XiaoE for the first time. Treatment allocation was masked from participants, investigators, and those involved in analyzing trial data, as it was saved in an encrypted electronic file form by multiple parties (the study designer, trial implementer, data processor, and technical development representative) and unblinded after the completion of data analysis. The intervention as well as the outcome measure were completed online, and none of the investigators had access to the participants' systems during the intervention period (single blind).

Procedure

The intervention lasted for 1 week. On the day of enrollment (T0), baseline data were collected, including a pretest of the primary outcome measure (9-item Patient Health Questionnaire [PHQ-9]) and demographic information. A posttest of the primary outcome was performed 1 week later (T1), accompanied by the secondary outcomes working alliance, usability, and acceptability. A final follow-up assessment of the primary outcome was carried out 30 days after enrollment (T2).

XiaoE

XiaoE is an unguided CBT-based chatbot developed for depression, which can be used in screening, prevention, and self-assistance for depressive symptoms through a fully automatic intelligent interaction with users (text, image, and voice). The technology of XiaoE is rooted in natural language processing (NLP) and deep learning [61]. The whole chatbot dialogue system has been constructed through the open-source framework RASA [62], with content about mental health produced, discussed, and supervised by a psychologist panel led by several experienced clinical and counseling experts from schools and hospitals. XiaoE provides self-assistance service via the WeChat Official Accounts Platform. The objective of the development of XiaoE is not to replace human therapists, but to provide a convenient self-help intervention to users failing to receive immediate mental health services. It can also serve as an auxiliary tool to cooperate with traditional psychological counseling and treatment, covering functions including campus and epidemic-related counseling, adolescent mental health

XSL•FO
RenderX

screening and diagnostics, automated CBT-based chatbot interventions, intelligent multiturn conversations, artificial psychological counseling, and "tree hole" (a place to share thoughts and secrets). Participants in this condition were exposed to only the automated CBT-based chatbot intervention. Based on the principles of CBT, multiturn dialogue [63] and personalized customization were taken as the main intervention forms by referring to the content and process of several mature CBT-based chatbots [20] and internet-delivered cognitive behavioral therapy (ICBT) apps [64]. The following 7 modules were designed: "Cognition Challenge," "Improve Self-esteem," "Learn to Relax," "Energy List," "Wonderful World," "Are You OK," and "Escape from Loneliness," and they correspond to the 7 concepts of psychology, cognitive distortions, self-esteem, mindfulness meditation, mental energy, natural connection, self-help, and loneliness, respectively. Participants were asked to complete a module per day in sequence during the 1-week intervention period, as well as a separate module called "Gratitude Journal" for recording positive events and mood every day.

XiaoE is equipped with complete process guidance and daily task reminders. During the implementation of the trial, the participants were only required to follow the guidance of XiaoE every day, where the staff only provided answers to technical or operational questions. In addition, the interaction data of engagement and use of XiaoE can be obtained in the background of the system. The data could not be obtained from the control groups because the interactions occurred outside the XiaoE system. As a result, the interaction frequency in the control groups was measured in the form of a self-rating questionnaire at the end of the trial.

e-Book

Participants in control group 1 were arranged to read an e-book about depression, I Had a Black Dog [65], which is a classic book that introduces depression knowledge to the public and guides to help depressed patients serve themselves from the first-person perspective of depressed patients and their companions. The World Health Organization adopted the animated version as its official promotional video [66] on the theme of depression. In addition, participants in the group were presented with a high-quality depression-related article daily, with the theme of each article corresponding to the daily theme of the functional modules of the intervention group.

Xiaoai

Participants in control group 2 were asked to communicate with Xiaoai at least once a day. Xiaoai is a chatbot in China designed to cater to the demands of a wider audience for small talk and not particularly for mental health services such as depression. The chat content between participants and Xiaoai was unrestricted. However, we limited the daily conversation topics (corresponding to the daily functional modules of the intervention group) and proposed specific chat tasks to the participants. For example, the topic on day 2 was self-esteem, and we endorsed that participants share their perspectives and feelings on self-esteem with Xiaoai, discuss "how self-esteem affects our emotional state and what is the relationship between it and depression," assess their current level of self-esteem with Xiaoai, and ask for advice on "how to improve it."

Outcomes

Primary Outcome

The primary outcome was the score of the PHQ-9 [67], which is one of the most widely used, reliable, and validated measures of depressive symptoms. It is a 9-item self-report questionnaire that assesses the frequency and severity of depressive symptoms within the previous 2 weeks based on the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) criteria for major depressive disorder on a 4-point scale from 0 (not at all) to 3 (nearly every day). Scores ranging from 0 to 5 indicate no symptoms of depression, and scores of 5-9, 10-14, 15-20, and 20 represent mild, moderate, moderately severe, and severe depression, respectively.

Secondary Outcomes

The secondary outcomes were the scores of the Working Alliance Questionnaire (WAQ) [68], the Usability Metric for User Experience-LITE (UMUX-LITE) [69], and the Acceptability Scale (AS). The WAQ is based on the Helping Alliance Questionnaire (HAq-II), WAI, and California Psychotherapy Alliance Scales (CAL-PAS), with three 4-item subscales assessing the development of an affective bond in treatment and the level of agreement with treatment goals and treatment tasks. The scores of all 12 items range from 0 (rarely) to 5 (always). Usability, as "the extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use" [70], was assessed by the UMUX-LITE, with 2 items to assess usefulness and ease of use, respectively, ranging from 0 (rarely) to 5 (always). Acceptability, referring to psychological acceptability for the therapeutic process and content, was assessed using a 5-point Likert scale (AS), referring to items used in previous studies on mental health chatbots [33,71] covering overall satisfaction, content satisfaction, emotional awareness, learning new knowledge, relevance to daily life, and promotion of the self-help process.

Statistical Analyses

Sample size calculation was conducted with G* Power (version 3) [72]. Latest research showed a large effect (d=0.83) of a chatbot intervention for depression in college students [37]. On the assumption that a replication study might be expected to achieve broadly similar results, we calculated that a sample size of 32 in each group would have 90% power to detect a net effect size of 0.83, using analysis of covariance (ANCOVA) with a 2-sided significance level of .05, while also allowing for a 20% loss to follow-up.

Difference tests were conducted with SPSS (version 26; IBM Corp). In order to determine whether any significant differences between groups existed at baseline, F tests with one-way analysis of variance (ANOVA) were performed on continuous baseline variables (PHQ-9 and age), and chi-square analyses were performed on categorical or nominal variables (gender, ethnicity, only child, single parent, religion, home location, and parental marriage). The same comparisons of baseline

XSL•FO
RenderX

characteristics were conducted between dropouts and participants who completed the study. Adjusted mean changes in the PHQ-9 score from baseline to T1 and T2 were analyzed as the primary efficacy endpoint using an ANCOVA model with the treatment group as the fixed effect and the corresponding baseline value as the covariate. A covariate was removed from the statistical model in case of significant interaction effects being found between this covariate and the group. A post-hoc test with Bonferroni correction was employed for multiple group comparisons. η^2 was calculated and converted to Cohen *d* to examine the effect size of the group difference [73]. A Cohen *d* of 0.2 represents a small effect; 0.5, a moderate effect; and 0.8, a large effect [74]. *F* tests with ANOVA were performed for the results of secondary outcomes.

The results of both the intention-to-treat (ITT) analysis [75] on the full analysis set (all enrolled participants) and the per-protocol (PP) analysis on the PP set (participants in the full analysis set without important protocol violations leading to exclusion) were reported by including all available observations in the analysis [76]. Using *mi impute* within Stata (version 15; StataCorp), we processed missing data via multiple imputation (MI) methods and performed further sensitivity analysis via δ-based methods [77].

There are 3 broad classes of missing data mechanism assumptions [78]: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). MI is based on MAR, where the probability of a datum being missing does not depend on the unobserved value of the datum, but only depends on the observed values of other recorded variables. Nevertheless, missing data may not necessarily conform to MAR. Instead, they may follow MNAR, where the probability of a datum being missing does depend on the unobserved value of the datum, even given the observed data. We cannot distinguish between MNAR and both MAR and MCAR since the true values of missing data are never known, which means the results of MI may be biased. The publication of ICH E9 (R1) [79], addendum on estimands and sensitivity analysis in clinical trials, states that sensitivity analysis of missing data should be performed to ensure the robustness of the results. As a result, we performed a sensitivity analysis with δ-based methods to see if the effect remained significant when missing data followed MNAR. δ-based MI entails modifying the MAR imputation distribution using a specified numerical delta parameter to make predicted responses better or worse than predicted under MAR. For a continuous outcome, δ, the offset parameter can represent the difference in the mean response between the observed and unobserved cases [80]. Usually, the sensitivity analysis will repeat for a range of δ values corresponding to 25%, 50%, 75%, and 100% of the absolute change from baseline of outcomes in all participants.

Adherence is revealed by chi-square analysis of the attrition of participants, and engagement is revealed by the frequency and duration of the interaction with the chatbot. An interaction was considered a session if there was engagement with the chatbot lasting at least 2 user inputs within 2 minutes and a break no longer than 1 minute. Mean interaction frequency was defined as the average number of sessions each participant had with the chatbot per day during the 1-week intervention period. Mean interaction duration was defined as the average response time of each session calculated in milliseconds between the first time the user inputs content and the last time the chatbot outputs content per day. The 1-week intervention period was divided into days 1 through 7, and each day's 24 hours were divided into 12 two-hour time periods. We recorded and calculated the mean interaction frequency and mean interaction duration for the 7 days and the 12 time periods. We recontacted all the enrolled participants after all the follow-up measurements were finished and opened access to XiaoE. They were asked 3 open-ended questions at the end of the trial: "What was your best experience using XiaoE?" "What was your worst experience using XiaoE?" and "Please make some personal comments or suggestions on XiaoE." We ran a thematic analysis on participants' feedback using Latent Dirichlet Allocation (LDA) [81], an unsupervised learning algorithm, with Pycharm (version 2020.2.2). In order to confirm the optimal number of themes for participants' feedback on each question, the perplexity under different numbers of themes should be calculated, and the topic model with the minimum perplexity should be selected. Five keywords were extracted from each theme, and each theme was named by combining keywords and original feedback text labeled as corresponding themes.

Ethics Approval

The study protocol was approved by the Medical Ethics Committee of Tianjin Anding Hospital (Tianjin Mental Health Center; number: 2021-21). All participants provided informed consent.

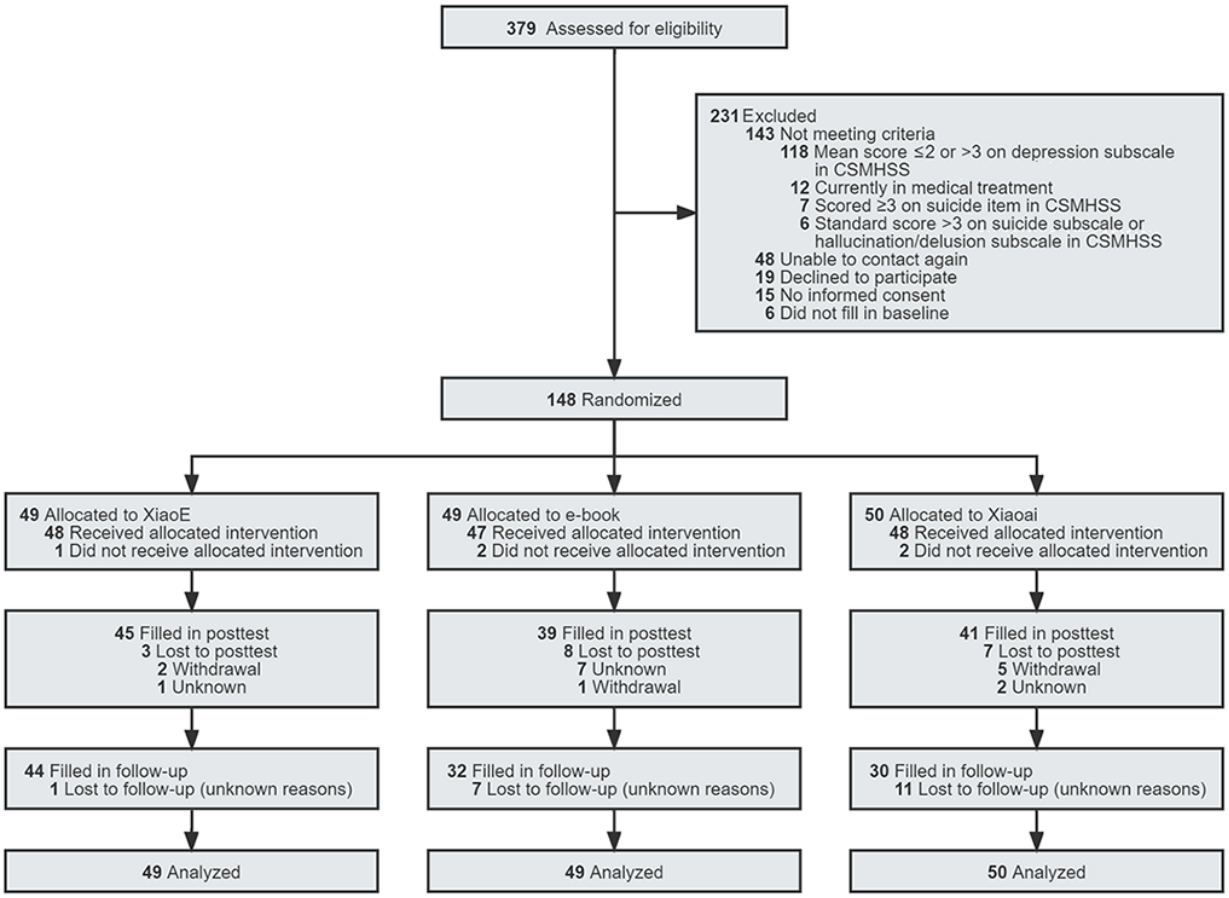
Results

Participant Characteristics

Figure 1 shows the participant flow (CONSORT flow diagram) [82]. A total of 379 college students were assessed for eligibility and enrolled between September 1, 2021, and November 15, 2021, of whom 143 did not meet the study criteria, 48 could not be contacted again, 19 declined to participate, 15 did not sign the written informed consent form, and 6 failed to complete the baseline measure. Ultimately, 148 participants were enrolled and randomized, of whom 49 were allocated to use the mental health chatbot (XiaoE), 49 were allocated to read the e-book, and 50 were allocated to use the general chatbot (Xiaoai). Participants were on average 18.78 years old (SD 0.89; range 17-21 years), and 37.2% (55/148) were female. The mean PHQ-9 score was 10.02 (SD 3.18; range 2-19) at baseline, just reaching the level of moderate depression. There were no significant differences in baseline characteristics among the 3 arms, as well as between dropouts and participants who completed the study (Table 1). Five participants (1 from the XiaoE group and 4 from the Xiaoai group) were identified by counseling psychologists as high risk during and after the course of the trial and underwent artificial psychological counseling.

XSL•FO
RenderX

Figure 1. Flow of participants (CONSORT). CSMHSS, College Students Mental Health Screening Scale.



XSL•FO
RenderX

Table 1. Baseline characteristics by randomization arm.

Characteristic	XiaoE (N=49)	e-book (N=49)	Xiaoai (N=50)	Total (N=148)	F/χ^2 (df)[b]	*P* value
PHQ-9[a] score, mean (SD)	10.10 (3.18)	9.18 (3.94)	10.76 (3.86)	10.02 (3.71)	2.294 (2,145)	.11
Age, mean (SD)	18.80 (0.89)	18.92 (0.84)	18.64 (0.90)	18.78 (0.88)	1.258 (2,145)	.29
Gender, n (%)					0.023 (2)	.99
Male	31 (63.3)	31 (63.3)	31 (62.0)	93 (62.8)		
Female	18 (36.7)	18 (36.7)	19 (38.0)	55 (37.2)		
Ethnicity, n (%)					3.239 (2)	.20
Han	44 (89.8)	44 (89.8)	49 (98.0)	137 (92.6)		
Non-Han	5 (10.2)	5 (10.2)	1 (2.0)	11 (7.4)		
Only child, n (%)					2.043 (2)	.36
Yes	16 (32.7)	13 (26.5)	10 (20.0)	39 (26.3)		
No	33 (67.3)	36 (73.5)	40 (80.0)	109 (73.7)		
Single parent, n (%)					0.450 (2)	.80
Yes	6 (12.2)	4 (8.2)	5 (10.0)	15 (10.1)		
No	43 (87.8)	45 (91.8)	45 (90.0)	133 (89.9)		
Religion, n (%)					1.912 (2)	.38
Yes	3 (6.1)	4 (8.2)	1 (2.0)	8 (5.4)		
No	46 (93.9)	45 (91.8)	49 (98.0)	140 (94.6)		
Home location , n (%)					5.057 (4)	.28
Urban	11 (22.5)	14 (28.6)	12 (24.0)	37 (25.0)		
Suburban	10 (20.4)	6 (12.2)	15 (30.0)	31 (20.9)		
Rural	28 (57.1)	29 (59.2)	23 (46.0)	80 (54.1)		
Parental marriage, n (%)					6.089 (4)	.19
Harmony	36 (73.5)	42 (85.7)	45 (90.0)	123 (83.1)		
Disharmony	7 (14.3)	5 (10.2)	2 (4.0)	14 (9.5)		
Divorced	6 (14.2)	2 (4.1)	3 (6.0)	11 (7.4)		

[a]PHQ-9: 9-item Patient Health Questionnaire.

[b]*F* value for PHQ-9 and age, and ^{2} for gender, ethnicity, only child, single parent, religion, home location, and parental marriage.

Adherence and Attrition

Of the 49 participants allocated to the XiaoE group, 4 dropped out over the 1-week period and 1 dropped out over the 1-month period. Of the 49 participants allocated to the e-book group, 10 dropped out over the 1-week period and 7 dropped out over the 1-month period. Of the 50 participants allocated to the Xiaoai group, 9 dropped out over the 1-week period and 11 dropped out over the 1-month period (Figure 1). There was a lower attrition in the intervention condition compared with the control conditions (37% vs 10%; $^2_1$=11.904; *P*<.001).

Effectiveness

ITT Analysis

At T1, no significant interaction effects were found between group and baseline PHQ-9 score (*P*=.86), age (*P*=.91), gender (*P*=.32), ethnicity (*P*=.20), only child (*P*=.33), single parent (*P*=.99), religion (*P*=.54), home location (*P*=.62), and parental marriage (*P*=.59) with the ANCOVA model. Similarly, at T2, no significant interaction effects were found between group and baseline PHQ-9 score (*P*=.16), age (*P*=.14), gender (*P*=.43), ethnicity (*P*=.96), only child (*P*=.27), single parent (*P*=.59), religion (*P*=.87), home location (*P*=.90), and parental marriage (*P*=.66) with the ANCOVA model.

Depressive symptoms significantly reduced more among participants in the XiaoE group in comparison with controls, and a moderate between-group effect size was reported at T1 ($F_{2,136}$=17.011; *P*<.001; *d*=0.51), while a small effect size was reported at T2 ($F_{2,136}$=5.477; *P*=.005; *d*=0.31) (Table 2). The post-hoc test with Bonferroni correction revealed significant treatment differences with XiaoE versus e-book and Xiaoai in the reduction of depression at T1 (*P*=.04 and *P*<.001, respectively) and T2 (*P*=.049 and *P*=.006, respectively) (Figure 2).

All results were robust under sensitivity analysis, except for the comparison with e-book at T2, which changed from significant to not significant (Table 3).

XSL•FO
RenderX

Table 2. Primary outcome measures and between-group differences in the full analysis set and per-protocol set.

Analysis and timepoint	XiaoE		e-book		Xiaoai		*F* (df)	*P* value	η^2	Cohen's *d*
	Adjusted[a] PHQ-9[b], mean (SE)	n	Adjusted[a] PHQ-9[b], mean (SE)	n	Adjusted[a] PHQ-9[b], mean (SE)	n				
ITT[c] analysis										
Postintervention	7.58 (0.30)	45	8.62 (0.30)	39	10.10 (0.30)	41	17.011 (2,136)	<.001	0.060	0.51
Change from baseline	−2.44 (0.30)		−1.40 (0.30)		0.08 (0.30)					
Follow-up	7.82 (0.34)	44	9.01 (0.35)	32	9.39 (0.35)	30	5.477 (2,136)	.005	0.024	0.31
Change from baseline	−2.20 (0.34)		−1.01 (0.35)		−0.63 (0.35)					
PP[d] analysis										
Postintervention	7.51 (0.28)	45	9.29 (0.30)	39	10.51 (0.30)	41	26.168 (2,113)	<.001	0.088	0.62
Change from baseline	−2.84 (0.28)		−1.06 (0.30)		0.16 (0.30)					
Follow-up	7.92 (0.37)	44	9.23 (0.43)	32	10.04 (0.46)	30	6.408 (2,94)	.002	0.044	0.43
Change from baseline	−2.41 (0.37)		−1.10 (0.43)		−0.29 (0.46)					

[a]Adjusted for baseline PHQ-9 score, age, gender, ethnicity, only child, single parent, religion, home location, and parental marriage.

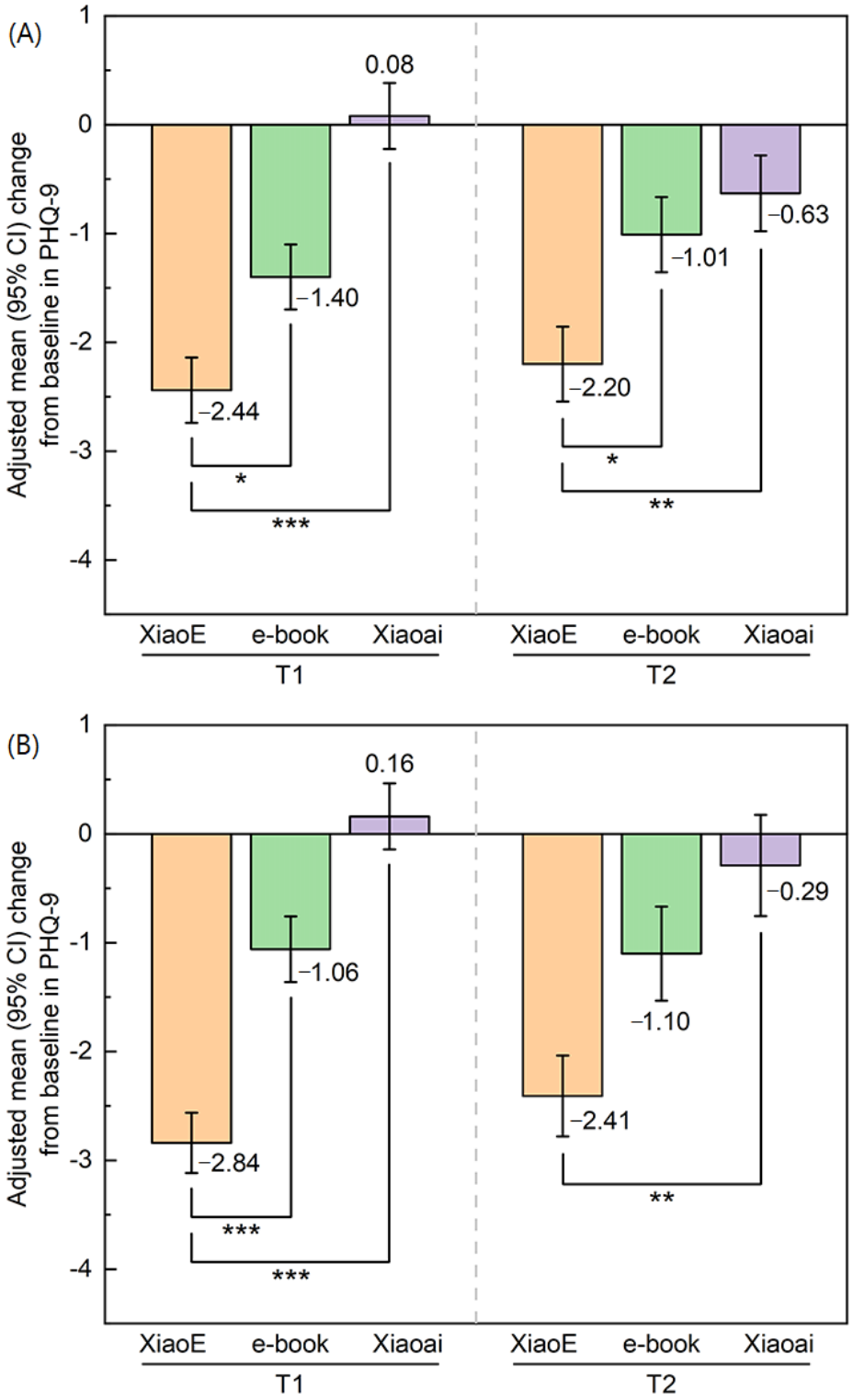
[b]PHQ-9: 9-item Patient Health Questionnaire.

[c]ITT: intention-to-treat.

[d]PP: per-protocol.

XSL•FO
RenderX

Figure 2. Efficacy for the reduction of depression symptoms in participants. The image presents the mean change from baseline in the primary outcome measure (9-item Patient Health Questionnaire [PHQ-9]) and the between-group differences in participants with XiaoE versus those with e-book and Xiaoai at postintervention and at follow-up. Means and standard errors are displayed. (A) Intention-to-treat analysis. (B) Per-protocol analysis. *P<.05; **P<.01; ***P<.001.



XSL•FO
RenderX

Table 3. δ-based sensitivity analysis.

Time and analysis[a]	Compared to e-book			Compared to Xiaoai		
	Group difference, value (SE)	95% CI	*P* value	Group difference, value (SE)	95% CI	*P* value
T1 (after 1 week)						
MI[b], MAR[c]	−1.52 (0.43)	−2.38 to −0.66	.001	−2.62 (0.42)	−3.45 to −1.78	<.001
δ=−0.31	−1.45 (0.43)	−2.31 to −0.59	.001	−2.52 (0.42)	−3.36 to −1.69	<.001
δ=−0.62	−1.38 (0.44)	−2.25 to −0.51	.002	−2.43 (0.43)	−3.27 to −1.58	<.001
δ=−0.93	−1.31 (0.44)	−2.19 to −0.43	.004	−2.33 (0.43)	−3.18 to −1.48	<.001
δ=−1.24	−1.24 (0.45)	−2.13 to −0.35	.007	−2.24 (0.44)	−3.10 to −1.37	<.001
T2 (after 1 month)						
MI, MAR	−1.11 (0.54)	−2.18 to −0.03	.043	−1.65 (0.55)	−2.74 to −0.56	.003
δ=−0.32	−1.03 (0.54)	−2.11 to 0.04	.06	−1.55 (0.55)	−2.64 to −0.46	.006
δ=−0.64	−0.96 (0.55)	−2.05 to 0.12	.08	−1.45 (0.55)	−2.55 to −0.35	.01
δ=−0.96	−0.89 (0.55)	−1.98 to 0.20	.11	−1.35 (0.56)	−2.46 to −0.24	.02
δ=−1.28	−0.81 (0.56)	−1.92 to 0.29	.15	−1.25 (0.56)	−2.37 to −0.13	.03

[a]The absolute mean change from baseline to postintervention in the PHQ-9 score of all participants was −1.24, and the absolute mean change from baseline to follow-up in the PHQ-9 score of all participants was −1.28.

[b]MI: multiple imputation.

[c]MAR: missing at random.

PP Analysis

At T1, no significant interaction effects existed between group and baseline PHQ-9 score (*P*=.59), age (*P*=.88), gender (*P*=.47), ethnicity (*P*=.44), only child (*P*=.39), single parent (*P*=.86), religion (*P*=.69), home location (*P*=.21), and parental marriage (*P*=.57) with the ANCOVA model. Similarly, at T2, no significant interaction effects existed between group and baseline PHQ-9 score (*P*=.34), age (*P*=.30), gender (*P*=.98), ethnicity (*P*=.95), only child (*P*=.11), single parent (*P*=.37), religion (*P*=.68), home location (*P*=.53), and parental marriage (*P*=.52) with the ANCOVA model.

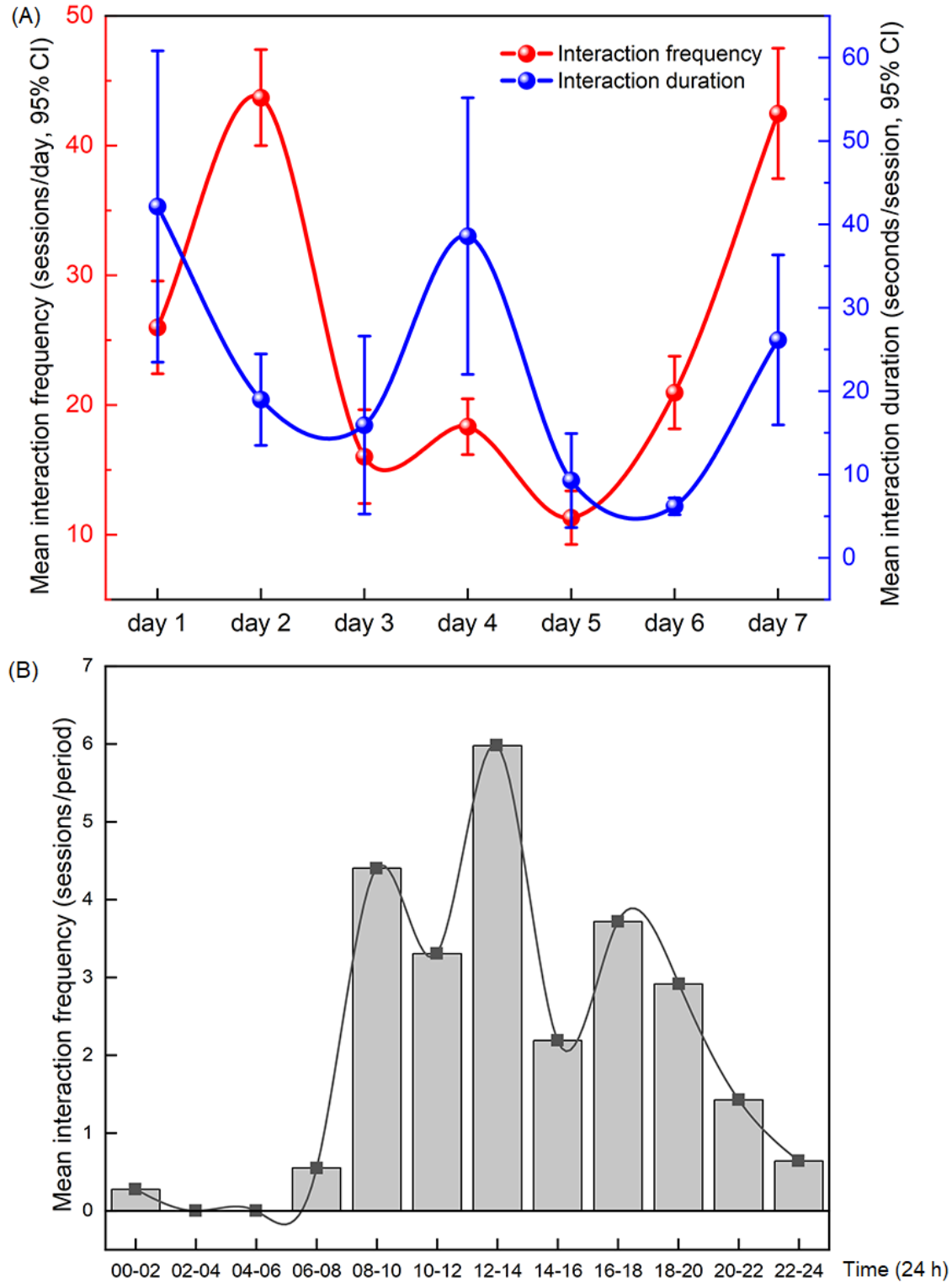
Depressive symptoms significantly reduced more among participants in the XiaoE group in comparison with controls, and a moderate between-group effect size was reported at T1 ($F_{2,113}$=26.168; *P*<.001; *d*=0.62), while a small effect size was reported at T2 ($F_{2,94}$=6.408; *P*=.002; *d*=0.43) (Table 2). The post-hoc test revealed significant treatment differences with XiaoE versus e-book and Xiaoai in the reduction of depression at T1 (*P*<.001 and *P*<.001, respectively) and a significant difference between XiaoE and Xiaoai (*P*=.003) but no significant difference between XiaoE and e-book (*P*=.08) at T2 (Figure 2).

Use and Engagement

As shown in Figure 3, participants in the XiaoE group interacted with the chatbot for 25.54 sessions (SD 26.45; range 0-172) on average per day, and each session lasted an average of 22.46 seconds (SD 79.88; range 0-758 seconds) over the 1-week period. The daily frequency and duration of the interaction were high on day 1, day 2, and day 7, while they were relatively low on day 3, day 5, and day 6, and rebounded to some extent on day 4. The frequency of the interaction reached peaks in the 3 time periods of 8-10 AM, 12-2 PM, and 4-6 PM per day. According to the answers of participants in the e-book group, 2% (1/49) had not read it once, 51% (25/49) had read it once, and 47% (23/49) had read it twice or more. In the Xiaoai group, 29% (14/48) said they interacted with Xiaoai once a day, 27% (13/48) said twice a day, and 44% (21/48) said 3 or more times a day.

XSL•FO
RenderX

Figure 3. Use and engagement with XiaoE. The image shows the frequency and duration of interaction with the chatbot and the trend of daily interactions and interactions for 12 time periods per day in the XiaoE group during the intervention. (A) Daily engagement. The x-axis represents each day of the 1-week intervention. (B) Engagement for 12 time periods. The x-axis represents each time period in 1 day.



Working Alliance, Usability, and Acceptability

Table 4 summarizes the results of the secondary outcomes. Participants in the XiaoE condition scored higher on the total WAQ ($F_{2,145}$=3.407; P=.04), as well as the subscales Bond ($F_{2,145}$=3.890; P=.02) and Engagement ($F_{2,145}$=3.925; P=.02) compared with the e-book group and the Xiaoai group. No significant difference among arms was found on the UMUX-LITE ($F_{2,145}$=0.968; P=.38). Better acceptability was discovered in the XiaoE group for total AS ($F_{2,145}$=4.322; P=.02), content satisfaction ($F_{2,145}$=5.093; P=.007), emotional awareness ($F_{2,145}$=3.636; P=.03), learning new knowledge ($F_{2,145}$=4.330; P=.02), and relevance to daily life ($F_{2,145}$=4.834; P=.009).

XSL•FO
RenderX

Table 4. Secondary outcome measures and differences between conditions.

Variable	XiaoE (n=49), mean (SD)	e-book (n=49), mean (SD)	Xiaoai (n=50), mean (SD)	*F* (df)	*P* value
WAQ[a] score					
Total	53.94 (5.96)	50.35 (9.38)	50.68 (6.87)	3.407 (2,145)	.04
Goal task	17.22 (2.71)	16.43 (3.10)	16.54 (2.48)	1.188 (2,145)	.31
Bond	18.47 (1.92)	17.06 (3.26)	17.32 (2.64)	3.890 (2,145)	.02
Engagement	18.24 (2.25)	16.86 (3.54)	16.82 (2.69)	3.925 (2,145)	.02
UMUX-LITE[b] score					
Total	8.61 (1.43)	8.31 (1.52)	8.24 (1.30)	0.968 (2,145)	.38
Usefulness	4.16 (0.94)	4.14 (0.76)	4.08 (0.78)	0.135 (2,145)	.87
Ease of use	4.45 (0.71)	4.16 (0.87)	4.16 (0.77)	2.192 (2,145)	.12
AS[c] score					
Total	27.86 (3.25)	25.82 (5.04)	25.48 (4.53)	4.322 (2,145)	.02
Overall satisfaction	4.67 (0.75)	4.43 (0.89)	4.32 (0.89)	2.264 (2,145)	.11
Content satisfaction	4.76 (0.52)	4.45 (0.79)	4.30 (0.81)	5.093 (2,145)	.007
Emotional awareness	4.57 (0.74)	4.20 (1.00)	4.12 (0.90)	3.636 (2,145)	.03
Learning new knowledge	4.63 (0.64)	4.27 (0.95)	4.16 (0.89)	4.330 (2,145)	.02
Relevance to daily life	4.67 (0.63)	4.14 (1.10)	4.30 (0.81)	4.834 (2,145)	.009
Promotion of self-help process	4.55 (0.77)	4.33 (0.94)	4.28 (0.83)	1.429 (2,145)	.24

[a]WAQ: Working Alliance Questionnaire.

[b]UMUX-LITE: Usability Metric for User Experience-LITE.

[c]AS: Acceptability Scale.

Thematic Analysis

According to the chart of themes-perplexity of LDA (Figure 4), the number of themes reported in the question "What was your best experience using XiaoE?" was set to 4 and the number of themes reported in the question "What was your worst experience using XiaoE?" was set to 2. Table 5 lists all the themes and keywords for participants' feedback. The last question "comments or suggestions" was analyzed with a qualitative method because the result of LDA was not ideal.

The following 4 themes emerged in respect to the feedback to the question regarding the *best experience*: "relationship" (n=25), "emotion" (n=12), "personalization" (n=31), and "practicability" (n=80). The keywords extracted from the relationship theme were "company," "care," "loneliness," "favor," and "attending," and the corresponding labeled example text was "XiaoE is very sweet, I like to talk to XiaoE, he will accompany and accept me, so I don't feel lonely." The keywords for the emotion theme were "happy," "relax," "stress," "catharsis," and "company," and the example text was "always makes me laugh! Ha ha ha ha, the pressure suddenly disappeared, and I am so happy." The keywords for the personalization theme were "thinking," "learning," "depression," "mood," and "intelligence," and the example was "The best experience is that sometimes XiaoE's answers are indeed valuable and can really target some of my questions, which is very intelligent and promotes thinking." The keywords for the practicability theme were "convenience," "help," "reality," "method," and "usability," and the example was "practical, real and convenient, can help me."

The following 2 themes emerged in respect to the feedback to the question regarding the *worst experience*: "content" (n=120) and "technology" (n=28). The keywords extracted from the content theme were "inflexible," "response," "tedious," "repetitive," and "mechanical," and the corresponding labeled example text was "The content is too rigid. It will make people feel bored and irritable if used for a long time." The keywords extracted from the technology theme were "glitches," "lag," "system," "crash," and "inflexible," and the example text was "crashed when I just entered the interface, and some glitches need to be optimized."

The participants' feedback on the question "comments or suggestions" can be mainly extracted into the following 3 themes: hope for a more fluent process of dialogue, more emotional response and interaction, and server upgrade.

XSL•FO
RenderX

Figure 4. Chart of themes and perplexity. The image shows the perplexity under different number of themes for participant feedback of the 2 questions, "What was your best experience using XiaoE?" and "What was your worst experience using XiaoE?." (A) Themes-perplexity chart of “best experience.” (B) Themes-perplexity chart of “worst experience”.

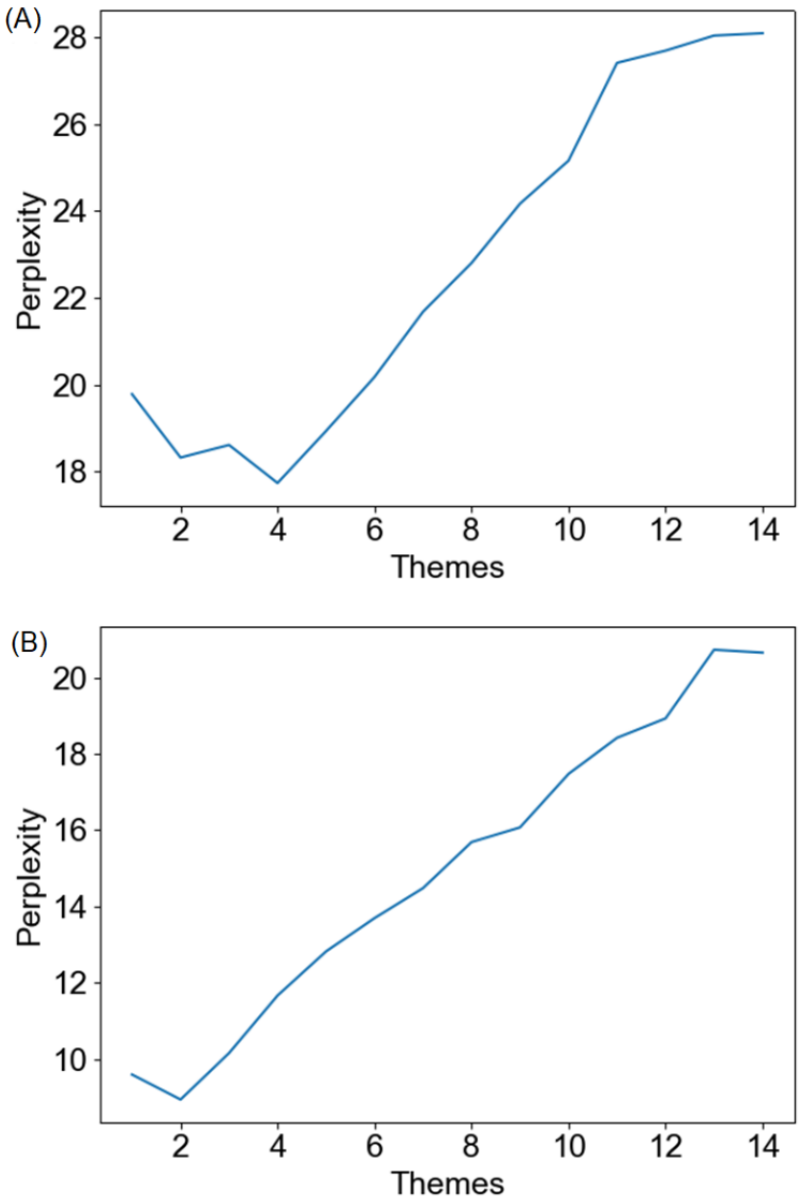


Table 5. Themes and keywords for participants’ feedback.

Question and theme	Keywords
Best experience	
Relationship	Company, care, loneliness, favor, and attending
Emotion	Happy, relax, stress, catharsis, and company
Personalization	Thinking, learning, depression, mood, and intelligence
Practicability	Convenience, help, reality, method, and usability
Worst experience	
Content	Inflexible, response, tedious, repetitive, and mechanical
Technology	Glitches, lag, system, crash, and inflexible

XSL•FO
RenderX

Discussion

Principal Findings

To our knowledge, this is the first study to directly compare the clinical efficacy of a mental health chatbot with a general chatbot performing automated teletreatment for depressive symptoms. We tested both the short- and long-term effectiveness of XiaoE via a single-blind, 3-arm randomized controlled trial and established a systematic evaluation of nonclinical metrics for mental health chatbots so as to offer references for future research.

Participants in this trial were on average 18.78 years old, and they were younger than samples of typical studies with adults or college students, indicating that research on mental health chatbots is translating to samples of adolescents. In addition, more men took part in this study than in previous studies, where the majority of participants were women. Given that there are currently no well-done studies on adolescents, we hope to see more of them in the future.

In terms of attrition, participants in the XiaoE group dropped out at a lower rate than those in the e-book and Xiaoai groups. XiaoE was associated with a high level of engagement, which rose to the highest level particularly at the beginning and toward the end of the trial, demonstrating that XiaoE was attractive to participants and could quickly establish relationships when participants came into contact with this novel AI. The participants using XiaoE were most active from 12 to 2 PM every day, which may be related to the automatic task reminder set after 12 PM once a day by XiaoE. However, a large fluctuation in engagement could be observed regarding the trend of weekly activation and daily activation, which indicated that the relationships between participants and XiaoE were not steady and firm enough.

ITT analysis showed a significantly better effectiveness of XiaoE for depression in comparison with that of the 2 controls for 1 week, achieving a moderate effect size (d=0.51), which was between the effect sizes of 2 previous studies [32,33] (Woebot: d=0.44; Tess: d=0.68) and remained robust in sensitivity analysis. The results of the long-term reduction of depressive symptoms 1 month later were statistically significant as well, while achieving a small effect size (d=0.31). PP analysis also showed significant short- and long-term effectiveness (T1, d=0.62; T2, d=0.43). As in previous studies [76], the results of the ITT analysis were lower than those of the PP analysis. However, opposite results were found in separate comparisons of the XiaoE and e-book groups. The difference between XiaoE and e-book was significant in the ITT analysis (despite failing to pass the test of the sensitivity analysis), but not in the PP analysis. Protocol deviations and the interaction between compliance and the intervention, which can lead to better outcomes for compliers in the active group but just the opposite (better for noncompliers) in the control group, are commonly thought to be the causes of the bias in the PP analysis. In this study, nevertheless, the effectiveness for compliers of the e-book group may also be overestimated due to the favorable impact of compliance, which may be more significant than that in the XiaoE group. Therefore, the difference between the 2 groups was not significant in the PP analysis. This shows that mental health chatbots should fortify the therapeutic alliance even more to increase the intervention compliance of participants.

It is necessary to note that while there was a significant improvement in symptoms via the mental health chatbot intervention, the magnitude of the improvement was small. As a result, the mental health chatbot is better suited as an auxiliary tool to work in conjunction with traditional psychological counseling and treatment or as the primary care approach for the treatment of mental illness. Although it is challenging to swiftly implement the intervention in real clinical practice, at least for the time being, the intervention is effective and convenient for individuals who desire to access self-help mental health services. This makes sense, since those represent a much larger group of people, and the spread of this unguided tool will greatly reduce the cost of human and financial resources.

XiaoE exhibited a significant high level of acceptability and work alliance with participants but a nonsignificant high level of usability. This shows that XiaoE has preliminarily reached the standard of capacity in relationship establishment, but some aspects, such as the user interface and the operating system, still need to be further simplified for users. Participants reported having received the best experience with XiaoE in the 4 themes of “relationship,” “emotion,” “personalization,” and “practicality.” The theme of “relationship” reflected the establishment, development, and function of the relationship between XiaoE and participants, as Dosovitsky et al [52] found that individuals can form a positive bond with an AI chatbot owing to its personality traits, such as being caring, open to listening, and nonjudgmental. The theme of “emotion” reflected that communication with XiaoE was helpful for emotional expression and catharsis, and made users feel accompanied and understood. The same themes were also observed in previous studies [32,33]. The theme of “personalization” reflected that XiaoE can make different suggestions for different emotional distresses put forward by the participants, which can trigger more thinking and learning of the participants. At the beginning of the content design of XiaoE, in order to avoid an overly sermonizing feeling, we added many simple and specific tips. This could be the reason why participants considered XiaoE to be practical (“practicality”). The worst experiences reported mainly focused on “content” and “technology.”

As mentioned earlier, the use of psychology in chatbots is still superficial. Despite the fact that our content was based on CBT, we discovered through our thematic analysis that participant comments barely made any mention of it. XiaoE’s conversations do not always emphasize CBT itself to participants, similar to how patients receiving therapy from a human therapist may feel like they are improving but not know what kind of therapy they are actually receiving. On the other hand, it is evident that CBT has its limitations. Even though CBT is a highly structured therapy, translating a typical CBT-based psychotherapy into a chatbot setting is difficult.

Comparison With Prior Work

We added a general chatbot (Xiaoai) intervention as a control condition to demonstrate the significance of psychological design and content for mental health chatbots. Interestingly,

XSL•FO
RenderX

participants who interacted with Xiaoai showed a small worsening of depressive symptoms after receiving the 1-week intervention. This indicates that using a general chatbot to treat mental health problems may be harmful, and a specifically designed chatbot for mental health may be required to alleviate depressive symptoms. In this study, follow-up was added to investigate the long-term effectiveness, and δ-based sensitivity analysis was performed to ensure the robustness of the conclusion. We established an innovative systematic evaluation of nonclinical metrics for mental health chatbots, and LDA was applied for the first time in the thematic analysis of users' feedback as the sample size increased.

Limitations and Future Directions

There were some limitations in this study: First, due to the particularity of the tool and the consideration of actual recruitment, it was below capacity to double-blind both the investigators and the participants. For the convenience of management, an online group was set up for the 148 enrolled participants to provide important information and technical solutions during the implementation process of the trial, which, as a potential risk, may have resulted in an attempt to reveal different contents of their own interventions, thus imposing subjective influence on the effectiveness for other participants. Special attention should be paid to this in future online research. Second, the 1-week intervention period in this study was relatively short, and the results might have shown some difference if the intervention was prolonged. It can be concluded from the trend of weekly activation that engagement with XiaoE had a wide fluctuation range and XiaoE showed strong attractiveness, but it rapidly faded in the middle of the trial. It may have resulted from repeated interactions with the inflexible and tedious content, as well as technical problems such as glitches and lag. It can be speculated that the chatbot may be more suitable for a short-term intervention rather than a long-term intervention, which needs to be explored in further studies with a longer treatment period. Third, for the control condition, the strength of evidence for the intervention itself was still limited, and the e-book intervention, as a self-help approach, only involved the concept of psychological education, and it was not equipped with a complete set of programs for psychological therapy [83] or designed for multiple or recurring sessions. Therefore, it is better to choose other active control approaches whose efficacy has been clearly proven, including traditional face-to-face therapy, online psychological counseling, ICBT, and VR. Fourth, as we could only gather self-reported involvement in the control groups as opposed to comprehensive objective data in the treatment group, it was not possible to directly compare the engagement of the XiaoE group with that of the control groups. Future research should also collect behavioral data in control groups corresponding to the data in the treatment group as the basis for comparison. Finally, this study involved students from a single university in Tianjin, China, and it was not determined whether the conclusion can be extended to a larger group. This can be addressed by attempting to perform multicenter randomized controlled trials in the future.

In the postepidemic era, people's lifestyles have undergone profound transformations, and digital technology and internet informatization have drawn more attention than ever. It is reasonable to predict that in the future, chatbot-based digital psychotherapy will play a significant role in the field of mental health care [84]. This will provide new clinical guidelines and technical viewpoints to relevant psychologists, psychiatrists, and AI researchers and practitioners.

At present, there are many digital therapeutic approaches with excellent psychological content, with little attention to the effective factors in the psychological therapeutic process, such as emotional response, therapeutic alliance, empathy, and personalization. Despite people's doubt regarding whether machines can provide emotional experiences, they typically respond better to agents that express emotions than those that do not [85], illustrating the importance of a positive therapeutic alliance in the internet environment in the absence of therapist support [86]. Chatbots with sophisticated empathic capabilities can enrich user experience and affinity. The concept of empathic chatbots has been proposed [87], accompanied with system design and development [88], but there is no mature product present and effectiveness has not been tested yet. The utilization of user profiles or user models to support personalized and adaptive features, and assessments for personalization are still limited in mental health chatbots [89]. Thus, the technologies of chatbots, particularly NLP [61] and multiturn dialogue [60], require to be constantly upgraded, and the user interface and operating system should be modified to improve user experience. Future chatbots can be targeted at more mental health problems, such as anxiety, insomnia, well-being, stress, and addiction. Meanwhile, ethical issues with AI, such as privacy, security, information disclosure, and harm avoidance need to be carefully considered [90].

Conclusions

The mental health chatbot XiaoE can be used as a feasible, engaging, and effective digital intervention for college students with depressive symptoms. Compared with a general chatbot, XiaoE exhibited significant short-term and long-term effectiveness that remained robust after sensitivity analysis, illustrating the unique role of psychological design and process in the field of digital mental health. XiaoE showed special capacity for building relationships with users, enhancing engagement, and improving user experience during the therapeutic process. Further evidence is required to confirm the long-term effectiveness via trails replicated with a longer dose, as well as exploration of its greater efficacy in comparison with other active controls.

Acknowledgments

This study was supported by the Artificial Intelligence for Sustainable Development Goals (AI4SDGs) Research Program and the National Social Science Foundation of China (grant number: 14AZD111 and grant number: 21B[GL0031]0). We thank Xiaomi

XSL•FO
RenderX

Corporation for technical support of chatbots, Tianjin Anding Hospital for supervision on the content for mental health, professor Peng Zhang and associate professor Bo Wang for their valuable suggestions on the design and development of XiaoE, Qiang Zhang and Tong Li for their statistical expertise, and Zheng Fang and Xiaojun Hu for their efforts in the implementation of the trial.

Authors' Contributions

YH, LY, and BW contributed to the conception of the study. YH, LY, and TT designed the process and content of the XiaoE intervention. SZ contributed to the development of XiaoE. YH designed the trial and drafted the manuscript. XZ contributed to the recruitment and screening of participants. YH, XZ, CQ, and TT conducted the trial. LY, BW, and SZ commented on the design of the trial and supervised data collection and analysis. YH and CQ analyzed the data. All authors revised the content critically and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1258 KB - jmir_v24i11e40719_app1.pdf]

References

1. Mental Health and COVID-19: Early evidence of the pandemic's impact: Scientific brief, 2 March 2022. World Health Organization. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1 [accessed 2022-05-30]
2. Christensen H. Computerised therapy for psychiatric disorders. The Lancet 2007 Jul;370(9582):112-113. [doi: 10.1016/s0140-6736(07)61065-4]
3. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
4. Corrigan P. How stigma interferes with mental health care. Am Psychol 2004 Oct;59(7):614-625. [doi: 10.1037/0003-066X.59.7.614] [Medline: 15491256]
5. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
6. Wu T, Jia X, Shi H, Niu J, Yin X, Xie J, et al. Prevalence of mental health problems during the COVID-19 pandemic: A systematic review and meta-analysis. J Affect Disord 2021 Feb 15;281:91-98 [FREE Full text] [doi: 10.1016/j.jad.2020.11.117] [Medline: 33310451]
7. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
8. Loades ME, Chatburn E, Higson-Sweeney N, Reynolds S, Shafran R, Brigden A, et al. Rapid Systematic Review: The Impact of Social Isolation and Loneliness on the Mental Health of Children and Adolescents in the Context of COVID-19. J Am Acad Child Adolesc Psychiatry 2020 Nov;59(11):1218-1239.e3 [FREE Full text] [doi: 10.1016/j.jaac.2020.05.009] [Medline: 32504808]
9. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
10. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. The Lancet Psychiatry 2020 Mar;7(3):228-229. [doi: 10.1016/s2215-0366(20)30046-8]
11. Xu C, Xu Y, Xu S, Zhang Q, Liu X, Shao Y, et al. Cognitive Reappraisal and the Association Between Perceived Stress and Anxiety Symptoms in COVID-19 Isolated People. Front Psychiatry 2020 Sep 2;11:858 [FREE Full text] [doi: 10.3389/fpsyt.2020.00858] [Medline: 32982809]
12. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
13. McTear M, Callejas Z, Griol D. The Conversational Interface: Talking to Smart Devices. Switzerland: Springer Publishing Company; 2016.
14. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 1966 Jan;9(1):36-45. [doi: 10.1145/365153.365168]

XSL•FO
RenderX

15. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
16. Kazdin AE. Annual Research Review: Expanding mental health services through novel models of intervention delivery. J Child Psychol Psychiatry 2019 Apr 13;60(4):455-472. [doi: 10.1111/jcpp.12937] [Medline: 29900543]
17. Ho A, Hancock J, Miner A. Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot. J Commun 2018 Aug;68(4):712-733 [FREE Full text] [doi: 10.1093/joc/jqy026] [Medline: 30100620]
18. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput.-Hum. Interact 2005 Jun 01;12(2):293-327. [doi: 10.1145/1067860.1067867]
19. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
20. Gaffney H, Mansell W, Tai S. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
21. Xu B, Zhuang Z. Survey on psychotherapy chatbots. Concurrency and Computation 2020 Dec 28;34(7):e6170 [FREE Full text] [doi: 10.1002/cpe.6170]
22. 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. [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
23. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
24. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. NPJ Digit Med 2020 May 4;3(1):65 [FREE Full text] [doi: 10.1038/s41746-020-0280-0] [Medline: 32377576]
25. Depression and Other Common Mental Disorders. World Health Organization. 2017. URL: https://www.who.int/publications/i/item/depression-global-health-estimates [accessed 2022-03-15]
26. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. The Lancet Psychiatry 2019 Mar;6(3):211-224 [FREE Full text] [doi: 10.1016/S2215-0366(18)30511-X]
27. Gao L, Xie Y, Jia C, Wang W. Prevalence of depression among Chinese university students: a systematic review and meta-analysis. Sci Rep 2020 Sep 28;10(1):15897 [FREE Full text] [doi: 10.1038/s41598-020-72998-1] [Medline: 32985593]
28. Deng J, Zhou F, Hou W, Silver Z, Wong CY, Chang O, et al. The prevalence of depressive symptoms, anxiety symptoms and sleep disturbance in higher education students during the COVID-19 pandemic: A systematic review and meta-analysis. Psychiatry Res 2021 Jul;301:113863 [FREE Full text] [doi: 10.1016/j.psychres.2021.113863] [Medline: 33984824]
29. Lu J, Xu X, Huang Y, Li T, Ma C, Xu G, et al. Prevalence of depressive disorders and treatment in China: a cross-sectional epidemiological study. The Lancet Psychiatry 2021 Nov;8(11):981-990. [doi: 10.1016/s2215-0366(21)00251-0]
30. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
31. Malhi GS, Mann JJ. Depression. The Lancet 2018 Nov;392(10161):2299-2312. [doi: 10.1016/s0140-6736(18)31948-2]
32. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
33. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
34. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial Intelligence-Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial. JMIR Form Res 2021 Aug 12;5(8):e20678 [FREE Full text] [doi: 10.2196/20678] [Medline: 34092548]
35. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
36. Greer S, Ramo D, Chang Y, Fu M, Moskowitz J, Haritatos J. Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
37. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. Internet Interv 2022 Mar;27:100495 [FREE Full text] [doi: 10.1016/j.invent.2022.100495] [Medline: 35059305]
38. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]

XSL•FO
RenderX

39. Suganuma S, Sakamoto D, Shimoyama H. An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial. JMIR Ment Health 2018 Jul 31;5(3):e10454 [FREE Full text] [doi: 10.2196/10454] [Medline: 30064969]
40. Loveys K, Sagar M, Pickering I, Broadbent E. A Digital Human for Delivering a Remote Loneliness and Stress Intervention to At-Risk Younger and Older Adults During the COVID-19 Pandemic: Randomized Pilot Trial. JMIR Ment Health 2021 Nov 08;8(11):e31586 [FREE Full text] [doi: 10.2196/31586] [Medline: 34596572]
41. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, Acceptability, and Effectiveness of Web-Based Conversational Agents to Facilitate Problem Solving in Older Adults: Controlled Study. J Med Internet Res 2020 May 27;22(5):e16794 [FREE Full text] [doi: 10.2196/16794] [Medline: 32384055]
42. Bird T, Mansell W, Wright J, Gaffney H, Tai S. Manage Your Life Online: A Web-Based Randomized Controlled Trial Evaluating the Effectiveness of a Problem-Solving Intervention in a Student Sample. Behav Cogn Psychother 2018 Sep;46(5):570-582. [doi: 10.1017/S1352465817000820] [Medline: 29366432]
43. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. Behav Cogn Psychother 2014 Nov;42(6):731-746. [doi: 10.1017/S135246581300060X] [Medline: 23899405]
44. Burton C, Szentagotai Tatar A, McKinstry B, Matheson C, Matu S, Moldovan R, Help4Mood Consortium. Pilot randomised controlled trial of Help4Mood, an embodied virtual agent-based system to support treatment of depression. J Telemed Telecare 2016 Sep;22(6):348-355. [doi: 10.1177/1357633X15609793] [Medline: 26453910]
45. Gardiner P, Bickmore T, Yinusa-Nyahkoon L, Reichert M, Julce C, Sidduri N, et al. Using Health Information Technology to Engage African American Women on Nutrition and Supplement Use During the Preconception Period. Front Endocrinol (Lausanne) 2020;11:571705 [FREE Full text] [doi: 10.3389/fendo.2020.571705] [Medline: 33584534]
46. Ma T, Sharifi H, Chattopadhyay D. Virtual Humans in Health-Related Interventions: A Meta-Analysis. In: CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 1-6. [doi: 10.1145/3290607.3312853]
47. Lim SM, Shiau CWC, Cheng LJ, Lau Y. Chatbot-Delivered Psychotherapy for Adults With Depressive and Anxiety Symptoms: A Systematic Review and Meta-Regression. Behav Ther 2022 Mar;53(2):334-347. [doi: 10.1016/j.beth.2021.09.007] [Medline: 35227408]
48. 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]
49. Cuijpers P, Reijnders M, Huibers MJ. The Role of Common Factors in Psychotherapy Outcomes. Annu Rev Clin Psychol 2019 May 07;15(1):207-231. [doi: 10.1146/annurev-clinpsy-050718-095424] [Medline: 30550721]
50. Abdulrahman A, Richards D. Modelling Therapeutic Alliance using a User-aware Explainable Embodied Conversational Agent to Promote Treatment Adherence. In: IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. 2019 Presented at: 19th ACM International Conference on Intelligent Virtual Agents; July 2-5, 2019; Paris, France p. 248-251. [doi: 10.1145/3308532.3329413]
51. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
52. Dosovitsky G, Bunge EL. Bonding With Bot: User Feedback on a Chatbot for Social Isolation. Front Digit Health 2021 Oct 6;3:735053 [FREE Full text] [doi: 10.3389/fdgth.2021.735053] [Medline: 34713203]
53. Beatty C, Malik T, Meheli S, Sinha C. Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study. Front Digit Health 2022 Apr 11;4:847991 [FREE Full text] [doi: 10.3389/fdgth.2022.847991] [Medline: 35480848]
54. Darcy A, Daniels J, Salinger D, Wicks P, Robinson A. Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study. JMIR Form Res 2021 May 11;5(5):e27868 [FREE Full text] [doi: 10.2196/27868] [Medline: 33973854]
55. Nißen M, Rüegger D, Stieger M, Flückiger C, Allemand M, V Wangenheim F, et al. The Effects of Health Care Chatbot Personas With Different Social Roles on the Client-Chatbot Bond and Usage Intentions: Development of a Design Codebook and Web-Based Study. J Med Internet Res 2022 Apr 27;24(4):e32630 [FREE Full text] [doi: 10.2196/32630] [Medline: 35475761]
56. Prochaska JJ, Vogel EA, Chieng A, Baiocchi M, Maglalang DD, Pajarito S, et al. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. Drug Alcohol Depend 2021 Oct 01;227:108986 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108986] [Medline: 34507061]
57. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A Smartphone-Based Health Care Chatbot to Promote Self-Management of Chronic Pain (SELMA): Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]

XSL•FO
RenderX

58. 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]
59. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 2018 Nov 28;78(11):15169-15211. [doi: 10.1007/s11042-018-6894-4]
60. Yi X, Ding X, Xue J, Wu X. Optimization of the explanation of the results of chinese college students' mental health screening scale. Chinese Journal of Clinical Psychology 2021;29(5):962-966 [FREE Full text]
61. Otter DW, Medina JR, Kalita JK. A Survey of the Usages of Deep Learning for Natural Language Processing. IEEE Trans. Neural Netw. Learning Syst 2021 Feb;32(2):604-624. [doi: 10.1109/tnnls.2020.2979670]
62. Bocklisch T, Faulkner J, Pawlowski N. Rasa: Open Source Language Understanding and Dialogue Management. arXiv. 2017. URL: https://arxiv.org/abs/1712.05181 [accessed 2022-11-11]
63. Kao C, Chen C, Tsai Y. Model of Multi-turn Dialogue in Emotional Chatbot. 2019 Presented at: 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI); November 21-23, 2019; Kaohsiung, Taiwan. [doi: 10.1109/taai48200.2019.8959855]
64. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, 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]
65. Johnstone M. I Had a Black Dog. Australia: Constable & Robinson; 2007.
66. Depression. World Health Organization. URL: https://www.who.int/health-topics/depression [accessed 2021-07-30]
67. Kroenke K, Spitzer RL, Williams JBW. 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]
68. Xu Z, Guangrong J. Development of the Working Alliance Questionnaire. Chinese Journal of Clinical Psychology 2011(4):449-453 [FREE Full text]
69. Borsci S, Buckle P, Walne S. Is the LITE version of the usability metric for user experience (UMUX-LITE) a reliable tool to support rapid assessment of new healthcare technology? Appl Ergon 2020 Apr;84:103007. [doi: 10.1016/j.apergo.2019.103007] [Medline: 31785449]
70. ISO 9241-11:2018(en) Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. International Organization for Standardization. URL: https://www.iso.org/obp/ui/fr/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2022-04-10]
71. Jang S, Kim J, Kim S, Hong J, Kim S, Kim E. Mobile app-based chatbot to deliver cognitive behavioral therapy and psychoeducation for adults with attention deficit: A development and feasibility/usability study. Int J Med Inform 2021 Jun;150:104440. [doi: 10.1016/j.ijmedinf.2021.104440] [Medline: 33799055]
72. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
73. Computation of Effect Sizes. Psychometrica. URL: https://www.psychometrica.de/effect_size.html [accessed 2021-12-02]
74. Cohen J. A power primer. Psychological Bulletin 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
75. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ 2011 Feb 07;342(feb07 1):d40-d40 [FREE Full text] [doi: 10.1136/bmj.d40] [Medline: 21300711]
76. Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. J Clin Epidemiol 2007 Jul;60(7):663-669. [doi: 10.1016/j.jclinepi.2006.09.013] [Medline: 17573981]
77. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Stat Med 2020 Sep 20;39(21):2815-2842. [doi: 10.1002/sim.8569] [Medline: 32419182]
78. Rubin DB. Inference and Missing Data. Biometrika 1976 Dec;63(3):581. [doi: 10.2307/2335739]
79. E9(R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. ICH. URL: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf [accessed 2021-12-05]
80. Tan P, Cro S, Van Vogt E, Szigeti M, Cornelius VR. A review of the use of controlled multiple imputation in randomised controlled trials with missing outcome data. BMC Med Res Methodol 2021 Apr 15;21(1):72 [FREE Full text] [doi: 10.1186/s12874-021-01261-6] [Medline: 33858355]
81. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. The Journal of Machine Learning Research 2003;3:993-1022. [doi: 10.5555/944919.944937]
82. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340(jul21 1):c869-b2535 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
83. Gregory RJ, Schwer Canning S, Lee TW, Wise JC. Cognitive Bibliotherapy for Depression: A Meta-Analysis. Professional Psychology: Research and Practice 2004 Jun;35(3):275-280. [doi: 10.1037/0735-7028.35.3.275]
84. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]

XSL•FO
RenderX

85. de Gennaro M, Krumhuber EG, Lucas G. Effectiveness of an Empathic Chatbot in Combating Adverse Effects of Social Exclusion on Mood. Front Psychol 2019;10:3061 [FREE Full text] [doi: 10.3389/fpsyg.2019.03061] [Medline: 32038415]
86. Clarke J, Proudfoot J, Whitton A, Birch M, Boyd M, Parker G, et al. Therapeutic Alliance With a Fully Automated Mobile Phone and Web-Based Intervention: Secondary Analysis of a Randomized Controlled Trial. JMIR Ment Health 2016 Feb 25;3(1):e10 [FREE Full text] [doi: 10.2196/mental.4656] [Medline: 26917096]
87. Casas J, Spring T, Daher K, Mugellini E, Abou KO. Enhancing Conversational Agents with Empathic Abilities. In: IVA '21: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents. 2021 Presented at: 21st ACM International Conference on Intelligent Virtual Agents; September 14-17, 2021; Virtual Event Japan. [doi: 10.1145/3472306.3478344]
88. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
89. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The Personalization of Conversational Agents in Health Care: Systematic Review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
90. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019 Sep 02;1(9):389-399. [doi: 10.1038/s42256-019-0088-2]

Abbreviations

AI: artificial intelligence
ANCOVA: analysis of covariance
AS: Acceptability Scale
CBT: cognitive behavioral therapy
CSMHSS: College Students Mental Health Screening Scale
ICBT: internet-delivered cognitive behavioral therapy
ITT: intention-to-treat
LDA: Latent Dirichlet Allocation
MAR: missing at random
MCAR: missing completely at random
MI: multiple imputation
MNAR: missing not at random
NLP: natural language processing
PHQ-9: 9-item Patient Health Questionnaire
PP: per-protocol
UMUX-LITE: Usability Metric for User Experience-LITE
VR: virtual reality
WAQ: Working Alliance Questionnaire

Edited by T Leung, V Arnold, H Gouda; submitted 02.07.22; peer-reviewed by E Bunge, I Ezawa; comments to author 23.09.22; revised version received 14.10.22; accepted 03.11.22; published 21.11.22.

Please cite as:
He Y, Yang L, Zhu X, Wu B, Zhang S, Qian C, Tian T
Mental Health Chatbot for Young Adults With Depressive Symptoms During the COVID-19 Pandemic: Single-Blind, Three-Arm Randomized Controlled Trial
J Med Internet Res 2022;24(11):e40719
URL: https://www.jmir.org/2022/11/e40719
doi:10.2196/40719
PMID:36355633

©Yuhao He, Li Yang, Xiaokun Zhu, Bin Wu, Shuo Zhang, Chunlian Qian, Tian Tian. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

The Reconstruction of Human Fingerprints From High-Resolution Computed Tomography Data: Feasibility Study and Associated Ethical Issues

Orestis L Katsamenis[1], BSc, MSc, PhD; Charles B Burson-Thomas[2], MEng, PGCE, PhD; Philip J Basford[1], MEng, MIET, PhD; J Brian Pickering[3], MA, PGCert, MSc, MBPsS, MBCS, DPhil; Martin Browne[2], BSc, PhD

[1]µ-VIS X-ray Imaging Centre, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom
[2]Bioengineering Science Research Group, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom
[3]Electronics and Computer Science, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Orestis L Katsamenis, BSc, MSc, PhD
µ-VIS X-ray Imaging Centre
Faculty of Engineering and Physical Sciences
University of Southampton
Highfield Campus
Southampton, SO17 1BJ
United Kingdom
Phone: 44 2380593468
Email: o.katsamenis@soton.ac.uk

(J Med Internet Res 2022;24(11):e38650) doi:10.2196/38650

KEYWORDS

microcomputed tomography; µCT; microCT; HRpQCT; fingerprints; fingerprint; ethics; ethical; image; data; biometric; information; human; hands; x-ray; CT; care

Introduction

Volumetric imaging with modalities like magnetic resonance imaging (MRI) or x-ray computed tomography (CT) has been a valuable tool in many areas of clinical, preclinical, and basic research. Image files are data rich, often with metadata containing (or linked to) both *personal* and *sensitive* data like the name of the participant or diagnoses. When it comes to sharing these data for research, current ethics consensus and legislation require all data that can be used to link the shared data set directly or indirectly to the participant to be carefully removed [1]. This is of high importance as the deidentified data sets ensure the protection of participant confidentiality. However, in some cases, the information that can link the participant to the data is not simply accompanying the data as metadata. It *is* (part of) the data as, for example, in the case of head imaging where the data set can be used to reconstruct an image of participants' faces at a sufficient quality that can be then used for the reidentification of the participants [2].

However, the head is not the only site from where unique biometric information can be extracted. Perhaps the most widely used biometric information for human individualization is the friction ridge pattern on the fingertips, often referred to as fingerprints [3]. As with the face in head imaging, friction ridges are also recorded as part of the imaging data when a participant's hand is imaged. High-resolution x-ray CT imaging modalities such as high-resolution peripheral quantitative CT and microfocus CT (µCT) are used to image human hands for accessing the trabecular structure or various pathological changes (eg, osteoarthritis) with a voxel size <100 µm [4]. Here we show for the first time that the spatial resolution provided by these technologies enables retrieval of the friction ridge pattern.

Methods

A fixed cadaveric human right hand from a human body donor was imaged using a µCT system designed for 3D x-ray histology [5] at the 3D x-ray histology laboratory [6] at the University of Southampton, United Kingdom. Imaging was conducted using an isotropic voxel (edge) size of 72 µm at 160 kVp, maintaining an electron beam spot size <15 µm (Table 1). Image processing was conducted using Fiji/ImageJ [7] and volumetric renderings using VG Studio Max (v2.1.4 64-bit, Volume Graphics GmbH). Extraction of the 2D net of the ridge pattern was performed in Fiji/ImageJ, and it involved a workflow to isolate the ridge pattern–containing voxels of the outer surface of the finger, that is, the epidermis, the dermis, and part of the subcutis layer, and a method to project the pattern onto a 2D plane, which is

XSL•FO
RenderX

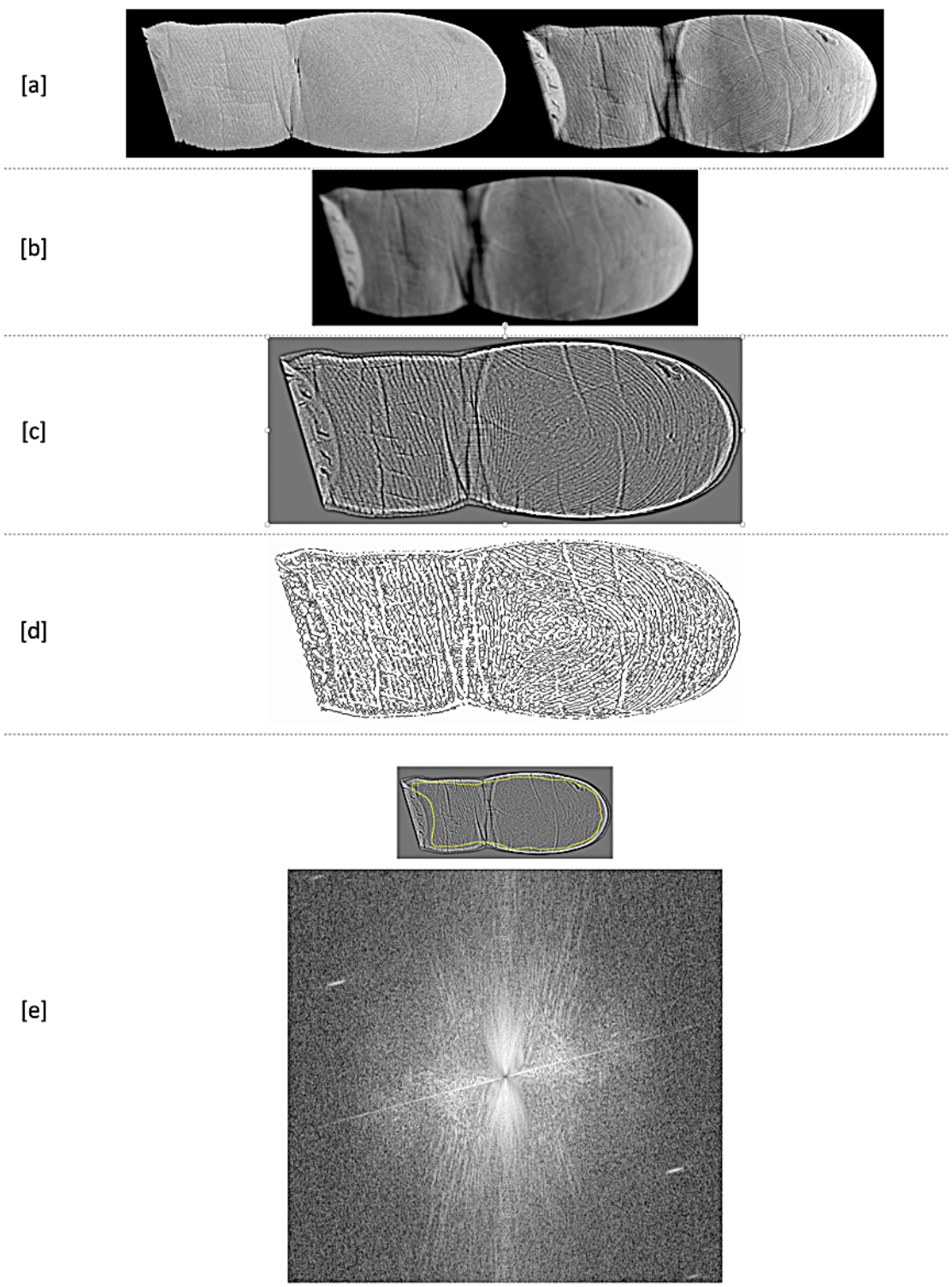
outlined in Figure 1 (see Multimedia Appendix 1 for more details).

Table 1. Microfocus computed tomography imaging parameters.

Scanner used	X-ray histology scanner (Custom Nikon XT H 225 ST)
Approximate total scan time (h:min:s)	0:23:37
Detector binning	1 × (2850 × 3850 dexels)
Target material	Tungsten (W)
Acceleration voltage (kVp)	160
Current (µA)	112
Power (W)	17.92
Angular projections	2001
Frames per projection	4
Exposure per frame (ms)	177
Analog gain (dB)	24
Filter material	None
Voxel size (µm)	72.00 (isotropic)

XSL•FO
RenderX

Figure 1. Fiji/ImageJ processing workflow for extracting the friction ridge information from the processed volume resulting from the workflow described in Multimedia Appendix 1. (A) Maximum intensity projection of the skin layer (left) and SD projection (right), (B) blurred image of the SD projection, (C) edge image resulting from the subtraction of the blurred and SD image, (D) skeletonized image of (C), (E) frequency domain Fourier image (bottom) of the selected friction ridge area shown in the insert image (inner area of outlined region) (Multimedia Appendix 2).



Ethical Considerations

The study was performed in accordance with the University of Southampton's ethics policies and ethical guidelines (ERGO/FEPS/67396). The sample was obtained by cadaveric donors who have given consent for their bodies to be used for scientific research and imaging.

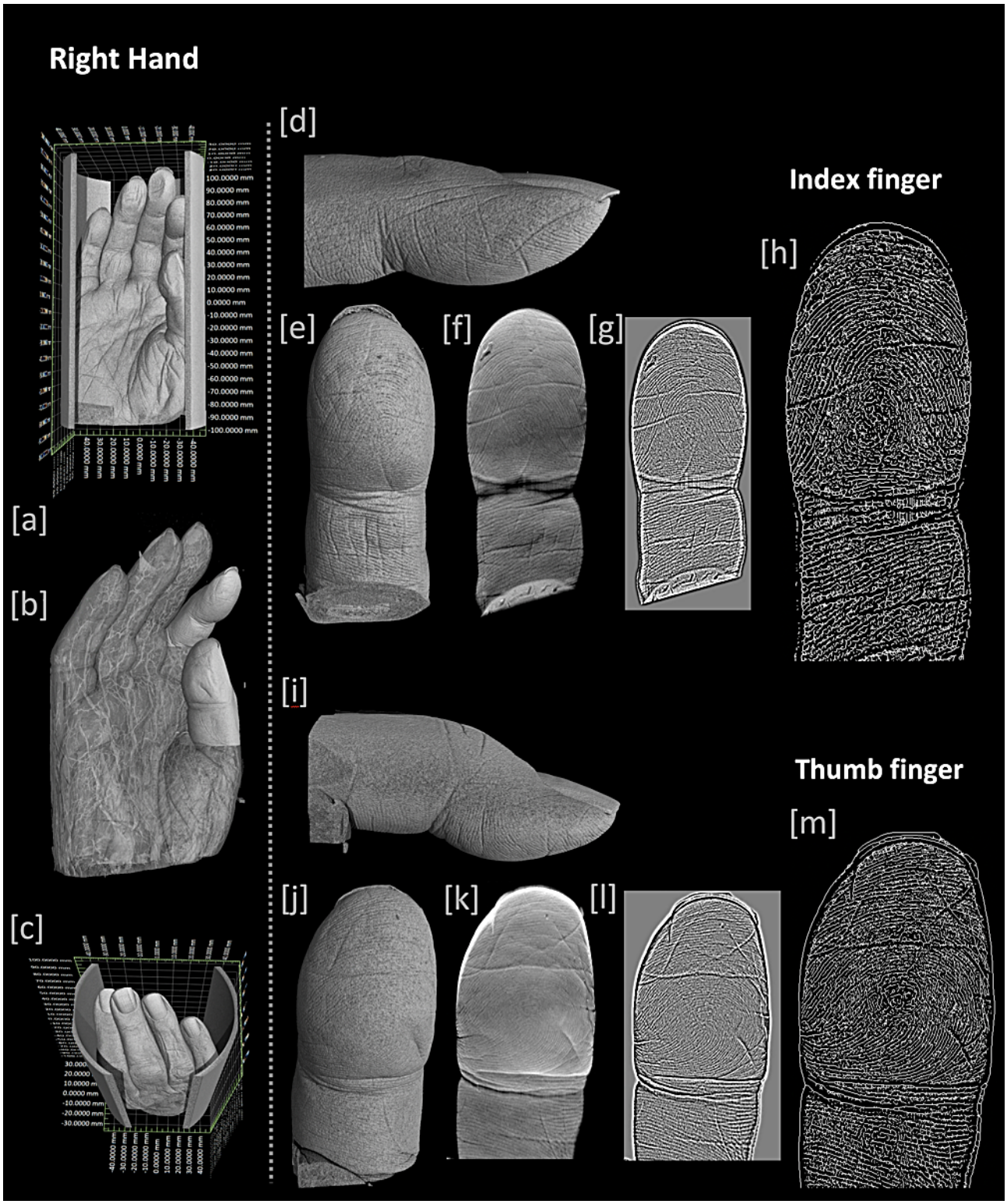
Results

Following the workflow described above and in Multimedia Appendix 1, one is left with a single volume per finger that contains only the volumetric gray-level information of the friction ridges, which can now be projected onto a single 2D

XSL•FO
RenderX

plane so that the net of the friction ridge pattern can be recorded and further processed (Figures 1 and 2). Two examples of further processing are shown in Figure 1D and E, where a skeleton (1D) of the Fourier (frequency domain) image (1E) of the edge image is presented.

Figure 2. Cadaveric human right hand imaged for the needs of the Anatomically Precise Revolutionary Implant for Bone Conserving Osteoarthritis Treatment (APRICOT) project using a high-resolution x-ray microfocus computed tomography system at the x-ray histology laboratory, University of Southampton. Left pane (A-C): volume renderings of the hand depicting the positioning and the segmentation of the index and thumb fingers for further analysis. Right pane: (D,I) top and (E,J) bottom view of volumetric rendering of the segmented finger volumes. (F,K) SD projection across the segmented volume after the removal of all information deeper than 10 voxels from the surface (skin). (G,L) Edge image of (F,K) and (H,M) skeleton of the binarized (G,L) images (Multimedia Appendix 3).



XSL•FO
RenderX

Discussion

µCT imaging was sufficient to resolve the friction ridges of the fingertips. In silico reconstructed fingerprints of the thumb and the index fingers using the whole hand µCT data set are shown in Figure 2. With fingerprint-based identification being widely used outside forensics in a range of applications such as accessing personal accounts and devices, the risk of reidentification, impersonation, and password compromisation should be considered and discussed in an ethics application. Regarding the latter point, imaging scientists should be made aware that processing of such data might potentially have implications on data protection obligations; since biometric data are used for identification (eg, authentication), they become *special category personal data* (General Data Protection Regulation, article 9-1) [8]. Our study complements previous imaging studies that assessed the ethical implication associated with the risk of reidentification of the participants using MRI data [2] and highlights the risks of modern high-resolution imaging modalities, such as x-ray µCT, accidentally producing identifiable information. We argue that data sets containing high-resolution CT images of the human hands should be considered "sensitive" and thus handled and shared with appropriate care, as data subjects and imaging scientists may not be able to identify a priori all implications of further processing [9]. Relevant bodies (eg, institutional ethics committees or data protection committees) should consider this aspect of high-resolution CTs when reviewing research or clinical imaging protocols and make their recommendations according to the currently applicable law and local code of practice.

Acknowledgments

The work was supported by European Union Horizon 2020 grant 863183, Wellcome Trust grant 212940/Z/18/Z, and the NVIDIA Corporation who donated the RTX A6000 used to generate all images. OLK and BJP are a member and chair, respectively, of the University of Southampton's research ethics committee. We are grateful to Dr David Walker and Ms Ellen Adams at the Anatomical Sciences Laboratory, Centre for Learning Anatomical Sciences at the University of Southampton for providing the specimen.

Data Availability

Data generated during the study are available from the corresponding author by request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information.
[DOCX File , 2813 KB - jmir_v24i11e38650_app1.docx]

Multimedia Appendix 2
Supplementary Figure 1.
[PNG File , 3710 KB - jmir_v24i11e38650_app2.png]

Multimedia Appendix 3
Supplementary Figure 2.
[PNG File , 3493 KB - jmir_v24i11e38650_app3.png]

References

1. GDPR and research – an overview for researchers. UK Research and Innovation. 2020. URL: https://www.ukri.org/about-us/policies-standards-and-data/gdpr-and-research-an-overview-for-researchers/ [accessed 2021-12-31]
2. Schwarz CG, Kremers WK, Therneau TM, Sharp RR, Gunter JL, Vemuri P, et al. Identification of anonymous MRI research participants with face-recognition software. N Engl J Med 2019 Oct 24;381(17):1684-1686 [FREE Full text] [doi: 10.1056/NEJMc1908881] [Medline: 31644852]
3. Holder EH, Robinson LO, Laub JH. The Fingerprint Sourcebook. Office of Justice Programs. 2011. URL: https://www.ojp.gov/pdffiles1/nij/225320.pdf [accessed 2022-11-18]
4. Ohs N, Collins CJ, Atkins PR. Validation of HR-pQCT against micro-CT for morphometric and biomechanical analyses: A review. Bone Rep 2020 Dec;13:100711 [FREE Full text] [doi: 10.1016/j.bonr.2020.100711] [Medline: 33392364]
5. Katsamenis OL, Olding M, Warner JA, Chatelet DS, Jones MG, Sgalla G, et al. X-ray micro-computed tomography for nondestructive three-dimensional (3D) x-ray histology. Am J Pathol 2019 Aug;189(8):1608-1620 [FREE Full text] [doi: 10.1016/j.ajpath.2019.05.004] [Medline: 31125553]

XSL•FO
RenderX

6. 3D x-ray histology. University of Southampton. 2022. URL: https://www.xrayhistology.org [accessed 2022-11-18]
7. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 2012 Jun 28;9(7):676-682 [FREE Full text] [doi: 10.1038/nmeth.2019] [Medline: 22743772]
8. 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). EUR-Lex. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2022-11-18]
9. Pickering B. Trust, but verify: informed consent, AI technologies, and public health emergencies. Future Internet 2021 May 18;13(5):132. [doi: 10.3390/fi13050132]

Abbreviations

CT: computed tomography
MRI: magnetic resonance imaging
µCT: microfocus computed tomography

Edited by R Kukafka; submitted 11.04.22; peer-reviewed by O Arthurs, M Roguljić; comments to author 24.07.22; revised version received 04.08.22; accepted 02.09.22; published 23.11.22.

Please cite as:
Katsamenis OL, Burson-Thomas CB, Basford PJ, Pickering JB, Browne M
The Reconstruction of Human Fingerprints From High-Resolution Computed Tomography Data: Feasibility Study and Associated Ethical Issues
J Med Internet Res 2022;24(11):e38650
URL: https://www.jmir.org/2022/11/e38650
doi:10.2196/38650
PMID:36416872

©Orestis L Katsamenis, Charles B Burson-Thomas, Philip J Basford, J Brian Pickering, Martin Browne. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Role of Information Infrastructures in Scaling up Video Consultations During COVID-19: Mixed Methods Case Study Into Opportunity, Disruption, and Exposure

Joseph Wherton[1], PhD; Trisha Greenhalgh[1], PhD; Gemma Hughes[1], DPhil; Sara E Shaw[1], PhD

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

Corresponding Author:
Joseph Wherton, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Oxford, OX2 6GG
United Kingdom
Phone: 44 865 617855
Email: joseph.wherton@phc.ox.ac.uk

Abstract

Background: Until COVID-19, implementation and uptake of video consultations in health care was slow. However, the pandemic created a "burning platform" for scaling up such services. As health care organizations look to expand and maintain the use of video in the "new normal," it is important to understand infrastructural influences and changes that emerged during the pandemic and that may influence sustainability going forward.

Objective: This study aims to draw lessons from 4 National Health Service (NHS) organizations on how information infrastructures shaped, and were shaped by, the rapid scale-up of video consultations during COVID-19.

Methods: A mixed methods case study of 4 NHS trusts in England was conducted before and during the pandemic. Data comprised 90 interviews with 49 participants (eg, clinicians, managers, administrators, and IT support), ethnographic field notes, and video consultation activity data. We sought examples of infrastructural features and challenges related to the rapid scale-up of video. Analysis was guided by Gkeredakis et al's 3 perspectives on crisis and digital change: as opportunity (for accelerated innovation and removal of barriers to experimentation), disruption (to organizational practices, generating new dependencies and risks), and exposure (of vulnerabilities in both people and infrastructure).

Results: Before COVID-19, there was a strong policy push for video consultations as a way of delivering health care efficiently. However, the spread of video was slow, and adopting clinicians described their use as ad hoc rather than business as usual. When the pandemic hit, video was rapidly scaled up. The most rapid increase in use was during the first month of the pandemic (March-April 2020), from an average of 8 video consultations per week to 171 per week at each site. Uptake continued to increase during the pandemic, averaging approximately 800 video consultations per week by March 2021. From an opportunity perspective, participants talked about changes to institutional elements of infrastructure, which had historically restricted the introduction and use of video. This was supported by an "organizing vision" for video, bringing legitimacy and support. Perspectives on disruption centered on changes to social, technical, and material work environments and the emergence of new patterns of action. Retaining positive elements of such change required a judicious balance between managerial (top-down) and emergent (bottom-up) approaches. Perspectives on exposure foregrounded social and technical impediments to video consulting. This highlighted the need to attend to the materiality and dependability of the installed base, as well as the social and cultural context of use.

Conclusions: For sustained adoption at scale, health care organizations need to enable incremental systemic change and flexibility through agile governance and knowledge transfer pathways, support process multiplicity within virtual clinic workflows, attend to the materiality and dependability of the IT infrastructure within and beyond organizational boundaries, and maintain an overall narrative within which the continued use of video can be framed.

(J Med Internet Res 2022;24(11):e42431) doi:10.2196/42431

XSL•FO
RenderX

KEYWORDS

information infrastructure; video consultations; organizational ethnography; COVID-19; health care organization; telehealth; health care delivery; agile governance; knowledge transfer

Introduction

Video Consulting Before, During, and Beyond the Pandemic

There has been growing interest in the use of video as a method of consultation between clinician and patient over the past 10 years, and numerous studies have shown such consultations to be acceptable, safe, and effective in selected patients [1-6]. However, until the pandemic, the uptake of video in the National Health Service (NHS), as in health care organizations in many countries, was slow. Sustained adoption at scale requires new ways of organizing clinical and administrative roles, new organizational routines, new approaches to privacy and information governance, and new clinical and communication skills. However, in published research trials of video consultations, the services were usually available only as part of the trial and discontinued thereafter, so the challenges of embedding video in business as usual were never addressed [7].

COVID-19 created a "burning platform" for the mainstreamed use of such services, as health care organizations worldwide halted face-to-face appointments for nonurgent care. The global emergency prompted strategic (policy decisions and legal changes), operational (increasing capacity and delivery by building skills and resources at pace and scale), and regulatory (eg, pandemic-related unofficial workarounds with unregulated products) changes with regard to the delivery of telehealth across different national contexts [8-10]. We initially flagged the pandemic as an "opportunity in a crisis" for giving video consulting the push it needed [11].

During the first wave of the pandemic in 2020, most countries saw a rapid reduction in face-to-face consultations and an increase in remote ones in both primary and secondary care [12-15]. Guidance for conducting video consultations were initially produced based on early examples of implementation [16-22]. More definitive guidance was later developed, based on learning during the pandemic, with the view of informing best practices beyond COVID-19 [23-26].

As health care organizations are looking to expand and maintain the use of video in the postpandemic "new normal," it is important to acknowledge and harness wider system changes that have emerged. In this paper, we focus on the infrastructural aspects of the rapid scale-up of video and draw lessons for ongoing developments and sustainability. To this end, we explore the impact of the pandemic in 4 NHS organizations, where we had previously identified ways in which infrastructural features influenced the limited use of video in the years preceding the pandemic [27].

Attending to the Health IT Infrastructure

Star [28] defines infrastructure as the stuff other things run on. It consists of hardware and software, the buildings, wires, connections, clinical records, organizational routines, standards, and other aspects that make an information system work. A defining characteristic of infrastructure is its transparency (invisible, taken for granted, and ready to hand), and so often, it is not considered within health technology projects until it breaks down or gets in the way. Health IT systems are also patch-worked and path-dependent, in which components emerge incrementally and so cannot be replaced wholescale. Star's [28] conceptualization of infrastructure challenged the common view of technology as stand-alone artifacts, emphasizing the situated and practical work of implementation. Technology cannot be merely designed and installed, but instead, it must emerge from, and build on, the *installed base* (ie, on systems and practices already in place).

In the years before COVID-19, we conducted ethnographic research in 4 NHS organizations in England, all of which were seeking to introduce or scale up video consultations [7,27,29]. The case studies varied in size, geography, patient population, and the use of digital technology. Common across all was the way in which the "boring things" of infrastructure (eg, internal procedures; locally endorsed standards; aspects of software functionality; mundane administrative issues, such as room bookings; general pressures on the system; and interoperability and compatibility issues across new and legacy systems) greatly influenced the fortunes of this initiative.

In our application and retheorization of Star's [28] notion of infrastructure, we identified several infrastructural features related to the limited uptake and use of video: (1) intricacy and lack of dependability of the installed base; (2) interdependencies of technologies, processes, and routines; (3) the inertia of established routines; (4) the constraining (and, occasionally, enabling) effect of legacy systems; and (5) delays and conflicts relating to clinical quality and safety standards [27].

A crisis is a low-probability, high-impact event that threatens social and life-sustaining systems, creates deep uncertainty, and requires international and government intervention [30]. Gkeredakis et al [31] applied 3 different perspectives to shed light on the varied uses of digital technology, and associated tensions, during the COVID-19 crisis: *opportunity* for accelerated innovation and removal of barriers to experimentation; *disruption* to organizational and occupational practices, generating new dependencies and risks; and *exposure* of vulnerabilities in both people and infrastructures that have previously gone unnoticed. Gkeredakis et al [31] observed that although the pandemic accelerated and expanded the use of digital technology, these shifts were "fast-paced, dramatic and not well understood."

In this paper, we provide longitudinal qualitative and quantitative data (spanning before and during the pandemic) on 4 NHS organizations to understand infrastructural changes during the rapid scale-up of video consultations and how this may influence sustainability going forward. All 4 sites saw a significant reduction in face-to-face appointments, and immediate and substantial increase in remote ones (overall a 245-fold increase in video consultations), as part of a

XSL•FO
RenderX

systemwide response to the pandemic in 2020. Against this background, we sought to study how health information infrastructures shaped, and were shaped by, the rapid scale-up. In the remainder of this paper, we first describe the national context and aims of the study. In the Methods section, we describe the study setup and our analytical approach to understanding infrastructural features and challenges. We then describe our findings on crisis-engendered change as a time of opportunity, disruption, and exposure. Finally, we discuss the findings in the context of the wider literature, highlighting learning points for scaling up and sustaining the use of video beyond the pandemic.

National Context

Since 2010, there had been a growing policy emphasis on digital innovation and remote care in England [32-34]. Prepandemic, this provided impetus for innovation-driven change. However, the adoption of video consulting was slow, time-consuming, and resource intensive, with activity confined to specific clinical services and settings (typically with a local clinical enthusiast leading) [7,35,36].

In 2019, the year before COVID-19, NHS England and NHS Improvement (NHSEI, the national implementation arm of the NHS in England) set up several pilot video consulting services in secondary care using a video platform called Attend Anywhere (building on the learning of a similar program in Scotland) [37]. Attend Anywhere is an internet browser–based video technology that can be accessed by a staff member on a work computer or a member of the public using their own device. One defining feature is its inbound workflow, which seeks to emulate how patients attend in-person appointments. For example, a single button on a website (or a consistent weblink address on an appointment letter) offers a one-stop virtual front door for patients. On clicking that link, the patient enters a virtual waiting area, before being invited into the clinician's virtual consulting room.

When the pandemic hit in March 2020, the use of remote consultations (phone and video) formed a key element of the national response [38]. Building on the national pilot of Attend Anywhere, central procurement of the software was extended to provide all NHS trusts unlimited use of the software for 12 months. Staff training and materials to support swift deployment were made available through the NHS England website [39]. National, regional, and interorganizational materials were also shared through the FutureNHS platform (a virtual networking platform for health and social care staff). Within 3 weeks of the pandemic, the NHSEI established 7 regional implementation teams (to provide temporary setup support, webinars, and peer learning) and a national helpdesk (leveraging NHS England's existing IT call center). Alongside training and operational resources, every NHS trust in England was provided £20,000 (US $ 23120.40) capital funding to purchase video call equipment, and 5000 iPads (Apple Inc) were rapidly sourced and distributed for frontline staff.

NHSX (a cross-department unit within the NHS with responsibility for setting the national policy on digital and data management) advised on governance reviews of video platforms to facilitate organizational approvals, as well as the negotiation of zero-rated 4G with major mobile network providers (so patients could use Attend Anywhere without incurring mobile broadband charges).

The provision of Attend Anywhere across England was further supported by recent insights from prepandemic scale-up of Attend Anywhere in Scotland. In addition, the temporary use of the Scottish platform servers during the initial response allowed for immediate setup of the trust's Attend Anywhere accounts before they were transferred to a dedicated server.

Across England, the proportion of remote consultations (phone and video) surged from just 3.9% of all outpatient activity in January-February 2020 to 36.6% by the end of April 2020. Over the course of 1 year (April 2020-March 2021), remote consulting accounted for, on average, 28% of outpatient activity (with a proportionate drop largely due to a gradual increase in the number of in person appointments) [40]. Despite this increase in remote consultations, video constituted a relatively small proportion of the appointment activity, in which it made up approximately 2.4% of overall secondary care consultations, although with much variation in uptake across settings and specialties [41]. However, this presented an unprecedented shift in the scale of video consultations, which is considered to have a long-lasting role going forward [42]. For further details on the varied UK context, see Shaw et al [43].

The aim of this study was to draw lessons from 4 NHS organizations on how information infrastructures shaped, and were shaped by, the rapid scale-up of video consultations during COVID-19. Our research question was, "How did information infrastructures shape, and become shaped by, the rapid implementation and scaling-up of video consultations during the pandemic, and what does this mean for scale-up and sustainability going forward?"

Methods

Study Design

A naturalistic case study with an action research component was conducted in 4 NHS trusts in England, which we refer to by their pseudonyms: Petroc, Eastern, Southern, and Northern Trusts. All were seeking to introduce and scale up video consultations before and during the pandemic as part of a service improvement program, in which members of the Petroc team supported clinicians and managers in the 3 other trusts. Data sources included ethnographic field notes, interviews, consultation activity data, service evaluation reports, documents, and material artifacts.

Ethical Considerations

Research ethics approval was obtained from London – Camberwell St. Giles Research Ethics Committee (ref. 19/LO/0550). An advisory group was established from the start to oversee both phases of the project, with wide stakeholder representation (eg, policy makers, organizational stakeholders, and patient groups) and a lay chair. The group provided input on study progress during 6-monthly meetings, as well as comments on project outputs by phone and email.

XSL•FO
RenderX

Data Collection and Management

Data were collected in 2 phases, before and during the pandemic (the periods and data sources are presented in Table 1). The prepandemic phase extended data previously reported up until when COVID-19 was confirmed in the United Kingdom (January 2018-February 2020), providing context to the impact of the pandemic. The in-pandemic data were collected during March 2020-July 2021.

In total, 90 interviews were conducted with 49 participants, including doctors, nurses, allied health professionals (AHPs), service managers, and admin and IT support. This included 43 (48%) interviews with 29 (59%) participants prepandemic and 47 (52%) interviews with 37 (41%) participants during the pandemic. In addition, 17 (35%) participants were interviewed in both phases, and 6 (12%) key informants were interviewed on multiple occasions within each phase.

Fieldwork was conducted in person before the pandemic but was, of necessity, conducted remotely during the pandemic. Interviews lasted between 30 and 60 minutes. Participants were asked to talk about their experience of supporting or using video consultations (or why they had chosen not to support/use this medium), including the progress of and the challenge in using (or supporting the use of) video and the impact of the pandemic. When interviewees talked in the abstract about problems and challenges, we asked them to describe specific examples of these.

Our data set of qualitative interviews was supplemented by evaluation data captured by local teams (eg, aggregated data on patient experience surveys and demographics), internal audits (staff engagement reports), policy documents (eg, digital strategy documents, information governance, service recovery plans), and standard operating procedures and training resources (eg, implementation procedures, guidance materials).

Video consultation activity was captured during March 2020-March 2021 as part of a national NHS England pilot using Attend Anywhere, which was the main video platform used within the sites.

Table 1. Data sources for the 2 phases of the evaluation.

	Phase 1 (prepandemic)	Phase 2 (during the pandemic)	Total
Period of data collection	January 2018-February 2020	March 2020-July 2021	3 years, 7 months
Ethnographic observation (hours)	180 hours	No ethnography possible due to the pandemic	180 hours
Interviews by site			
Petroc	n=22 (54%)	n=19 (46%)	41
Eastern	n=5 (38%)	n=8 (62%)	13
Southern	n=12 (52%)	n=11 (48%)	23
Northern	n=4 (31%)	n=9 (69%)	13
Interview participants	• 7 doctors • 5 nurses • 3 AHPs[a] • 8 managers • 6 admin/IT staff	• 10 doctors • 3 nurses • 7 AHPs • 8 managers • 9 admin/IT staff	• 12 doctors • 6 nurses • 8 AHPs • 10 managers • 13 admin/IT staff
Uptake statistics for video consultations, by NHS[b] trust	July 2019-February 2020	March 2020-March 2021	21 months
Online patient survey reports	Postconsultations, user experience survey (Petroc Health)	Postconsultations, user experience survey (all 4 sites)	5 patient surveys (N=4050)

[a]AHP: allied health professional.

[b]NHS: National Health Service.

Analytical Approach

Interviews, field notes, and supporting materials collected during the pandemic phase were first used to gain familiarity with each case site and produce an organizational narrative on the impact of COVID-19. Following this familiarization phase, interview transcripts were then organized into a Microsoft Excel spreadsheet to identify emerging themes within Gkeredakis et al's [31] 3 perspectives on the use of digital technology in a crisis: as an *opportunity* (for the implementation, uptake, and use of video), as a *disruption* (to the existing working practice and the need to adapt), and as *exposure* (issues that had previously gone unnoticed or underplayed).

Analysis of emerging subthemes was guided by Star's [44] ethnographic approach to the study of information infrastructure to surface master narratives (the overarching discourses that shape decisions), infrastructural inversion (eg, foreground things that are usually kept in the background), surface invisible work (eg, work done by low-grade staff, such as secretaries and administrators), and study paradoxes (eg, why a simple change makes the whole system unworkable, perhaps because it generates additional hidden work) [44].

Analysis was further guided by the literature on crisis management [45] and routine dynamics [46]. Christianson et al [45] highlighted how the novel demands of *rare interruptions*

XSL•FO
RenderX

trigger organizational learning and reveal underlying weaknesses and potential strengths as events unfold. This includes organizational actors engaging in the process of *interpreting* (which initiates the conditions that guide responsive action), *relating* (when members of the organization understand their contribution to the overall outcome), and *restructuring* (reconfiguring social and materials structures to maintain core organizational functions). In relation to the latter, we also drew on the notion of organizational routines, defined as "recognisable, repetitive patterns of interdependent action carried out by multiple actors" [46]. Organizational routines coordinate work, reduce uncertainty, and are situated within a sociomaterial context, structured around time, physical spaces, and material and technological artifacts [47].

Pentland and Feldman [46] distinguish between *ostensive* routines (abstract understanding about how it is enacted) and *performative* routines (the range of ways in which it is carried out in practice). They propose the use of *narrative networks* to represent the generative tension between these 2 aspects of organizational routines and how this is mediated by *technology in use* [48]. Unlike other graphical representations of work, such as process mapping, narrative networks summarize the relations between actions as a networked graph of nodes (categories of actions) and linked edges (indicating sequential relations between the actions), providing a network of potential performances or "stories" within a process. We applied this approach to help us understand how video was used and embedded within routine practice.

Results

Case Site Overview of Rapid Scale-Up

Petroc Health is a multisite acute hospital trust located in a predominantly deprived and multiethnic part of London. It is 1 of the largest trusts in England, serving a population of 2.5 million. Since 2013, Skype had been used within a diabetes clinic to reduce did-not-attend rates. Building on the success of this pilot, the digital strategy team commenced a trustwide program in 2018 to spread use across outpatient services. However, progress was hampered by technical problems with Skype, including incompatibility with the new virtual desktop infrastructure (VDI, a virtualization technology that hosts a desktop operating system on a centralized server). Petroc joined the NHSEI Attend Anywhere pilot in October 2019, and 7 clinical services started using the platform by the time COVID-19 hit in March 2020. As part of the pandemic response, other members of the digital strategy team were drafted in to support scale-up (eg, training, deployment), with oversight from a cross-departmental covid executive group.

Southern Trust is a large multisite provider in a university city that had won awards as a digital innovator. It includes 4 hospitals serving a population of 655,000, with a relatively high proportion of young people aged 20-24 years (including university students). As part of the prepandemic scale-up program, 2 services (diabetes and orthopedics) started using Skype for Business. Although this software was supported by the trust network, licensing permissions and firewall restrictions created problems conducting video calls with people outside of the organization (eg, patients). A small group of clinicians sought to engage the information and communication technology (ICT) department into resolving these issues, but little progress was made due to other IT priorities. In July 2019, Southern Trust joined the NHSEI Attend Anywhere pilot, in which 4 services started using the platform. When the pandemic hit, the chief digital officer coordinated a trustwide deployment, with the support of 5 divisional digital leads to manage technical setup (eg, the provision of iPads to run video via secure Wi-Fi).

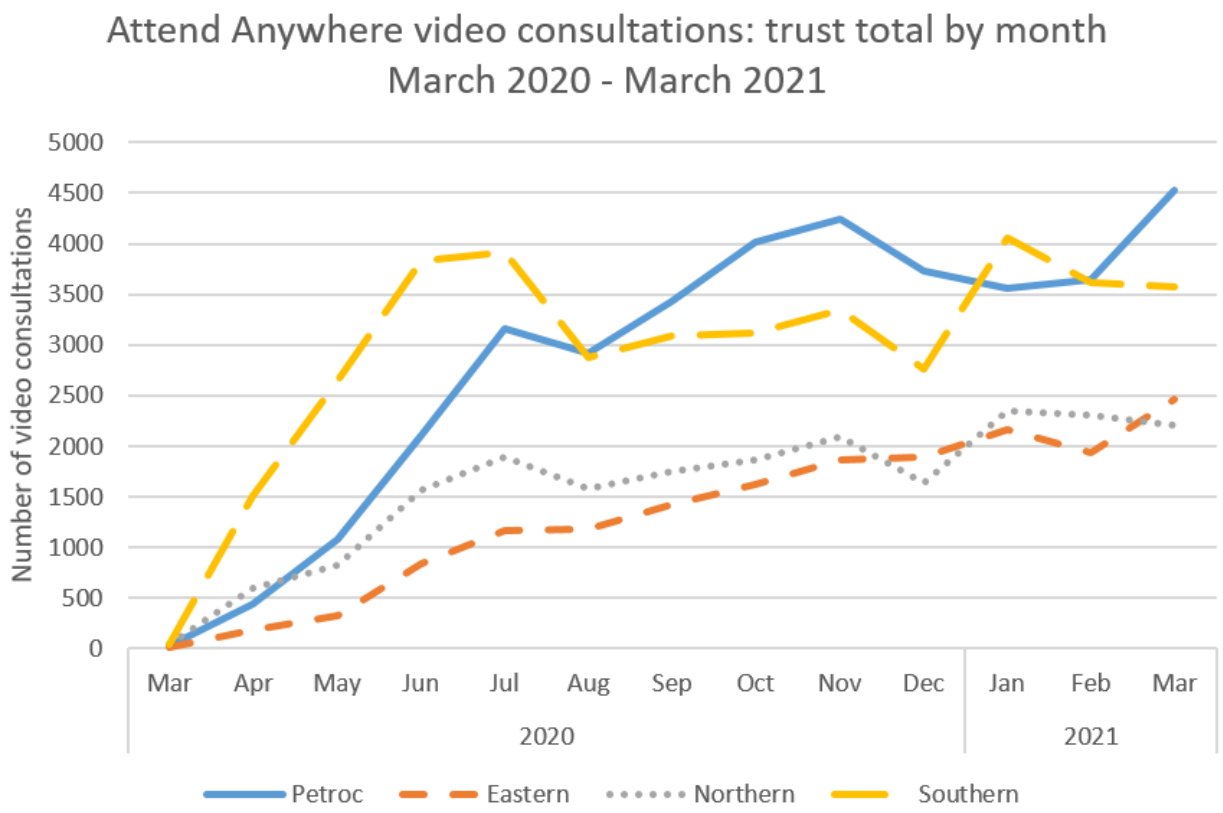
Eastern Trust runs 2 hospitals in a largely rural county. The catchment population of 1 million is predominantly White British, with a relatively high proportion of patients over 65 years. There was a prepandemic strategy to digitize patient records. However, the electronic patient record system had not yet been rolled out, and so the large majority of services still relied on paper-based records. A diabetes consultant's initial attempts to pilot the use of Skype, as part of the prepandemic scale-up, were halted by the ICT department due to governance concerns. The trust joined the NHSEI Attend Anywhere pilot in November 2019, but none of the services started using the software until COVID-19. This meant extensive work was needed to prepare clinicians and administrative teams using the software, with training from the regional support team set up by the NHSEI.

Northern Trust provides hospital and community services across a built-up metropolitan borough, as well as small towns and rural areas. Although geographically the largest catchment area, it has the smallest patient population of approximately 500,000. The population is predominantly White British, with socioeconomically diverse regions, including areas of high deprivation, unemployment, chronic illness, substance misuse, and mental health problems. As an early participating site for the NHSEI Attend Anywhere pilot, prepandemic use of the video software began in July 2019, with a focus on reducing patient travel and strong senior management buy-in. By the time the pandemic hit, Attend Anywhere was being used in 6 clinical services. The implementation team continued to lead the deployment, with additional operational and training support brought in from the IT and business units.

Figure 1 shows the number of video consultations for the 4 sites during the period of March 2020-March 2021. The most rapid increase in the use of video was seen during the first month of the pandemic (March-April 2020), from an average of 8 video consultations per week to 171 per week. Uptake continued to increase over the course of the pandemic, averaging approximately 800 video consultations per week by March 2021. As was observed nationally, video still constituted a relatively low proportion of overall outpatient activity, in which most remote appointments were conducted by phone.

XSL•FO
RenderX

Figure 1. Monthly video consultations across the 4 trusts during March 2020-March 2021.



In sum, the case studies present contextually different circumstances in relation to the geography and technical and organizational contingencies but show similar trends in the rapid scale-up of video consultations. The following sections highlight common themes in relation to how the information infrastructures shaped, and became shaped by, the rapid scale-up.

Opportunity

Viewing the pandemic as an opportunity for positive change highlights how the crisis helped accelerate processes that were stalling, questioned institutional norms and processes, and allowed for experimentation.

Before COVID-19, efforts to implement video were restricted by (to a varying degree across sites) pressures on human and financial resources, competing operational and strategic priorities, differing institutional logics (beliefs, assumptions, and practices that shape actions [49]), problems interfacing the new technology with local legacy systems and standards, and training and support requirements for staff and patients. As the project manager within Petroc expressed shortly before the pandemic:

> *It's difficult to make a financial case for video consultations, because we're not reducing any activity in the system, but we are requesting new technology…And it is difficult timing for the trust to take on something that doesn't even pay for itself…You need the senior leadership to see it as the next priority, but of course, there are hundreds of priorities. It is really challenging. It doesn't happen overnight…* [Bridget, program manager, Petroc Health, January 2020]

The prepandemic focus on the economic case echoed the dominant policy discourse that viewed digital solutions primarily in terms of enhancing efficiency [33]. Other confounding logics included frontline duty of care (professional standards of what excellent care looks like and any potential risks of harms the digital medium might bring), IT regulatory concerns on data security and service quality, and managerial focus on capacity and resource to support the change. In this context, the translational efforts were typically driven by local clinical champions' ability to define and frame the problem for which video was to be a solution, engage interest, and mobilize other key organizational actors to support the initiative [27]. Broadly speaking, such efforts were rarely successful and even more rarely sustained long term. Video consultations could be established as small-scale demonstration projects but almost never gained momentum as business as usual.

However, in the wake of the pandemic, a national mandate to avoid direct clinician-patient contact on the grounds of safety (infection control) engendered collective action toward the rapid rollout, as the same project manager described:

> *All of a sudden, we have a level of support we never had before. Like from a very senior level. As soon as this [pandemic] happened, they put in temporary governance structures—the Pandemic Covid Outpatient Group—headed up by one of the most senior doctors in the trust. I've moved full-time on this and report to them. So, if I have any problems, I*

XSL•FO
RenderX

> *have someone to escalate to, who can do something about it…Also, the clinical system team [responsible for building electronic clinic schedules] usually take[s] weeks to implement changes. But I've been given permission to, basically, get in touch with them to say, "This is a priority."* [Bridget, program manager, Petroc Health, April 2020]

These new relations illustrate how a shared *organizing vision* (clear and consistent vision among stakeholders as to what will be achieved) [50] brought legitimacy and support for the change, in which staff and resources were mobilized to address infrastructural constraints. As illustrated in the extract earlier, provisional management and communicative structures helped hold groups together around this narrative and increase awareness of their role within the context of the collective outcome.

An opportunity was also seen within the temporary suspension of regulative structures, allowing staff to bypass administrative burden and governance processes, which had previously thwarted attempts to introduce video:

> *Things kind of became, to an extent, easier to do. Previously, you would have a lot of bureaucratic hurdles and hoops to jump through. But we have come to a point where things just had to be done.* [Nathan, operational support, Northern Trust, March 2021]

This brought a welcome cultural shift, providing greater autonomy and openness to change. In accordance with Scott's [51] 3 interacting institutional forces that operate in health care, an easing of risk management protocols (*regulative*) enabled more agile, unformalized (*cultural-cognitive*) governance, underpinned by professional (*normative*) values:

> *Before [COVID-19], we had hearing aids where you could remotely program them, and I looked into setting up a remote programming clinic…I went to information governance, and they gave me this list, and I was like, "Oh it's not worth it—not for a few patients." So, I didn't bother. And that's terrible—I should have bothered. But it was too much work. But now, when we asked them, it went straight through to the medical director, and she just said, "Go ahead with it…" So, things progressed—which is what the NHS would struggle with sometimes.* [Karen, consultant, Southern Trust, November 2020]

As described before, regulatory constraints were also reduced through national measures by NHS England, including the central procurement of Attend Anywhere (enabling unrestricted use with no cost to trusts), endorsement of video platforms by national information governance bodies (enabling experimentation), and temporary tariffs for remote consultations (avoiding local commissioning requirements and associated system configurations).

However, although these were intended as *temporary* arrangements to deal with the pandemic, they set organizations on a particular infrastructural path, upon which they were now destined to build:

> *One challenge is, few of us want to change what we've got now [Attend Anywhere]. We will have to go through this procurement—because the region will be doing it on our behalf—and there is a slim chance another provider will be chosen, and that would be a real challenge for us.* [Teresa, program manager, Eastern Trust, February 2021]

Finally, the pandemic created an opportunity for clinician engagement. Before the pandemic, there was a striking difference between those who embraced the use of video with enthusiasm and other clinicians on the same teams who were reluctant or could not see the benefits of the change. Hurdles to adoption were less to do with needing to learn to use the technology and more related to professional concerns about patient safety, quality of care, and identity. However, during the pandemic, the perceived *relative advantage* [52] of video compared to existing alternatives meant that it was no longer seen as a suboptimal option:

> *There was quite a lot of resistance previously by medical colleagues. They felt it would affect the relationship. [COVID-19] has turned that on its head quite quickly.* [Eleanor, diabetes nurse, Southern Trust, June 2020]

Clinicians using video talked about the advantages over phone for basic visual assessments ("eyeballing" the patient) and nonverbal interaction. This was reflected in the video activity data across sites, in which a large proportion of consultations were conducted within psychiatry/psychology and mental health services (13,952/109,401, 12.75%, of overall video activity), pediatric services (12,964/109,401, 11.85%), musculoskeletal and orthopedic services (11,953/109,401, 10.93%), and physiotherapy (11,636/109,401, 10.64%).

Crucially, many clinicians described *reflection in action* [53] through the pandemic, in which the use of video reshaped prior assumptions on the role of such technology within their clinical practice:

> *You have to be very pragmatic [over video] about how you do your assessments. And all those special tests we used to do in the consultation room, we're suddenly finding that we were not getting any value out of them. And we are now coming to realize, it is what the patients tell us, it is how they got out their chair to get their medication or how they turn around to tell their partner to turn the telly down—all these little functional cues that we are seeing. I'm finding video a lot more helpful than these clever tests we used to do.* [Larry, physiotherapist, Northern Trust, August 2020]

Christianson et al [45] describe how the uncertainties generated by rare events provide such opportunities to question previously held assumptions about core organizational functions and roles and provide a *strategic opening* for supporters of the initiative. Other opportunities and extended uses of video included remote multidisciplinary and group consultations, which would have previously been difficult to establish as physical encounters.

XSL•FO
RenderX

However, although the pandemic provided a context to try out and reflect on the use of video, as predicted by Gkeredakis et al [31], an opportunity through a crisis also comes with inherent tensions going forward. A key challenge raised related to subsequent shifts in the social and political forces as the risk of COVID-19 decreased:

> *The big question for me is whether this change is permanent or just temporary…Are we going to see clinicians just not wanting to go back to the way they were working before, or are we going to see a slip in the use?* [Beth, program manager, Northern Trust, April 2021]

The absence of prepandemic groundwork within Eastern Trust also brought uncertainties about the extent to which the new service model can be socially and technically stabilized within the organization:

> *We built a bridge—and it was a pretty good bridge to jump over the river—but we didn't build the foundations. We suddenly had the technology, instantly threw ourselves into it, and got over that river. But the bridge is a little insecure…As a major change project, you would never do that—implement a new technology, try to get hundreds on board, and then build the system to support it.* [Teresa, program manager, Eastern Trust, December 2020]

Disruption

Viewing the pandemic as disruption highlights how staff had to adapt to widespread displacement of social and material environments, including the need to work from home. Technology shortages also demanded improvised use of computing equipment (an advantage of Attend Anywhere was that it could run on a personal device, as nothing needed to be downloaded). Staff needed to reorganize work routines to accommodate and support remote consultations in the face of prolonged uncertainty and disruption. Adaptive capabilities allowed some service provision to continue but also distorted work practices and created new demands and unintended consequences.

Clinic workflows are complex and structured around various interacting routines (eg, booking appointments, arranging prior tests, processing patients through the clinic) and associated spatial and material structuring devices [54], which have largely evolved around the *physical* copresence of patients and staff. Indeed, the inertia of established routines was a major constraint on the use of video before the pandemic, in which apparent embedding resulted from elaborate workarounds. Hence, the sudden and extensive transition to *virtual* created significant disruption, in which a fix for one problem generated problems elsewhere in the system.

Particularly during the early stages of the pandemic, clinicians relied on the telephone to consult with patients:

> *We turned all into blanket telephone—and in a sense that has been the legacy—and we haven't moved forward with video clinics, because the telephone clinics have been kind of working…There is a logistical problem. If you see people at clinic, it is really easy because they turn up, there is a slip [of paper] with their name on the trolley, and you just come out, pick up the slip…And when you've seen them, they will leave—you can't double-see them. With the phone, we don't have that. You may get someone who does not answer, so you call the next person, and come back to them—but then of course, that next person wasn't expecting a call at that time, so then doesn't pick up…So at the moment, I'll start the [in clinic] appointments, and the registrar starts on the phone list. And then, when there are some less complicated patients in clinic, the registrar sees them, and I start at the bottom of the phone list and work up.* [Frank, neuro-oncology surgeon, Petroc Health, October 2020]

This extract highlights how the "installed base" both enabled and constrained the rapid reorganization of clinic routines, which, in turn, became locally embedded. Feldman [55] describes how such "patterning" (the process of reinforcing old and creating new patterns simply by taking action) initially emerges through the stringing together of anticipated-orientated actions (situated and highly influenced by the fit between available opportunities and their current abilities) under prolonged disruptive conditions. Although these "provisional adjustments" are enacted to make a situation work, over time, they form new patterns, roles, and ways of doing things.

Similarly, the rapid expansion of video demanded a high degree of adaptation and emergence in accordance with local needs and contingencies. One of the main areas of focus was the management of patient "entry" into their video appointment. The main video platform (Attend Anywhere) uses a consistent URL for each clinic virtual waiting area. In the initial phase, most patients were sent the relevant waiting area URL on an appointment letter, which they would be required to write into their internet browser. Although quick to implement, this arrangement was prone to problems with typing errors and incomplete knowledge of the new process:

> *They send them a sheet. I guess it must say what they have to do, but it's very limited. In fact, I don't think they always even send that. Sometimes, they just send a letter saying log onto this [URL] address. And 9 times out of 10, the first time they use it, they'll go onto Google Search…which just comes up with [Attend Anywhere] websites. And then, we say 'No, you put it into the address bar at the top…"* [Betty, audiology therapist, Eastern Trust, February 2021]

In this extract, the clinician, patient, and administrative staff lack the mutual awareness [56] and embodied knowledge needed to coordinate actions within this new digital space—aspects that are deeply embedded and taken for granted in traditional outpatient waiting areas.

However, over the course of the pandemic, various changes were made to help orientate patients, depending on local human and technical resources in place. Within Petroc, for example, existing (outsourced) text reminding services adapted the text

XSL•FO
RenderX

message sent to patients so they could click directly onto the URL, although this relied on the patient having a smart phone.

Another route of entry, eventually established within Petroc, Northern, and Southern Trusts, was the trust website, in which patients could click a button for the relevant virtual waiting area, although this relied on the patient being able to navigate the website and locate the correct button. The following extract describes a common problem in which patients would turn up in the wrong virtual waiting area due to difficulties distinguishing between overlapping specialties. This created a high degree of "invisible routine articulation work" (work necessary for dealing with anticipated contingencies but that is not formalized or documented [57]) for the clinician to locate and orientate them accordingly:

> *Patients select a waiting room to go into. But often, they would be waiting in orthopedics because they had seen that department previously…But we don't have access to that waiting room. So, we would need to contact them and tell them to come out and go into the MSK [musculoskeletal] waiting area.* [Nick, physiotherapist, Petroc Health, March 2021]

Organizational capability to endure, and even harness, locally driven adaptations depended on the people and connections in place to monitor and embed new practices. For example, the Northern Trust implementation team continued to engage with clinicians shortly after the rapid deployment of Attend Anywhere, during which they discovered some practitioners using an alternative platform (accuRx). This platform's texting and video call functions aligned better with their particular workflows and systems, and so it became supported at the system level.

> *It was kind of by chance and through discussions that we found some people had started using [accuRx]...We realized it was building a bit of momentum amongst some clinicians. So, we officially approved it, did a bit of comms, and developed some training on how to use that platform.* [Nathan, operational support , Northern Trust, March 2021]

Similarly, over the course of the pandemic, the developers of Attend Anywhere incorporated new design features based on user feedback, including the "consult now" function, allowing clinicians to send the URL link directly to patients via text or email.

The various sociotechnical arrangements to support patient entry reflect the generative tension between ostensive (generalized understanding) and performative (specific actions taken) aspects of routines [46], driven by clinicians' efforts to balance efficiency (managing their time and capacity) with flexibility (to enhance patient access) in this new virtual space. Pentland et al [58] highlight the paradoxical tension of *process multiplicity* (a single process incorporates many possible "paths"), in which ICT greatly expands the "space of possible paths"; different ICTs may be used at various points within the same process. Even a simple change (eg, adding a new button to enter a virtual waiting area) creates a host of new actions and pathways.

Figure 2 presents a narrative network of known possible actions involved in the process of patient entry into a video appointment, based on participants' accounts of using Attend Anyway. Seeing the process as a network of related actions highlights the extent of performative variation within this component of the workflow. Each step leads to various alternative actions that could happen next, enabled by the video software and associated technology. For example, some clinicians talked about a tendency to phone the patient if they were not in the virtual waiting area. From there, they might explain to the patient how to access it, or send login details electronically. Sometimes, patients would not answer the phone, so the clinician might leave a voice message (if the technology allows); some would also leave their direct contact number (as caller IDs were often withheld). Upon hearing the voice message, several actions could occur next; the patient may attempt to (re)connect to the virtual waiting area, they could phone the clinician back, they might call the clinic reception, or they might do nothing.

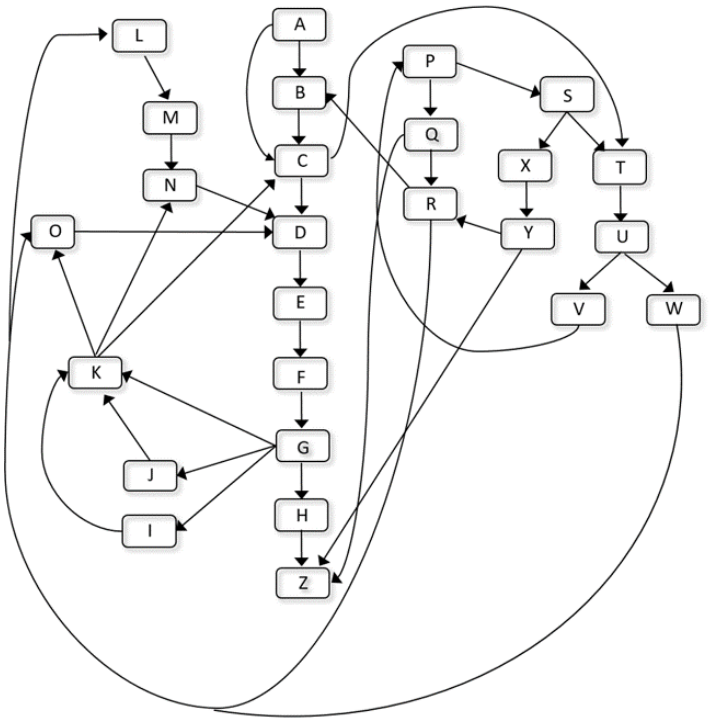
Another example of process extension was clinicians' use of the instant messaging feature to communicate with patients in the virtual waiting area (eg, to tell them whether the clinic is running late), in which patients may decide to wait or exit and re-enter at a later point. In some cases, the clinician and patient would jointly decide to abandon the video option and consult by phone instead (when clinically appropriate).

Although some service teams brought admin support into this process (eg, to monitor the virtual waiting area), it usually relied on the clinicians to manage. Participants talked about the need for a conceptual shift in the departmental management of these new digital spaces, if they are to be sustained:

> *If the patient has a face-to-face appointment, it isn't just them instantly in with a clinician and instantly out with a clinician…Outpatient departments manage the physical outpatient space, but now we've got all of this virtual outpatient space…which actually…other than the patient and the clinician linking up together, nobody is managing that space to make sure that link-up happens correctly…We need to be taking responsibility, full responsibility—like we do for the physical space—of that virtual space.* [Simon, ICT manager, Northern Trust, March 2021]

XSL•FO
RenderX

Figure 2. Narrative network analysis of patient entry into video consultations.



Actions
A: Patient retrieves login information [START]
B: Patient launches internet browser
C: Patient types URL address into internet browser
D: Patient inputs verification information
E: Patient completes audio/video tests
F: Patient clicks "start call" button
G: Patient enters virtual waiting area
H: Clinician selects patient listed in virtual waiting area
I: Clinician sends instant message to patient
J: Admin staff send instant message to patient
K: Patient leaves virtual waiting area
L: Patient goes to trust website
M: Patient locates clinic/speciality button
N: Patient clicks on clinic/speciality button
O: Patient clicks on URL link provided
P: Clinician phone calls patient
Q: Patient answers phone
R: Clinician provides login information to patient
S: Clinician leaves a phone voice message
T: Patient phone calls clinic
U: Admin staff answer phone
V: Admin staff notify clinician of patient phone call
W: Admin staff provide login information to patient
X: Patient phone calls clinician
Y: Clinician answers phone
Z: [END] Consultation begins

Exposure

Perspectives on exposure revealed infrastructural challenges and tensions to video consulting that have gone largely unnoticed. As Star [44] points out, infrastructure tends to exist in the background, invisible and taken for granted until it breaks down or gets in the way. Before the pandemic, clinicians using video described their use as being ad hoc, on a small scale, and with selected patients, rather than business as usual. The unprecedented demands of the rapid scale-up effort exposed important social, technical, and ethical issues to mainstreamed use—what Christianson et al [45] refer to as the "brutal audit of weakness" brought about by the novel demand of rare events.

Networks and servers came under significant strain during the first few months of the pandemic, resulting in periods of poor service reliability. Although these problems were dealt with promptly in collaboration with technology providers, local IT issues were poorly understood and more difficult to resolve:

> *It is difficult to test the consultation capacity, because it is very dependent on the volume of calls, how the platform is doing on that day…It is really difficult to test the actual performance because it can vary so much.* [Tanvi, operational support, Petroc Health, January 2021]

Additionally, access and dependability issues extended beyond the boundaries of the organization's IT infrastructure, including "digital poverty" (eg, no smartphone, no webcam, limited data package). It was reported that video tended to be lower among groups with a greater health care need and already facing health inequalities, such as older people, low health literacy, weak social networks, limited mobility, limited transport, and psychological and mental health problems:

> *Some people don't have [smartphones], they don't have laptops, iPads. You have to be kind of fairly well off to be able to access video consultation. We've definitely seen a difference in what type of patients can and can't access.* [Nick, physiotherapist, Petroc Health, March 2021]

Clinicians talked about limited privacy in family households, as well as differences in network connectivity. Particularly for patients living in rural areas, audio-video quality was unreliable, impacting the quality of the consultation and sometimes resulting in the use of the telephone as a backup:

> *Because if you're talking to someone and you're trying to explain something or you ask them a question and then you have to wait 10 seconds for them to answer, then you think, "Are they not answering, or is it they've not heard me or, you know, is it the delay?"* [Betty, audiology therapist, Eastern Trust, February 2021]

It was often difficult to anticipate the needs and wishes of patients in advance, requiring flexible use, alongside nondigital alternatives. Digital exclusion was not merely down to whether a person could use technology, but cut across multiple technical and social elements, as described next:

> *I had to say to her, "I can see your background [on video]." And I don't think she was comfortable with that…We have a large Asian population, where there is an element of larger families and smaller living spaces. If I access their space, I will not only be seeing them, I would be seeing their family. And they may not be comfortable with that…You have to be*

XSL•FO
RenderX

> *careful.* [Tanisha, hepatologist, Petroc Health, October 2020]

Mitigating digital exclusion was an effortful accomplishment between clinicians, patients, and their support networks. However, some early strategic work had begun within Northern Trust to establish local hubs for video appointments in partnership with community centers:

> *So now, if they need help accessing technology, they can get in touch with the community center—and we trained people in these community hubs to help people access the remote consultation. But also, in the community hub, they will get access to a full package of support, such as food packages…It has taken a while just engaging with these groups, making sure they are happy.* [Tessa, PPI engagement lead, Northern Trust, April 2021]

This extract reveals the role of collegial partnerships in order to coordinate and interface between the digital and material aspects of health care and community-based infrastructures. Perspectives on exposure, therefore, reveal not only weakness within the system but also strengths to be leveraged and built upon. In a similar vein, Sanner et al [59] draw on the horticultural notion of grafting to describe how infrastructural mergence between existing systems (from separate organizational boundaries) can be supported through the mutual adjustment and careful alignment of available resources and interests. Such perspective draws attention to how health IT organizations may need to interface with other organizational and professional boundaries and associated infrastructure in order to enhance the accessibility of remote digital care.

Discussion

Principal Findings

Gkeredakis et al's [31] 3 perspectives on crisis and digital change provide different vantage points to foreground how health IT infrastructures shaped, and were shaped by, the rapid scale-up of video consultations. Considering COVID-19 as an *opportunity* highlighted the institutional elements of infrastructure, in which conventional assumptions, norms, and governance structures were challenged. The pandemic brought about a collective understanding or *organizing vision* [50] for video, in which the organizational application shifted from a focus on efficiency and access to an issue of safety (infection control). This played a key role in legitimizing the technology and mobilizing resources to support adoption at scale. Beyond the pandemic, the organizing vision remains uncertain. Although many perceive a long-term role for video, the balance between the benefits and harms of such a service model will change, as the context moves from one dominated by risk of infection to concerns about the impact of remote consulting on patient safety and quality of care.

The focus on *disruption* drew attention to the structures and practices involved in the urgent response to the pandemic. With the displacement of deeply embedded routines, new performative patterns of action emerged, in which the virtual environment greatly expanded the "space of possible paths" within consultation workflows. A large space of possible paths makes routines more resilient to disruption (by minimizing the single points of failure), but it also makes the processes more difficult to replicate and manage [58]. The capacity to retain and mainstream positive elements of the change beyond the pandemic will require a judicious balance between the managerial top-down (clear processes and milestones) and the emergent bottom-up (responsive and contingent) approaches.

Accounts related to *exposure* reflected inherent properties of information infrastructures, as backgrounded and taken for granted. The accelerated scale-up of video revealed social and technical impediments to video consulting and the potential to accentuate unresolved health inequalities. This raises practical and ethical challenges to scaling up such a service model and conflicts with the policy talk of such technology as state of the art, efficient, and accessible [33]. Improving access will be an extensive and ongoing accomplishment, requiring close attention to the materiality and dependability of the installed base, as well as the social and cultural context of use.

Enabling Infrastructure Growth in the New Normal

Research on information infrastructures highlights how technology-supported change needs to be cultivated in a way that acknowledges challenging organizational needs and the inertia of the installed base [60-62]. The extreme and stochastic nature of the pandemic has provided a unique insight into the processes and mechanisms that surround infrastructural change, including the path-dependent way in which the system is built and added to, and, in turn, set organizations on a particular infrastructural path for remote consulting.

Despite the policy focus on digital [63], the reality is that most remote consultations were conducted using the telephone. The pandemic nevertheless brought about a significant shift in the use of video, which shows promise for particular clinical contexts. Building on our previous research on the challenges in introducing video consultations in health care settings, this study provides further insights for development and sustainability going forward.

First is the need to balance stability (and integration, security, and centralized control) on the one hand and openness to change (and emergence and local adaptability) on the other. Although regulatory changes during the pandemic may have brought a new-found agility to experiment and innovate, health care organizations must still ensure a high level of safety, security, and dependability [8]. Bygstad and Iden's [64] notion of *bimodal* governance offers a potential approach, in which "heavyweight" IT systems (traditional and sequential, emphasizing safety and accuracy) and "lightweight" IT applications (exploratory and nonlinear, emphasizing agility and speed) are managed separately under differing organizational structures, cultures, and regulatory mechanisms. Such a model has been shown to work well in hospital settings, in which a partitioning of IT departments—one part dealing with the traditional core IT system and the other with new applications—allowed for a "two-speed" approach to the co-evolution of these loosely couple subsystems [65].

XSL•FO
RenderX

Second, the extended use of remote consultations highlights a need to conceptualize information infrastructures beyond the boundaries of the health care organization. The ubiquitous use of extra-organizational ICT within professional work settings has prompted calls for more theoretical and empirical research into the intersection between institutional infrastructures and individual digital assemblages [66]. Video consulting during the pandemic has been found to occur across multiple settings outside of the usual organizational rules and sociomaterial arrangements [67]. Accordingly, our study suggests that more accessible, inclusive, and dependable approaches to remote care require a focus on how health IT infrastructures can merge with other systems across professional, technological, and organizational boundaries. This occurred at multiple levels within our study, including individual configurations (eg, improvised use of personal computing devices), interorganizational partnerships (eg, establishing community hub video facilities), and national strategies (eg, temporary use of Scottish platform servers, negotiations with mobile network providers). Such infrastructural "grafting" [59] involves the ability to identify opportune moments and parts of the installed base to leverage, and work effectively with stakeholders who retained some control over those parts of the infrastructure. This aligns with other studies that have focused on equity and virtual care during COVID-19, drawing attention to the need for a multilevel approach to improving access, including policy level strategies (eg, investment in internet and device access), organizational system design (eg, training and capacity building programs to support users), and community engagement (eg, digital literacy education and supporting resources) [68,69].

Third, close attention needs to be paid to routines and patterns of actions underpinning virtual appointments. Our findings on the disruptive forces of the pandemic resonate with other studies focusing on the redesign of care pathways using process-mapping methods [70,71]. Traditional workflow-style process maps can be useful in many respects. However, our analysis highlights that *virtual* consultations vastly expand the set of possible paths within an appointment process, and therefore, it is important to visualize and understand the multiple performances that *could* be generated in a process (as opposed to a single version of how the process *should* occur). Narrative networks may provide a useful methodological device to see the terrain of possible paths and purposefully increase variation in performance (for flexibility) or try to reduce it (for standardization and control). This could help identify areas to increase resilience and choice (eg, choosing between different remote consulting modalities or platforms), while also ensuring associated roles and processes are consistent and coordinated.

Finally, it is important to bridge competing institutional logics with an overall narrative or organizing vision [50] for video, through ongoing cross-stakeholder dialogue, involving policy, organizational, and patient engagement. The pandemic saw the emergence of intra- and interorganizational "communities of practice" [72] to share experiences and validate knowledge on the use of video. However, by their very nature, communities of practice tend to be informal and unstructured and so difficult to establish and sustain. As we have observed elsewhere [37], quality improvement collaboratives (structured approaches to meeting, sharing resources, and evaluating and informing strategic change) can help promote such learning, leverage national resources, and inform institutional elements of the infrastructure (eg, policies, regulation, funding, training). The varied use of video across clinical specialties highlights a need for further work in this regard to support shared learning as to how, and in what circumstances, this modality works well for patients. In addition, from a managerial standpoint, it will be important to help various organizational actors envisage how their actions interrelate and contribute to the overall narrative or collective outcome. Osmundsen and Bygstad [73] highlight the importance of sense making (process by which organizational actors engage in retrospective and prospective thinking to interpret reality [74]) and sense giving (process whereby organizational actors attempt to influence the meaning construction of other organizational actors [75]) to support a collective understanding of ongoing infrastructural changes. To guide the growth and expansion of remote consultations, it will be important to understand and foster communicative dynamics that promote "collective minding" [76] across various aspects of the supporting infrastructure.

Strengths and Limitations

The strength of this study is that we undertook research at 4 sites before the pandemic, through which we had already identified the importance of information infrastructures on the implementation and use of video. This provided a unique opportunity to study crisis-engendered changes from the early stages. In addition, the case studies were conducted within a wider program of research on remote consultations across the United Kingdom, providing a wider national context to the organizational settings.

Pandemic restrictions meant that we could not undertake ethnographic work during this phase, and our data collection during this time was affected by the unprecedented pressures on NHS staff, raising potential sample biases toward those individuals with time available to speak with us. However, our positive and longstanding relationship with these sites helped mitigate such issues; the researcher had previously visited all sites on numerous occasions before the pandemic, adapted interview schedules, and was able to draw on multiple sources of data.

A further limitation of this study includes the question of how far we can generalize from our 4 case studies on the rapid scale-up of video consultations. Our theoretical analysis helped explain the empirical data on information infrastructures in our case sites but may not explain all aspects of such infrastructure in all contexts. We chose to focus on video consultations as a topic of academic interest because it exemplifies a promising innovation that has taken decades to scale up. We encourage others to apply the same theoretical lens to explore crisis engendered changes in other digital health contexts.

Conclusion

Gkeredakis et al's [31] 3 perspectives on crisis and digital change (as opportunity, disruption, and exposure) provided a useful lens to understanding the role of information infrastructures during the rapid scale-up for video consultations,

XSL•FO
RenderX

as well as the challenges and tensions going forward. Foregrounding crisis-engendered change helps explain how the infrastructures constrained and enabled the rapid scale-up and highlights how health care organizations can build on the gains going forward. To extend and sustain the use of video in the long term, it will be important to enable incremental systemic change through agile governance and knowledge transfer pathways, allow greater flexibility and process multiplicity within virtual clinic workflows, attend to the materiality and dependability of the installed base both within and beyond organizational boundaries, and maintain an overall narrative within which the continued use of video can be framed.

Acknowledgments

We thank the research participants, including patients, health care staff, and wider stakeholders, who all gave generously of their time while working under considerable pressure.

Fieldwork for this study was funded by the Health Foundation, an independent charity committed to bringing about better health care for people in the United Kingdom under 2 research programs (496288 and 2133488). Additional funding for TG's, SES's, and JW's salaries over the course of the study was provided by the UK National Institute for Health Research Oxford Biomedical Research Centre (BRC-1215-20008).

Conflicts of Interest

None declared.

References

1. Chongmelaxme B, Lee S, Dhippayom T, Saokaew S, Chaiyakunapruk N, Dilokthornsakul P. The effects of telemedicine on asthma control and patients' quality of life in adults: a systematic review and meta-analysis. J Allergy Clin Immunol Pract 2019 Jan;7(1):199-216.e11. [doi: 10.1016/j.jaip.2018.07.015] [Medline: 30055283]
2. Armfield NR, Bradford M, Bradford NK. The clinical use of Skype: for which patients, with which problems and in which settings? A snapshot review of the literature. Int J Med Inform 2015 Oct;84(10):737-742. [doi: 10.1016/j.ijmedinf.2015.06.006] [Medline: 26183642]
3. Hansen C. Video consultations as add-on to standard care among patients with type 2 diabetes not responding to standard regimens: a randomized controlled trial. Eur J Endocrinol 2017;176(6):727-736. [doi: 10.1530/eje-16-0811]
4. Harris MA, Freeman KA, Duke DC. Seeing is believing: using Skype to improve diabetes outcomes in youth. Diabetes Care 2015 Aug;38(8):1427-1434. [doi: 10.2337/dc14-2469] [Medline: 26033508]
5. Katz IJ, Pirabhahar S, Williamson P, Raghunath V, Brennan F, O'Sullivan A, et al. iConnect CKD - virtual medical consulting: a web-based chronic kidney disease, hypertension and diabetes integrated care program. Nephrology (Carlton) 2018 Jul 24;23(7):646-652. [doi: 10.1111/nep.13070] [Medline: 28474361]
6. Gentry MT, Lapid MI, Rummans TA. Geriatric telepsychiatry: systematic review and policy considerations. Am J Geriatr Psychiatry 2019 Feb;27(2):109-127. [doi: 10.1016/j.jagp.2018.10.009] [Medline: 30416025]
7. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
8. Basu A, Kuziemsky C, de Araújo Novaes M, Kleber A, Sales F, Al-Shorbaji N, et al. Telehealth and the covid-19 pandemic: international perspectives and a health systems framework for telehealth implementation to support critical response. Yearb Med Inform 2021 Aug 21;30(1):126-133 [FREE Full text] [doi: 10.1055/s-0041-1726484] [Medline: 33882598]
9. Nittari G, Savva D, Tomassoni D, Tayebati SK, Amenta F. Telemedicine in the covid-19 era: a narrative review based on current evidence. Int J Environ Res Public Health 2022 Apr 22;19(9):5101 [FREE Full text] [doi: 10.3390/ijerph19095101] [Medline: 35564494]
10. Due TD, Thorsen T, Andersen JH. Use of alternative consultation forms in Danish general practice in the initial phase of the COVID-19 pandemic - a qualitative study. BMC Fam Pract 2021 Jun 02;22(1):108 [FREE Full text] [doi: 10.1186/s12875-021-01468-y] [Medline: 34078281]
11. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
12. Day K, Hunter I, Rao V, Russell G, Roskvist R, Moir F, et al. Survey protocol for exploring video and phone use in Aotearoa New Zealand general practice: considerations for future telehealth. BMJ Health Care Inform 2021 Mar 22;28(1):e100309 [FREE Full text] [doi: 10.1136/bmjhci-2020-100309] [Medline: 33753326]
13. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the covid-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]

XSL•FO
RenderX

14. Taylor A, Caffery LJ, Gesesew HA, King A, Bassal A, Ford K, et al. How Australian health care services adapted to telehealth during the covid-19 pandemic: a survey of telehealth professionals. Front Public Health 2021;9:648009 [FREE Full text] [doi: 10.3389/fpubh.2021.648009] [Medline: 33718325]
15. Huibers L, Bech BH, Kirk UB, Kallestrup P, Vestergaard CH, Christensen MB. Contacts in general practice during the COVID-19 pandemic: a register-based study. Br J Gen Pract 2022 Aug 30:BJGP.2021.0703. [doi: 10.3399/bjgp.2021.0703]
16. 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 Leaders 2020 May 18;4(3):120-123. [doi: 10.1136/leader-2020-000262]
17. Smith K, Ostinelli E, Macdonald O, Cipriani A. COVID-19 and telepsychiatry: development of evidence-based guidance for clinicians. JMIR Ment Health 2020 Aug 28;7(8):e21108 [FREE Full text] [doi: 10.2196/21108] [Medline: 32658857]
18. Johns G, Tan J, Burhouse A, Ogonovsky M, Rees C, Ahuja A. A visual step-by-step guide for clinicians to use video consultations in mental health services: NHS examples of real-time practice in times of normal and pandemic healthcare delivery. BJPsych Bull 2020 Dec 08;44(6):277-284 [FREE Full text] [doi: 10.1192/bjb.2020.71] [Medline: 33213560]
19. Duncan LJ, Cheng KF. Modifications to the delivery of NHS face-to-face general practice consultations during the COVID-19 pandemic in England. F1000Res 2021 Aug 20;10:261. [doi: 10.12688/f1000research.52161.2]
20. Nitkunan A, Paviour D, Nitkunan T. COVID-19: switching to remote neurology outpatient consultations. Pract Neurol 2020 May 04;20(3):222-224 [FREE Full text] [doi: 10.1136/practneurol-2020-002571] [Medline: 32366596]
21. Jayadev C, Mahendradas P, Vinekar A, Kemmanu V, Gupta R, Pradhan Z, et al. Tele-consultations in the wake of COVID-19 – suggested guidelines for clinical ophthalmology. Indian J Ophthalmol 2020;68(7):1316. [doi: 10.4103/ijo.ijo_1509_20]
22. Hohenschurz-Schmidt D, Scott W, Park C, Christopoulos G, Vogel S, Draper-Rodi J. Remote management of musculoskeletal pain: a pragmatic approach to the implementation of video and phone consultations in musculoskeletal practice. Pain Rep 2020 Nov 15;5(6):e878. [doi: 10.1097/pr9.0000000000000878]
23. Greenhalgh T, Rosen R, Shaw SE, Byng R, Faulkner S, Finlay T, et al. Planning and evaluating remote consultation services: a new conceptual framework incorporating complexity and practical ethics. Front Digit Health 2021 Aug 13;3:726095 [FREE Full text] [doi: 10.3389/fdgth.2021.726095] [Medline: 34713199]
24. Galway N, Stewart G, Maskery J, Bourke T, Lundy CT. Fifteen-minute consultation: a practical approach to remote consultations for paediatric patients during the COVID-19 pandemic. Arch Dis Child Educ Pract Ed 2021 Aug 24;106(4):206-209. [doi: 10.1136/archdischild-2020-320000] [Medline: 32839193]
25. Ferreira D, Azevedo E, Araújo R. Teleneurology in Parkinson's disease: a step-by-step video guide. Acta Neurol Scand 2021 Aug 09;144(2):221-225. [doi: 10.1111/ane.13429] [Medline: 33835482]
26. Sandbæk A, Christensen LD, Larsen LL, Primholdt Christensen N, Kofod FG, Guassora AD, et al. Guidance for implementing video consultations in Danish general practice: rapid cycle coproduction study. JMIR Form Res 2021 Nov 24;5(11):e27323 [FREE Full text] [doi: 10.2196/27323] [Medline: 34821560]
27. Greenhalgh T, Wherton J, Shaw S, Papoutsi C, Vijayaraghavan S, Stones R. Infrastructure revisited: an ethnographic case study of how health information infrastructure shapes and constrains technological innovation. J Med Internet Res 2019 Dec 19;21(12):e16093 [FREE Full text] [doi: 10.2196/16093] [Medline: 31855184]
28. Star SL. The ethnography of infrastructure. Am Behav Sci 2016 Jul 27;43(3):377-391. [doi: 10.1177/00027649921955326]
29. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Serv Deliv Res 2018 Jun;6(21):1-136. [doi: 10.3310/hsdr06210] [Medline: 29939519]
30. Boin A, Comfort L, Demchack C. The rise of resilience. In: Comfort L, Boin A, Demchack C, editors. Designing Resilience: Preparing for Extreme Events. Pittsburgh, PA: University of Pittsberg; 2010:1-12.
31. Gkeredakis M, Lifshitz-Assaf H, Barrett M. Crisis as opportunity, disruption and exposure: exploring emergent responses to crisis through digital technology. Inf Organ 2021 Mar;31(1):100344. [doi: 10.1016/j.infoandorg.2021.100344]
32. The Topol Review. Preparing the Healthcare Workforce to Deliver the Digital Future. URL: https://topol.hee.nhs.uk [accessed 2022-10-17]
33. NHS. NHS Long Term Plan. URL: https://www.longtermplan.nhs.uk/ [accessed 2022-10-27]
34. Honeyman M, Dunn P, McKenna H. Digital NHS? An Introduction to the Digital Agenda and Plans for Implementation. URL: https://www.kingsfund.org.uk/publications/digital-nhs [accessed 2022-10-27]
35. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC Med 2018 May 14;16(1):66 [FREE Full text] [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
36. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021 Jan 26;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
37. 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]

XSL•FO
RenderX

38. Ham C. Covid-19 exposes NHS fault lines in England. BMJ 2020 Oct 14;371:m3922. [doi: 10.1136/bmj.m3922] [Medline: 33055106]
39. NHS England. Video Consultations in Secondary Care. URL: https://www.england.nhs.uk/outpatient-transformation-programme/video-consultations-in-secondary-care/ [accessed 2022-10-17]
40. Burn S. What Happened to English NHS Hospital Activity during the COVID-19 Pandemic?. URL: https://ifs.org.uk/publications/what-happened-english-nhs-hospital-activity-during-covid-19-pandemic [accessed 2022-10-17]
41. Edge Health. Video consultations in Secondary Care: A Report by Edge Health for NHS England and NHS Improvement. URL: https://www.edgehealth.co.uk/outpatient-video [accessed 2022-10-17]
42. Bagenal J. The delivery plan for tackling the covid-19 backlog of elective care falls short. BMJ 2022 Apr 19;377:o995. [doi: 10.1136/bmj.o995] [Medline: 35440437]
43. Shaw SE, Hughes G, Wherton J, Moore L, Rosen R, Papoutsi C, et al. Achieving spread, scale up and sustainability of video consulting services during the covid-19 pandemic? findings from a comparative case study of policy implementation in England, Wales, Scotland and Northern Ireland. Front Digit Health 2021 Dec 20;3:754319 [FREE Full text] [doi: 10.3389/fdgth.2021.754319] [Medline: 34988546]
44. Star SL. Infrastructure and ethnographic practice: working on the fringes. Scand J Inf Syst 2002;14(2):107-122 [FREE Full text]
45. Christianson MK, Farkas MT, Sutcliffe KM, Weick KE. Organ Sci 2009 Oct;20(5):846-860. [doi: 10.1287/orsc.1080.0389]
46. Feldman MS, Pentland BT. Reconceptualizing organizational routines as a source of flexibility and change. Admin Sci Q 2016 Jun 24;48(1):94-118. [doi: 10.2307/3556620]
47. Swinglehurst D, Greenhalgh T, Myall M, Russell J. Ethnographic study of ICT-supported collaborative work routines in general practice. BMC Health Serv Res 2010 Dec 29;10(1):348 [FREE Full text] [doi: 10.1186/1472-6963-10-348] [Medline: 21190583]
48. Pentland BT, Feldman MS. Narrative networks: patterns of technology and organization. Organ Sci 2007 Oct;18(5):781-795. [doi: 10.1287/orsc.1070.0283]
49. Thornton P, Ocasio W, Lounsbury M. The Institutional Logics Perspective: A New Approach to Culture, Structure and Process. Oxford, UK: Oxford University Press; 2012.
50. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012 Jul;2(4):e001574 [FREE Full text] [doi: 10.1136/bmjopen-2012-001574] [Medline: 22815469]
51. Scott WR. Institutions and Organizations: Ideas, Interests, and Identities. 4th Edition. Thousand Oaks, CA: Sage; 2014.
52. Rogers E. Diffusion of Innovations. 5th Edition. New York, NY: Free Press; 2003.
53. Schon D. The Reflective Practitioner: How Professionals Think in Action. New York, NY: Basic Books; 1983.
54. Greenhalgh T. Role of routines in collaborative work in healthcare organisations. BMJ 2008 Nov 17;337:a2448. [doi: 10.1136/bmj.a2448] [Medline: 19015186]
55. Feldman MS, Worline M, Baker N, Lowerson Bredow V. Continuity as patterning: A process perspective on continuity. Strateg Organ 2021 Oct 25;20(1):80-109. [doi: 10.1177/14761270211046878]
56. Reddy M, Dourish P, Pratt W. Coordinating heterogeneous work: Information and representation in medical care. 2001 Presented at: Proceedings of the Seventh Eurpoean Conference on Computer Supported Cooperative Work; September 16-20, 2001; Bonn, Germany p. 16-20. [doi: 10.1007/0-306-48019-0_13]
57. Hampson I, Junor A. Invisible work, invisible skills: interactive customer service as articulation work. New Tech Work Empl 2005 Jul;20(2):166-181. [doi: 10.1111/j.1468-005x.2005.00151.x]
58. Pentland BT, Mahringer CA, Dittrich K, Feldman MS, Wolf JR. Process multiplicity and process dynamics: weaving the space of possible paths. Organ Theory 2020 Sep 29;1(3):263178772096313. [doi: 10.1177/2631787720963138]
59. Sanner T, Manda T, Nielsen P. Grafting: balancing control and cultivation in information infrastructure innovation. JAIS 2014 Apr;15(4):220-243. [doi: 10.17705/1jais.00356]
60. Grisot M, Hanseth O, Thorseng A. Innovation of, in, on infrastructures: articulating the role of architecture in information infrastructure evolution. J Assoc Inf Syst 2014 Apr;15(4):197-219. [doi: 10.17705/1jais.00357]
61. Maniatopoulos G, Procter R, Llewellyn S, Harvey G, Boyd A. Moving beyond local practice: reconfiguring the adoption of a breast cancer diagnostic technology. Soc Sci Med 2015 Apr;131:98-106. [doi: 10.1016/j.socscimed.2015.02.036] [Medline: 25769108]
62. Berg M. On distribution, drift and the electronic medical record. Dordrecht: Springer; 1997 Presented at: Proceedings of the Fifth European Conference on Computer Supported Cooperative Work; September 7-11, 1997; Lancaster, UK. [doi: 10.1007/978-94-015-7372-6_10]
63. Hancock M. The Future of Health Care (Speech, 30th July 2020). URL: https://www.gov.uk/government/speeches/the-future-of-healthcare [accessed 2022-10-17]
64. Bygstad B, Iden J. A governance model for managing lightweight IT. 2017 Presented at: Proceedings of the World Conference on Information Systems and Technologies; April 11-13, 1997; Maderia, Portugal. [doi: 10.1007/978-3-319-56535-4_39]
65. Bygstad B, Øvrelid E. Architectural alignment of process innovation and digital infrastructure in a high-tech hospital. Eur J Inf Syst 2020 Feb 27;29(3):220-237. [doi: 10.1080/0960085x.2020.1728201]

XSL•FO
RenderX

66. Jarrahi MH, Sawyer S, Erickson I. Digital assemblages, information infrastructures, and mobile knowledge work. J Inf Technol 2021 Oct 28;37(3):230-249. [doi: 10.1177/02683962211050943]
67. Hughes G, Moore L, Maniatopoulos G, Wherton J, Wood GW, Greenhalgh T, et al. Theorising the shift to video consulting in the UK during the COVID-19 pandemic: analysis of a mixed methods study using practice theory. Soc Sci Med 2022 Sep 14;311:115368 [FREE Full text] [doi: 10.1016/j.socscimed.2022.115368] [Medline: 36152402]
68. 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]
69. 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]
70. Peynetti VP, Gupta G, Gupte G, Carson N, Venter J. Rapid implementation of telepsychiatry in a safety-net health system during Covid-19 using Lean. NEJM Catal Innov Care Deliv 2020;1(4):1-20. [doi: 10.1056/CAT.20.0319]
71. Yang L, Brown-Johnson CG, Miller-Kuhlmann R, Kling SM, Saliba-Gustafsson EA, Shaw JG, et al. Accelerated launch of video visits in ambulatory neurology during COVID-19. Neurology 2020 Jul 01;95(7):305-311. [doi: 10.1212/wnl.0000000000010015]
72. Lave J, Wenger E. Situated Learning: Legitimate Peripheral Participation. Cambridge, UK: Cambridge University Press; 1991.
73. Osmundsen K, Bygstad B. Making sense of continuous development of digital infrastructures. J Inf Technol 2021 Oct 09;37(2):144-164. [doi: 10.1177/02683962211046621]
74. Weick K. Sensemaking in Organizations. Thousand Oaks, CA: Sage; 1995.
75. Gioia DA, Chittipeddi K. Sensemaking and sensegiving in strategic change initiation. Strateg Manag J 1991 Sep;12(6):433-448. [doi: 10.1002/smj.4250120604]
76. Cooren F. The communicative achievement of collective minding. Manag Commun Q 2016 Aug 17;17(4):517-551. [doi: 10.1177/0893318903262242]

Abbreviations

AHP: allied health professional
ICT: information and communication technology
NHS: National Health Service
NHSEI: NHS England and NHS Improvement

Edited by T Leung, G Eysenbach; submitted 04.09.22; peer-reviewed by U Kirk, PSS Lee; comments to author 10.10.22; revised version received 24.10.22; accepted 24.10.22; published 10.11.22.

Please cite as:
Wherton J, Greenhalgh T, Hughes G, Shaw SE
The Role of Information Infrastructures in Scaling up Video Consultations During COVID-19: Mixed Methods Case Study Into Opportunity, Disruption, and Exposure
J Med Internet Res 2022;24(11):e42431
URL: https://www.jmir.org/2022/11/e42431
doi:10.2196/42431
PMID:36282978

©Joseph Wherton, Trisha Greenhalgh, Gemma Hughes, Sara E Shaw. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX